

4005-893-01
MS Thesis Seminar
Thesis Pre-Proposal

Kurt Alfred Kluever (kurt@klover.com)
Committee Chair: Dr. Richard Zanibbi

January 11, 2008

1 Problem, Hypothesis, or Question

HIPs, or Human Interactive Proofs, are a method used to differentiate between humans and machines on the internet, and are typically implemented as distorted text which the user must correctly transcribe. HIPs should be easy for a machine to automatically generate, easy for a human to solve, and difficult, or impossible, for a machine to solve. Current implementations include recognizing a string of distorted characters [7], choosing the correct annotation for an image [6] [8], or transcribing an audio clip with noise [9]. The use of video in HIP challenges has not yet been explored. The major questions to be answered are: can a video based HIP be an effective and user friendly method of exploiting the gap between human intelligence and machine learning, and what types of attacks might be used against such a system?

2 Importance of Research

Differentiating between a user and machine over the internet has significant importance in the fields of internet security, artificial intelligence, and machine learning. Currently, HIPs prevent robots from signing up for free online services (such as email accounts), abusing online polls, providing biased feedback, and spamming innocent users.

3 Theory Base for Research

In 1950, Alan Turing provided an operational definition of intelligence using the imitation game [13]. The proposed Turing test is administered by a human and taken by a machine. If the machine can trick the human into believing that the machine is also a human, the machine is said to have passed the Turing test. In contrast, HIPs are administered by a machine but taken by a human. The burden is on the human to convince the machine that he/she is in fact human. In this way, HIPs are analogous to *reverse* Turing tests.

In 1996 unpublished draft, Moni Naor theorized nine possible sources for HIPs [11]. They included text based, image based, and speech based challenges. These three media formats (text, image, audio) have served as popular choices for HIP implementations. The current trend in online media is towards streaming video. The next logical media format choice for HIPs is video.

4 Significant Prior Research

Prior research has shown that requiring a user to recognizing characters in a string of distorted text is not the only reliable method. Image based HIPs, such as those developed by Monica Chew and J. D. Tygar [6], have proved both secure and easy to use. Due to the amount of challenge data involved, video HIPs have the potential to be more secure than image based HIPs.

There has been extensive research done in the area of HIPs at the Palo Alto Research Center [7] [5], Bell Laboratories [9], Microsoft Research [8] [4], Carnegie Mellon University [14], Lehigh University [7] [5] [2], University of California at Berkeley [6], and the University of Buffalo, CEDAR [12] . There have been two international workshops on HIPs [3] [1]. Independent user studies have been conducted to determine the “user friendliness” of the different HIPs. Most text based HIPs [14] [5] are no longer secure against attacks such as recognition via neural networks [4] and shape matching [10].

5 Possible Research Approach or Methodology

Once an implementation of the video HIP has been developed, a user study with at least 30 participants will be conducted to evaluate the user friendliness of the HIP. The study will evaluate how long it takes a user to solve the HIP and the user’s success rate, followed by a subjective user ranking of how easy the HIP was to solve. Attempts will be made to attack the implementation using methods such as database replication, brute force, and pattern recognition techniques.

6 Potential Outcomes and Importance of Each

A new method for differentiating between humans and machines will be developed. The vulnerability of the system against one or more attacks will be explored and documented. The results of the user study will determine whether users are satisfied with the proposed approach.

Regardless of the outcomes of the user study and the attempts at attacking the video HIP, a knowledge contribution will be made:

	Vulnerable to Attack	Resilient Against Attack
Easy for Humans to Solve	Minor contribution	Significant contribution
Difficult for Humans to Solve	Minor contribution	Minor contribution

If the video HIP is resilient against attack and easy for humans to solve, the contribution will be a secure and user friendly HIP. All other outcomes will result in a minor contribution, which demonstrate the limitations of the video HIP.

References

- [1] Henry S. Baird and Daniel P. Lopresti, editors. *Human Interactive Proofs, Second International Workshop*, volume 3517 of *Lecture Notes in Computer Science*, Bethlehem, PA, USA, May 19–20 2005. Springer.
- [2] Henry S. Baird, Michael A. Moll, and Sui-Yu Wang. Scattertype: A legible but hard-to-segment captcha. In *ICDAR '05: Proceedings of the Eighth International Conference on Document Analysis and Recognition*, pages 935–939, Washington, DC, USA, 2005. IEEE Computer Society.
- [3] Manuel Blum and Henry S. Baird, editors. *Human Interactive Proofs, First International Workshop*, Palo Alto, CA, January 2002. Xerox Palo Alto Research Center.
- [4] Kumar Chellapilla and Patrice Y. Simard. Using machine learning to break visual human interaction proofs (hips). In *NIPS*, 2004.
- [5] M. Chew and H. Baird. Baffletext: A human interactive proof. In *Proceedings of SPIE-IS&T Electronic Imaging, Document Recognition and Retrieval X*, pages 305–316, January 2003.
- [6] Monica Chew and J. D. Tygar. Image recognition captchas. Technical Report UCB/CSD-04-1333, EECS Department, University of California, Berkeley, Aug 2004.
- [7] Allison L. Coates, Henry S. Baird, and Richard J. Fateman. Pessimial print: A reverse turing test. In *Proceedings, IAPR 6th Int'l Conf. on Document Analysis and Recognition*, pages 1154–1158, Seattle, WA, Sept 2001. IEEE Computer Society.
- [8] John Douceur, Jeremy Elson, Jon Howell, and Jared Saul. Asirra: a captcha that exploits interest-aligned manual image categorization. In *CCS '07: Proceedings of the 14th ACM conference on Computer and communications security*, pages 366–374, New York, NY, USA, 2007. ACM.
- [9] Daniel P. Lopresti, C. Shih, and G. Kochanski. Human interactive proofs for spoken language interfaces. In *Proceedings of the 1st Workshop on Human Interactive Proofs*, pages 30–34, Palo Alto, CA, January 2002.
- [10] G. Mori and J. Malik. Recognizing objects in adversarial clutter: Breaking a visual captcha. In *CVPR*, volume 1, pages 134–141, 2003.
- [11] Moni Naor. Verification of a human in the loop or identification via the turing test. sdf, Sept 1996.
- [12] A. Rusu and V. Govindaraju. Handwritten captcha: using the difference in the abilities of humans and machines in reading handwritten words. *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, pages 226–231, October 2004.
- [13] A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, October 1950.
- [14] Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford. Captcha: Using hard ai problems for security. In *EUROCRYPT*, pages 294–311, 2003.