BREAKING AND REDESIGINING AUDIO CAPTCHAs

Liam Crawford
CS:4640 || Omar Chowdhury
05/12/2017

Motivation

One of the most popular ways for websites to authenticate humans is by presenting a CAPTCHA, or Completely Automated Turing test for telling Computers and Humans Apart. Many people think of CAPTCHA's as being visual puzzles, but most websites offer an audio alternative for visually impaired users. Unfortunately, most audio CAPTCHA's used today are either insecure (too easy for computers), difficult to pass (too hard for humans), or sometimes both (Bigham & Cavendar, 2009)!

This report was written in pursuit of a better design for audio CAPTCHA's, one is that is not only secure, but also accessible. It is written as a condensed/informal literature review followed by a proposed design framework.

History of Breaking Audio CAPTCHAs

In 2009, researchers at Carnegie Mellon University came up with the first comprehensive paper outlining the insecurity of audio CAPTCHA's. Their approach was to take various language-based CAPTCHAs from around the web and train machine learning algorithms to be able to segment and recognize the data. They were able to achieve up to a 71% success rate, which greatly varied based on factors such as the number of different voices, the size of the vocabulary, etc. Their results suggested that the algorithms had the most trouble when there were different speakers for each of the words, and a constant, varied background noise derived from human speech (Tam, et al., 2009).

In 2011, Bursztein et al. developed Decaptcha, a "two-phase audio CAPTCHA solver" designed to break into all non-continuous audio CAPTCHA that used numbers. Like the CMU researchers, they worked with machine learning algorithms to achieve audio segmentation and recognition. At the time of their writing the paper, they managed to defeat many popular websites' – including Microsoft and Yahoo – audio CAPTCHA's with a up to a 89% success rate! Their findings reaffirmed the need for background noise similar to human speech, as well as discouraged simple distortion of sounds (Bursztein, et al., 2011).

In 2014, Meutzner et al. developed an approach based on then "state of the art" Automatic Speech Recognition which broke the 2014 version of audio reCAPTCHA with a 62.8% success rate. In addition, they highlighted that the human success rate of the same CAPTCHA was 24.4%. From these findings, he suggested that the security benchmark for future CAPTCHAs be based on Automatic Speech Recognition, and not the segment and recognize scheme used before.

Liam Crawford
CS:4640 || Omar Chowdhury
05/12/2017

Recent Approaches

Interest in increasing the security of audio CAPTCHA's has gained a lot of traction in recent years, as they seem to be the weak link in human authentication. Here are some proposed/implemented improvements:

Decrease the number of CAPTCHAs a user sees

Google recently pushed out their new NoCAPTCHA reCAPTCHA, which allows users to click a simple "I'm not a robot" button to verify them. It supposedly does a risk analysis based on a number of factors, and you are only presented with a more traditional CAPTCHA if your activity is suspicious. This new system works with screen readers (Accessibility-reCAPTCHA), but still falls back on a relatively weak audio CAPTCHA in the end.

Utilize new qualities of sound

In 2013, researchers at Google filed a patent for "Systems and Methods for Three-Dimensional Audio CAPTCHA," which were aimed at utilizing humans' ability to detect auditory distance (Google, Inc. 2016). The design approach would involve trying to confuse computers by providing both decoy and target sounds, which humans would be better at distinguishing between. Not much else in terms of literature exists on this.

Creating language agnostic audio CAPTCHAs

In 2016, Muetzner and Kolossa worked on a CAPTCHA scheme that utilized types of sounds rather than language itself. This removed the need for semantic background noise, and in its place, they put environmental background noise. After listening to the audio, users were asked to identify the sounds that they perceived. By their own metrics, they were able to achieve a human success rate of 72% and a computer success rate of 21% (using a popular strategy that, on average, had a 50% success rate) (Meutzner, Kolossa, 2016).

