Observing CAPTCHAS “in the Wild”

Andrew Searles *Gene Tsudik*

Abstract

For nearly two decades, CAPTCHAS have been widely used as a means of protection against bots. As their use grew, techniques to defeat or bypass them continued to improve, while, CAPTCHAS evolved in terms of sophistication and diversity, becoming increasingly difficult to solve for both bots (machines) and humans. Given this long-standing and still-ongoing arms race, it is both timely and important to investigate the user burden of solving current CAPTCHAS, and how they are perceived by users.

This work [27] explores CAPTCHAS *in the wild* by evaluating solving performance and user perceptions of *unmodified currently-deployed* CAPTCHAS. We obtain this data through manual inspection of popular websites and a large-scale user study wherein $1,400$ participants collectively solved $14,000$ CAPTCHAS. Results show significant differences between most popular types of CAPTCHAS: surprisingly, solving time and user perception are not always correlated. We performed a comparative analysis of effects of *experimental context*, focusing on the difference between solving CAPTCHAS directly as opposed to solving them as part of a more natural task, such as account creation. While there were several potential confounding factors, results show that experimental context might impact this task, and must be taken into account in future CAPTCHA studies. Finally, we investigated CAPTCHA-induced user task *abandonment* by analyzing sessions where participants began, and did not complete, the task.

1 Introduction

Bots pose significant challenges for, and dangers to, many website operators and service providers. Masquerading as legitimate human users, they are often programmed to scrape content, create accounts, post fake comments or reviews, consume scarce resources (e.g., tickets or reservations), or generally (ab)use other website functionality intended for human use [10, 20]. If left unchecked, bots can perform these nefarious actions at scale. CAPTCHAS are a widely-deployed defense mechanism that attempts to prevent bots from interacting with websites by forcing each (purported) user to perform a task that usually involves responding to a challenge [2]. Ideally, the task should be straightforward for humans, yet difficult for machines [35].

Earliest CAPTCHAS asked users to transcribe random distorted text from an image. However, advances in computer vision and machine learning have dramatically increased ability of bots to recognize distorted text [13,17,37], and, by 2014, automated tools achieved $> 99\%$ accuracy [16,29]. Some bot operators outsource solving to so-called CAPTCHA *farms* – sweatshop-like operations where humans are paid to solve CAPTCHAS [24] in real time.

In response, CAPTCHA technology evolved significantly over the years. Popular current CAPTCHA tasks include object recognition (e.g., “select squares with...”), parsing distorted text, puzzle solving (e.g., “slide the block...”), and user behavior analysis [16, 29]. It is therefore critical to understand and quantify how long users take to solve current CAPTCHAS, and how CAPTCHAS are perceived by users.

Several prior research efforts explored CAPTCHA solving times, e.g., [5, 8, 12, 15, 25, 34]. For example, over a decade ago, Bursztein et al. [8] performed a large-scale user study, with $> 1,100$ participants from Amazon Mechanical Turk (MTurk) [1] as well as CAPTCHA farms. Results showed that CAPTCHAS were often more difficult, or took longer to solve, than expected. There was a loose correlation between time-to-annoyance and abandonment, with higher abandonment rates observed for CAPTCHAS that took longer to solve. The same study also showed several demographic trends, e.g., users outside the US typically took longer to solve English-language CAPTCHAS. However, since the study from Bursztein et al. [**?**], the CAPTCHA ecosystem changed substantially: new CAPTCHA types emerged, input methods evolved, and Web use boomed.

Building upon, and complementing, prior work, this effort evaluates CAPTCHAS *in the wild* – specifically, solving times and user perceptions of *unmodified* (i.e., not re-implemented) *currently-deployed* CAPTCHA types. The first phase innvolved

Table 1: Summary of research questions and main findings.

	Findings supporting prior work	Findings contradicting prior work	New findings on CAPTCHAS
RQ1: How fast users solve different types of CAPTCHAS?	Solving time across CAPTCHA types has a large degree of variance. [5, 8, 12]		Humans are slower than bots at solving CAPTCHAS.
RQ2: What CAPTCHA types do users prefer?	Solving time is not correlated with user preference. [12, 21, 33]		
RQ3: Does experimental context affect solving time?			Solving time is heavily influenced by experimental context, with differences in means up to 57.5%.
RQ4: Do demographics affect solving time?	Age has an effect on solving time. [8]	Self-reported education level does not correlate with solving time. [8]	
RQ5: Does experimental context influence abandonment?	High abandonment rates observed in CAPTCHA user studies. [8]		Experimental context directly affects the rate of abandonment.

a manual inspection of 200 popular websites, based on Alexa top websites list, in order to ascertain: (1) *how many* websites use CAPTCHAS, and (2) *what types* of CAPTCHAS they use. Next, a $1,000$-participant user study was conducted using Amazon MTurk, wherein each participant was required to solve 10 different types of CAPTCHAS. Collected information included CAPTCHA solving times, relative preferences for CAPTCHA types, kinds of devices used, and certain demographic information.

One notable aspect of the user study is measuring the impact of experimental context on CAPTCHA solving times. Half of the participants were directly asked to solve CAPTCHAS, while the other half were asked to create accounts, which involved solving CAPTCHAS as part of the task. The latter setting was designed to measure CAPTCHA solving times *in the context* of a typical web activity.

One inherent limitation of any user study, especially when using MTurk, is the inability of ensuring that all participants who begin the task will actually complete it. Thus, all our results should be interpreted as referring to *users who were willing to solve* CAPTCHAS, rather than users in general.

Having observed that some participants began, though did not complete, the study, we conducted a secondary MTurk study in order to quantify how many users abandon their intended web activity when confronted with various types of CAPTCHAS. We believe that CAPTCHA-induced *user abandonment* is an important – yet understudied – consideration, since every abandoned task (e.g., purchase, search, or account creation) represents a potential loss for the website.

2 Results & Analysis

We now present user study results. Note that, unless indicated otherwise, results are based on the full set of participants.

2.1 Solving times

Recall **RQ1:** *How long do human users take to solve different types of* CAPTCHAS*?* Figure 1 shows the distribution of solving times for each CAPTCHA type. We observed a few extreme outliers where the participant likely switched to another task before returning to the study. Therefore we filtered out the highest 50 solving times per CAPTCHA type, out of $1,000$ total.

For reCAPTCHA, selection between image- or click-based tasks is made dynamically by Google. While we know that 85% (easy) and 71% (hard) of participants were shown (respectively) only a click-based CAPTCHA, the exact task-to-participant mapping is not revealed to website operators. We therefore assume that slowest solving times correspond to image-based tasks since they have to have involved both a click-based task and an image-based task. After disambiguation, click-based reCAPTCHA had the lowest median solving time at 3.7 seconds. Curiously, there was little difference between easy and difficult settings.

The next lowest median solving time were for distorted text CAPTCHAS. As expected, simple distorted text CAPTCHAS were solved the fastest. Masked and moving versions had very similar, though higher, solving times. For hCAPTCHA, there is a clear distinction between easy and difficult settings. The latter consistently served either a harder image-based task or increased the number of rounds. However, for both settings, fastest solving times are similar to those of reCAPTCHA and distorted text. Finally, game-based and slider-based CAPTCHAS generally yielded higher median solving times, though some participants still solved these relatively quickly, e.g., in < 10 seconds.

With the exception of reCAPTCHA (click) and distorted text, solving times exhibit relatively high variance. Some variance is expected, especially, since these results encompass all input modalities across both direct and contextualized settings. However, *relative differences in variances* indicate that, while some types of CAPTCHAS are consistently solved