New directions in language based audio CAPTCHA design

Also in 2016, Muetzner et al. published a paper outlining a design approach based on "auditory perception and language understanding" as an extension of Muetzner's 2014 work. He implemented a design based on human's ability (and computers coinciding inability) to separate out simultaneously occurring and reverberating sounds. This approach – which relies heavily on speech – reduced the ASR's accuracy to 5.33% and increased humans to 56.38% (Meutzner, et al., 2016).

BREAKING AND REDESIGINING AUDIO CAPTCHAs

Liam Crawford
CS:4640 || Omar Chowdhury
05/12/2017

Proposal

I propose a new scheme for audio CAPTCHAs that combines elements of recent approaches as well as "no-knowledge" based authentication techniques.

Scheme

- When appropriate, a user will be presented with an audio CAPTCHA to solve
- First, you are given a spoken-language based instruction
    - This instruction details what you will be listening for in the second half of the audio clip
        - The audio equivalent of "Click all pictures that arew part of stop sign" in today's Audio reCAPTCHA
        - The user will NOT be expected to type these words into
    - Instructions will vary
        - Which sounds (1-5) were car horns?
        - Count the number of baby sounds you hear?
        - Which sounds (1-5) were not dogs?
    - Security elements
        - Use semantic background noise to distort the audio to computers
        - Superimpose words ever so slightly on top of one another
        - Rely on sentence structure to convey meaning
        - Multiple voices
- Second – challenge sounds and response
    - The user will listen to the sounds being presented to them
    - Follow the instruction given to them earlier to determine input
    - Security elements
        - Environmental background noise
        - Some, but little superimposition of different sounds

Challenges

- There is a language element, thus not all users can use it.
    - Solution: begin by just using English, crowdsourcing the audio being used
    - Over time, use services like Amazon's Mechanical Turk to create a translation base for the first half of instruction
- Picking sounds that are universally recognizable and categorizable
- Determining the appropriate AI technologies to combine to test the new scheme

Paper	Approach	Design Principles		
		Language Agnostic	ASR Protection	Varied Prompts
Google Patent	3D Audio Vectors	Theoretic (No product)		
Meutzner & Kolossa	Noises	Yes	N/A	No
Meutzner et al.	Audio Perception…	No	Yes	No
---	No-Knowledge Hybrid	Partially	Yes	Yes

Bibliography

Accessibility - reCAPTCHA. (2017). Retrieved May 03, 2017, from https://support.google.com /recaptcha/answer/6175971?hl=en

Bigham, J. P., & Cavender, A. C. (2009, April). Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1829-1838). ACM.

Bursztein, E., Beauxis, R., Paskov, H., Perito, D., Fabry, C., & Mitchell, J. (2011, May). The failure of noise-based non-continuous audio captchas. In *Security and Privacy (SP), 2011 IEEE Symposium on* (pp. 19-31). IEEE.

Google, Inc. (2016). U.S. Patent No. US 9263055 B2. Washington, DC: U.S. Patent and Trademark Office.

Meutzner, H., Gupta, S., Nguyen, V. H., Holz, T., & Kolossa, D. (2016). Toward Improved Audio CAPTCHAs Based on Auditory Perception and Language Understanding. *ACM Transactions on Privacy and Security (TOPS)*, *19*(4), 10.

Meutzner, H., & Kolossa, D. (2016, August). A non-speech audio CAPTCHA based on acoustic event detection and classification. In *Signal Processing Conference (EUSIPCO), 2016 24th European* (pp. 2250-2254). IEEE.

Meutzner, H., Nguyen, V. H., Holz, T., & Kolossa, D. (2014, December). Using automatic speech recognition for attacking acoustic CAPTCHAs: The trade-off between usability and security. In *Proceedings of the 30th Annual Computer Security Applications Conference* (pp. 276-285). ACM.

Tam, J., Simsa, J., Hyde, S., & Ahn, L. V. (2009). Breaking audio captchas. In *Advances in Neural Information Processing Systems* (pp. 1625-1632)