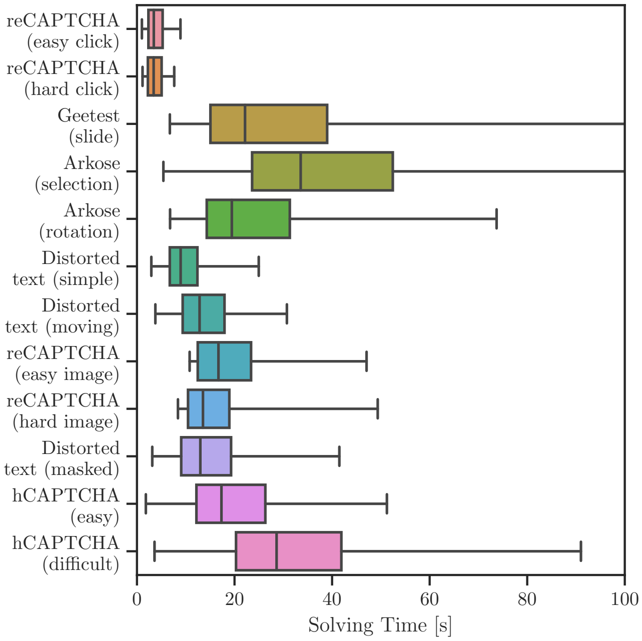

Figure 1: Solving times for various types of CAPTCHAS. Boxes show the middle 50% of participants, and whiskers show the filtered range. Black vertical lines show the median.

quickly, most have a range of solving times across the user population.

2.2 Preferences analysis

We now switch to **RQ2:** *What* CAPTCHA *types do users prefer?* Figure 2 shows participants' CAPTCHA preference responses after completing solving tasks. CAPTCHA types are sorted from most to least preferred by overall preference score, which is computed by summing numeric scores. Since easy and difficult settings of hCAPTCHA are visually indistinguishable, we could only ask participants for one preference.

As expected, participants tend to prefer CAPTCHAS with faster solving times. For example, reCAPTCHA (click) has the lowest median solving time and the highest user preference. However, surprisingly, this trend does not hold for game-based and slider-based CAPTCHAS, since they received some of the highest preference scores, even though they typically took longer than other types. This suggests that factors beyond solving time might contribute to participants' preference scores. Notably, no single CAPTCHA type is either universally liked or disliked. For example, even the top-rated click-based reCAPTCHA, was rated 1 or 2 out of 5 by 18.9% of participants. Similarly, over 31.0% rated hCAPTCHA 4 or 5, although it had the lowest overall preference score.

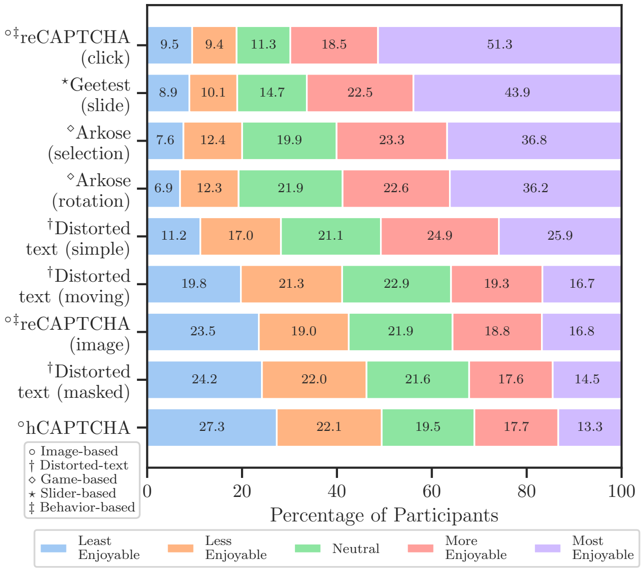

Figure 2: Participant-reported preference scores for different types of CAPTCHAS, sorted from highest to lowest.

2.3 Direct vs. contextualized setting

We now consider **RQ3:** *Does experimental context affect solving time?* Figure 3 shows histograms of CAPTCHA solving times for participants in the direct vs. contextualized settings. In every case except one, the mean solving time is lower in the direct setting. In most cases, the distribution from the contextualized setting has more participants with longer solving times, i.e., a longer tail.

The largest statistically significant difference is in reCAPTCHA (easy click), where the mean solving time grows by 1.8 seconds (57.5%). Second is Arkose (rotation), where it grows by 10 seconds (56.1%). Across all CAPTCHA types, the average increase from direct to contextualized is 26.7%. Similarly, the mean solving time for reCAPTCHA (easy image) increased by 63.6% in the contextualized setting, showing the largest increase. However that is not statistically significant. This is likely due to the skew of participants in direct and contextualized versions receiving image-challenges, which is controlled by Google. Easy images were shown to 8.9% of contextualized and 17.2% of direct setting participants, while hard images were shown to 25.5% and 30% respectively, resulting in different population sizes.

On the other hand, hCAPTCHA (difficult), which has the highest median solving time overall, showed no significant difference in mean solving time between direct and contextualized settings. This may be attributable to the difficulty of solving this type of CAPTCHA, regardless of the setting. Results of Kruskal-Wallis tests confirm that there are statistically significant differences in mean solving times for all CAPTCHA types ($p < 0.001$) except Geetest, reCAPTCHA (image) and hCAPTCHA (difficult).

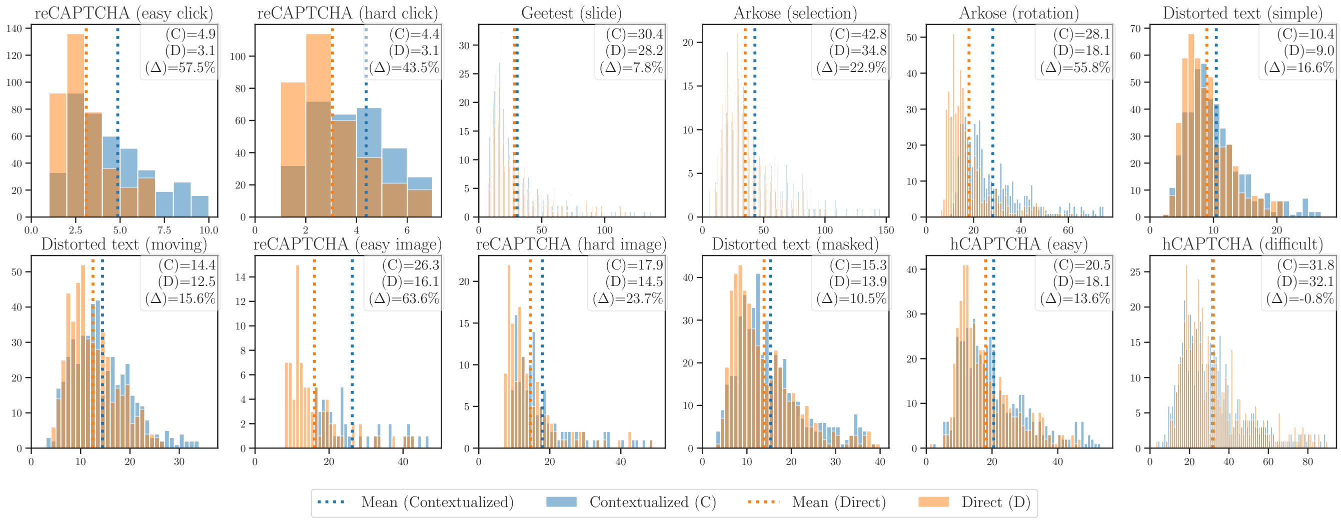

Figure 3: CAPTCHA solving times for direct (D) vs. contextualized (C) user study settings. Horizontal axis shows solving time in seconds, quantized into one-second buckets, and vertical axis shows number of participants.

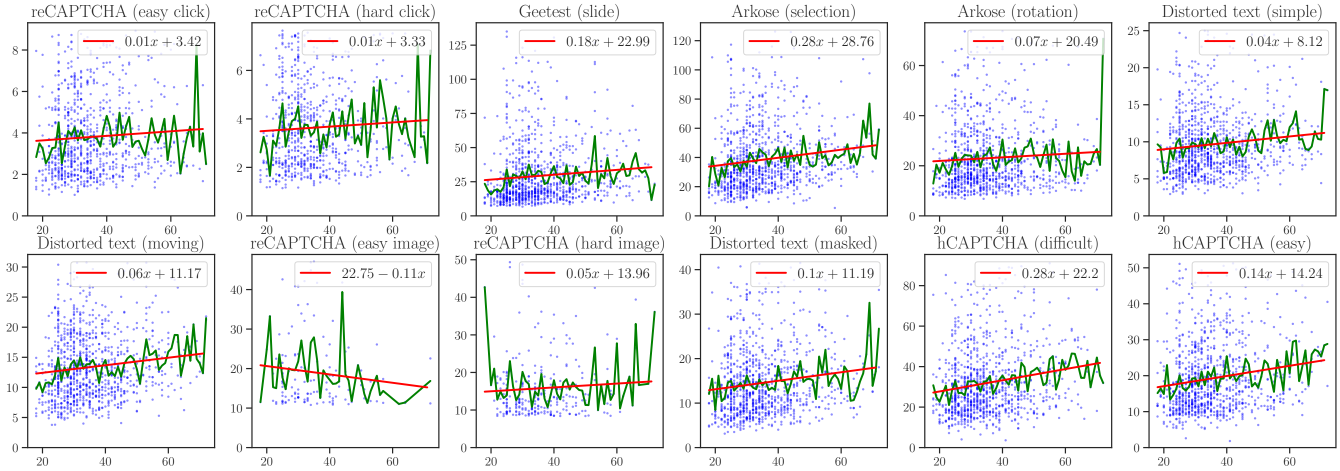

Figure 4: Effects of age in CAPTCHA solving time. Horizontal axis shows age and vertical axis shows solving time. Red line shows linear fit of data points and green line shows average solving time per age.

While there were several potential confounding factors in our study, results suggest that experimental context can significantly impact solving times, and must therefore be taken into account in the design of future user studies.

2.4 Effects of age

Next, we turn to **RQ4:** *Do demographics affect solving time?* The only statistically significant results we observed are related to age. Figure 4 shows its effect on solving time. The green line is the average solving time for each age, and the red line is a linear fit minimizing mean square error. For all types, except reCAPTCHA (easy image), there is a trend of younger participants having lower average solving times. This agrees with prior results [8] and is especially noticeable in hCAPTCHA, Arkose (selection), and Geetest.

2.5 Accuracy of CAPTCHAS

Table 2 contrasts our participants' solving times and accuracy against those of automated bots reported in the literature. Interestingly, these results suggest that bots *can* outperform humans, both in terms of solving time and accuracy, across all CAPTCHA types. Our use unmodified real-world CAPTCHAS means that we only have accuracy results for a subset of CAPTCHA types, e.g., neither Geetest nor Arkose provide accuracy information. For the same reason, our accuracy results also include participants who only partially completed the study.

reCAPTCHA: Accuracy of image classification was 81% and 81.7% for easy and hard settings, respectively. Surprisingly, the difficulty seems not to impact accuracy.

hCAPTCHA: Accuracy was 81.4% and 70.6% for easy and hard settings, respectively. This shows that, unlike reCAPTCHA, the difficulty has a direct impact on accuracy.

Distorted Text: We evaluated *agreement* among participants as a proxy for accuracy. Since each CAPTCHA was served to three distinct participants, we measured agreement between any two or more of them. We also observed that agreement increases dramatically (20% on average) if responses are treated as case insensitive, as shown in Table 3.

Table 2: Humans vs. bot solving time (seconds) and accuracy (percentage) for different CAPTCHA types.

	Human		**Bot**	
CAPTCHA Type	Time	Accuracy	Time	Accuracy
reCAPTCHA (click)	3.1-4.9	71-85%	1.4 [30]	100% [30]
Geetest	28-30	N/A	5.3 [36]	96% [36]
Arkose	18-42	N/A	N/A	N/A
Distorted Text	9-15.3	50-84%	<1 [38]	99.8% [16]
reCAPTCHA (image)	15-26	81%	17.5 [19]	85% [19]
hCAPTCHA	18-32	71-81%	14.9 [18]	98% [18]

Table 3: Agreement for distorted text CAPTCHAS.

	Average Agreement	Average Agreement (case insensitive)
Simple	84%	93%
Masked	50%	73%
Moving	62%	90%
Total	65%	85%

2.6 Measuring User Abandonment

Finally, we focus on **RQ5:** *Does experimental context influence abandonment?* We observed that the number of CAPTCHAS solved exceeded what would be expected, based on the number of participants who completed the study. We hypothesize that this was due to participants starting, yet not completing, the task. To assess this behavior, we conducted a second user study that collected timestamps between CAPTCHAS, regardless of whether the entire task was completed. We measured: (1) how many participants started the task; (2) how many abandoned the task when solving a CAPTCHA; and (3) if so, at which task and CAPTCHA.

This second study consisted of four groups, each with 100 unique participants. Two groups were presented with the direct, and another two with the contextualized, setting. We hypothesize that the amount of compensation might also impact abandonment. To this end, we doubled the compensation for one group in each setting. The studies were run sequentially to avoid prospective participants simply picking the higher-paying study.

Out of 574 participants who started the study, 174 abandoned prior to completion, corresponding to 30% abandonment rate. Several observations can be made: First, in the direct setting, 25% of the participants who ultimately abandoned the study did so before solving the first CAPTCHA; this rose to nearly 50% in the contextualized setting. Second, doubling the pay halved the abandonment rate for the contextualized setting (as expected), though increased it by 50% in the direct setting. Third, participants in the contextualized setting were 120% more likely to abandon than those in the direct setting. Fourth, in the contextualized setting, participants at the higher compensation level solved CAPTCHAS faster than those at the lower compensation level: 21.5% decrease in average solving time across all CAPTCHA types. Interestingly, in the direct setting, participants at the higher compensation level solved CAPTCHAS *slower* than those at the lower compensation level: 27.4% *increase* in average solving time across all CAPTCHA types. Finally, some CAPTCHA types (e.g., Geetest) exhibited higher rates of abandonment than others.

This initial investigation strongly motivates further exploration of CAPTCHA-induced abandonment. Although we studied the impact of compensation and experimental context, there may be other reasons for abandonment, such as: CAPTCHA type, CAPTCHA difficulty, and expected duration of study. Nevertheless, the trend of average users' unwillingness to solve a CAPTCHA during account creation (even for monetary compensation) is a relevant finding for websites that choose to protect account creation (and/or account access) using CAPTCHAS.

2.7 Security

Table 2 shows a comparison of our results to prior security analyses. Automated attacks on various CAPTCHA schemes have been quite successful [3, 4, 6, 7, 9, 11, 14, 16, 18, 19, 22, 23, 26, 28, 30–32, 36, 38]. accuracy of bots ranges from 85% to 100%, with the majority $> 96\%$.

This substantially exceeds observed human accuracy range: $50 - 85\%$. Furthermore bots' solving times are significantly lower in all cases, except reCAPTCHA (image), where human solving time (18 seconds) is similar to that of bots' – 17.5 seconds. However, in the contextualized setting, human solving time rises to 22 seconds, indicating that in this more natural setting, humans are slightly slower than bots.

3 Discussion

Based on results discussed above, we identify several key discussion points. First, the term CAPTCHA: *Completely Au-*

tomated Public Turing test to tell Computers and Humans Apart, has become a misnomer for two key reasons:

- Computer attacks and human inputs have become virtually indistinguishable.
- Popular current schemes using behavioral analysis use a black box (i.e., not public) approach to identifying suspected bot activity.

The heart of the Turing test is the imitation game, where a human judge would evaluate a natural language conversation between a human and a computer, without knowing which is which. If the judge can not distinguish among them, the computer is considered to have achieved a level of human intelligence. The main difference in CAPTCHAS is that the judge is a computer. This creates an entirely new problem: *Can a computer judge distinguish the difference between a live human and a replayed recording of a human when all related data is digital in form?*

In light of recent advances in bot technology and the consequent ease of attacks [3,4,6,7,9,11,14,16,18,19,22,23,26,28,30–32,36,38]: Image, Text, Game, Slider and Behavioral CAPTCHAS **can no longer distinguish** between live humans and bots. Therefore, to protect against bot activity at this day and age, there are only two options:

- Complicate CAPTCHAS even further, thus increasing user friction (i.e., annoyance and burden) without any concrete security guarantees.

  OR
- Deprecate all Image, Text, Game, Slider and, Behavioral CAPTCHAS, and invent new means of distinguishing among human and bot activity, especially, for high-value tasks, such as account creation and scarce resource consumption, e.g., tickets, reservations, etc.

Acknowledgments

We gratefully acknowledge our co-authors in the full version of this paper: Ai Enkoji, Andrew Paverd, and Yoshimichi Nakatsuka.We are also grateful to Usenix'Sec 2022 reviewers. This work was supported in part by: NSF Award 1840197, NSF Award 1956393, and NCAE-C CCR 2020 Award H98230-20-1-0345.

References

[1] Amazon Mechanical Turk. https://www.mturk.com/.

[2] CAPTCHA Usage Distribution on the Entire Internet. https://trends.builtwith.com/widgets/captcha/traffic/Entire-Internet.

[3] W. Aiken and H. Kim. POSTER: DeepCRACk: Using Deep Learning to Automatically CRack Audio CAPTCHAs. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ASIACCS '18, page 797–799, New York, NY, USA, 2018. ACM.

[4] F. H. Alqahtani and F. A. Alsulaiman. Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study. *Computers & Security*, 88:101635, 2020.

[5] J. P. Bigham and A. Cavender. Evaluating Existing Audio CAPTCHAs and an Interface Optimized for Non-Visual Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 1829–1838, New York, NY, USA, 2009. ACM.

[6] K. Bock, D. Patel, G. Hughey, and D. Levin. unCaptcha: A Low-Resource Defeat of reCaptcha's Audio Challenge. In *11th USENIX Workshop on Offensive Technologies (WOOT 17)*, Vancouver, BC, Aug. 2017. USENIX Association.

[7] E. Bursztein, R. Beauxis, H. Paskov, D. Perito, C. Fabry, and J. Mitchell. The Failure of Noise-Based Non-continuous Audio Captchas. In *2011 IEEE Symposium on Security and Privacy*, pages 19–31, 2011.

[8] E. Bursztein, S. Bethard, C. Fabry, J. C. Mitchell, and D. Jurafsky. How Good Are Humans at Solving CAPTCHAs? A Large Scale Evaluation. In *IEEE Symposium on Security and Privacy*, 2010.

[9] J. Chen, X. Luo, Y. Guo, Y. Zhang, and D. Gong. A Survey on Breaking Technique of Text-Based CAPTCHA. *Security and Communication Networks*, 12 2017.

[10] F. Consulting. State of online fraud and bot management. https://services.google.com/fh/files/misc/google_forrester_bot_management_tlp_post_production_final.pdf, 2021.

[11] M. Darnstädt, H. Meutzner, and D. Kolossa. Reducing the Cost of Breaking Audio CAPTCHAs by Active and Semi-supervised Learning. In *2014 13th International Conference on Machine Learning and Applications*, pages 67–73, 2014.

[12] Y. Feng, Q. Cao, H. Qi, and S. Ruoti. Sencaptcha: A mobile-first captcha using orientation sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(2), jun 2020.

[13] H. Gao, W. Wang, and Y. Fan. Divide and conquer: an efficient attack on Yahoo! CAPTCHA. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 9–16. IEEE, 2012.

[14] H. Gao, J. Yan, F. Cao, Z. Zhang, L. Lei, M. Tang, P. Zhang, X. Zhou, X. Wang, and J. Li. A Simple Generic Attack on Text Captchas. In *Network and Distributed System Security Symposium (NDSS)*, San Diego, California, United States, 2016.

[15] H. Gao, D. Yao, H. Liu, X. Liu, and L. Wang. A Novel Image Based CAPTCHA Using Jigsaw Puzzle. In *2010 13th IEEE International Conference on Computational Science and Engineering*, pages 351–356, 2010.

[16] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet. Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*, 2014.

[17] C. J. Hernandez-Castro and A. Ribagorda. Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study. *Computers & Security*, 29(1):141–157, 2010.

[18] M. I. Hossen and X. Hei. A Low-Cost Attack against the hCaptcha System. *CoRR*, abs/2104.04683, 2021.

[19] M. I. Hossen, Y. Tu, M. F. Rabby, M. N. Islam, H. Cao, and X. Hei. An Object Detection based Solver for Google's Image reCAPTCHA v2. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*, pages 269–284, San Sebastian, Oct. 2020. USENIX Association.

[20] Imperva. Imperva bad bot report. https://www.imperva.com/resources/resource-library/reports/bad-bot-report/, 2022.

[21] K. Krol, S. Parkin, and M. A. Sasse. Better the Devil You Know: A User Study of Two CAPTCHAs and a Possible Replacement Technology. In *2016 NDSS Workshop on Usable Security*, pages 1–10, 2016.

[22] C. Li, X. Chen, H. Wang, P. Wang, Y. Zhang, and W. Wang. End-to-end attack on text-based CAPTCHAs based on cycle-consistent generative adversarial network. *Neurocomputing*, 433:223–236, 2021.

[23] D. Lorenzi, J. Vaidya, E. Uzun, S. Sural, and V. Atluri. Attacking Image Based CAPTCHAs Using Image Recognition Techniques. In V. Venkatakrishnan and D. Goswami, editors, *Information Systems Security*, pages 327–342, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[24] M. Motoyama, K. Levchenko, C. Kanich, D. McCoy, G. M. Voelker, and S. Savage. Re: CAPTCHAs—Understanding CAPTCHA-Solving Services in an Economic Context. In *19th USENIX Security Symposium (USENIX Security 10)*, Washington, DC, aug 2010. USENIX Association.

[25] S. A. Ross, J. A. Halderman, and A. Finkelstein. Sketcha: A Captcha Based on Line Drawings of 3D Models. In *Proceedings of the 19th International Conference on World Wide Web*, page 821–830, New York, NY, USA, 2010. ACM.

[26] S. Sano, T. Otsuka, and H. G. Okuno. Solving Google's Continuous Audio CAPTCHA with HMM-Based Automatic Speech Recognition. In K. Sakiyama and M. Terada, editors, *Advances in Information and Computer Security*, pages 36–52, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[27] A. Searles, Y. Nakatsuka, E. Ozturk, A. Paverd, G. Tsudik, and A. Enkoji. An empirical study & evaluation of modern CAPTCHAs. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 3081–3097, Anaheim, CA, Aug. 2023. USENIX Association.

[28] H. Shekhar. Breaking Audio Captcha using Machine Learning/Deep Learning and Related Defense Mechanism. *San Jose State University Master's Projects*, 2019.

[29] V. Shet. Street View and reCAPTCHA technology just got smarter. https://security.googleblog.com/2014/04/street-view-and-recaptcha-technology.html, 2014.

[30] S. Sivakorn, I. Polakis, and A. D. Keromytis. I am Robot: (Deep) Learning to Break Semantic Image CAPTCHAs. In *2016 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 388–403, 2016.

[31] S. Solanki, G. Krishnan, V. Sampath, and J. Polakis. *In (Cyber)Space Bots Can Hear You Speak: Breaking Audio CAPTCHAs Using OTS Speech Recognition*, page 69–80. ACM, New York, NY, USA, 2017.

[32] M. Tang, H. Gao, Y. Zhang, Y. Liu, P. Zhang, and P. Wang. Research on Deep Learning Techniques in Breaking Text-Based Captchas and Designing Image-Based Captcha. *IEEE Transactions on Information Forensics and Security*, 13(10):2522–2537, 2018.

[33] N. Tanthavech and A. Nimkoompai. Captcha: Impact of website security on user experience. *ICIIT '19: Proceedings of the 2019 4th International Conference on Intelligent Information Technology*, pages 37–41, 02 2019.

[34] E. Uzun, S. Chung, I. Essa, and W. Lee. rtCaptcha: A Real-Time Captcha Based Liveness Detection System. In *Network and Distributed System Security Symposium (NDSS)*, San Diego, California, United States, 02 2018.

[35] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford. CAPTCHA: Using Hard AI Problems for Security. In E. Biham, editor, *Advances in Cryptology — EUROCRYPT 2003*, pages 294–311, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[36] H. Weng, B. Zhao, S. Ji, J. Chen, T. Wang, Q. He, and R. Beyah. Towards understanding the security of modern image captchas and underground captcha-solving services. *Big Data Mining and Analytics*, 2(2):118–144, 2019.

[37] J. Yan and A. S. El Ahmad. A Low-cost Attack on a Microsoft CAPTCHA. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 543–554, 2008.

[38] Y. Zi, H. Gao, Z. Cheng, and Y. Liu. An End-to-End Attack on Text CAPTCHAs. *IEEE Transactions on Information Forensics and Security*, 15:753–766, 2020.