Proceedings of the International Multiconference on Computer Science and Information Technology pp. 325–329
ISBN 978-83-60810-14-9
ISSN 1896-7094

SemCAPTCHA—user-friendly alternative for OCR-based CAPTCHA systems

Paweł Łupkowski, Mariusz Urbański
Chair of Logic and Cognitive Science
Institute of Psychology
Adam Mickiewicz University
Szamarzewskiego 89
60-568 Poznań, Poland
Email: {Pawel.Lupkowski, Mariusz.Urbanski}@amu.edu.pl

***Abstract*—In this paper we present a new CAPTCHA system (*Completely Automated Turing Test To Tell Computers and Humans Apart*). This proposal, SemCAPTCHA, is motivated by an increasing number of broken OCR-based CAPTCHA systems and it is based not only on text recognition but also on text understanding.**

We describe SemCAPTCHA from both user's perspective and system's perspective and compare it to some currently popular CAPTCHAs. We also briefly describe an experiment carried out to test our CAPTCHA on human users.

I. Introduction

IN MANY domains there is an increasing demand for simple and efficient way to differentiate real human users from malicious programs (bots). Just a few examples of such domains are: services offering free e-mail accounts, community portals, online polls etc.

One of the most popular ways to tell human users and bot users apart are so called CAPTCHA systems (this acronym stands for *Completely Automated Turing Test To Tell Computers and Humans Apart*).

Design of an effective CAPTCHA system is a difficult task, since two distant needs must be satisfied: it has to be really hard for a machine and at the same moment it has to be simple and friendly for a human. User friendliness is important as CAPTCHAs cannot engage to much of a user cognitive resources and cannot consume to much of her time. Registering a free e-mail account is a good example here. There are many alternative providers of such accounts on the market, so if you want a potential user to solve CAPTCHA on your site, it has to be as unproblematic for her as possible (and you want her to solve it in order to prove that she is a human, not a bot who will send tons of spam from your servers). If a potential user gets irritated, she will go away and pick another provider. To make things more difficult, there's also a third factor: a CAPTCHA has to be open, that is, the algorithms used by a system must be public. The idea is that CAPTCHA efectiveness should be based on hardness of an underlying AI problem and not on a secret cryptographic mechanism or other copyrighted mystery. Finally, test instances of a CAPTCHA should be generated automatically.

This research was partially supported by AMU Faculty of Social Sciences grant No. WSO/133/2006.

Internet users encounter CAPTCHAs very often. Most of them are visual CAPTCHAs where the task consists in recognition of a word or string of symbols (letters, numbers) from a distorted picture. To solve such CAPTCHA a user has to write down words or symbols from the picture. Such systems work e.g. on Yahoo, Gmail, Wirtualna Polska, Gazeta.pl and many other sites. Exemplary CAPTCHAs are presented in table V.

Currently it is an important issue that AI problem underlying such CAPTCHAs is challenged by constantly developing Optical Character Recognition (OCR) systems with increasing success rate. Mori and Malik [8] describe an attack on a visual CAPTCHA EZ-Gimpy used by Yahoo!, which enjoyed a success rate of 92%. In more difficult case of Gimpy they passed the test 33% of the time. As the authors claim: "with our 33% accuracy, this CAPTCHA would be ineffective in applications such as screening out "bots" since a computer could flood the application with thousands of requests." [8, p. 7]. After all, year 2008 seems to be a really bad year for visual CAPTCHAs: Yahoo! CAPTCHA was hacked again (http://osnews.pl, 21.01.2008), as well as Gmail one (http://osnews.pl, 27.02.2008) and MS Windows Live Hotmail (http://arstechnica.com, 15.04.2008). Many visual CAPTCHAs are broken 'out of the box' by PWNtcha system (see http://libcaca.zoy.org/wiki/PWNtcha—examples of 12 broken CAPTCHAs where success rate is from 49% to 100%).

As a consequence, there is a great need for more secure alternative CAPTCHAs, which are based not only on OCR problem. There are some proposals, like question-based CAPTCHA [7], ARTiFACIAL [12], PIX [1], sound oriented CAPTCHAs [3] etc. In our opinion the current situation offers a great motivation to look for an inspiration for CAPTCHA systems not only in simple sensory processing but in higher levels of human data processing.

II. SemCAPTCHA system

Our proposal is to base a CAPTCHA system on a combination of an OCR problem and some linguistic task, and to apply the effect of positive semantic priming to strengthen human odds against computers. Everything what is needed to break a simple visual CAPTCHA is an good OCR program. Breaking our system—SemCAPTCHA, where "Sem" stands for

Fig. 1. Sample instance of SemCAPTCHA test

"semantic"—is not that straightforward for a machine and still for a human user it remains quite simple. The process of solving SemCAPTCHA task consists of three steps, based on different cognitive activities (which must be completed within a certain amount of time): reading a text—understanding it—applying user's knowledge about the world.

A. SemCAPTCHA—a user's perspective

A SemCAPTCHA test instance consists of a distorted picture, on which three words are presented. All of them are the names of animals. One animal differs from the rest (e.g. it is a mammal among reptiles). The task is to recognize its name and point it by a mouse click. It has to be stressed that the words do not differ substantially as for their graphical properties (like, e.g. length). The difference is of semantic character: one word differs from the other two in its meaning.

An example of such test instance is given in figure 1: a user is presented with the words "kaczka" (a duck), "kukułka" (a cuckoo), "krowa" (a cow; SemCAPTCHA is designed in Polish). The proper answer is "krowa" and the semantic difference is based on taxonomy: ducks and cuckoos are birds while cows are mammals.

To solve this task a user first has to recognize the words from a distorted picture, then identify their meaning and finally find an underlying pattern and the word which does not fit it. The choice of words makes it easy even for not very fluent language users.

In order to make SemCAPTCHA even easier for humans we decided to employ the positive semantic priming effect. Each test instance is preceded by a prime (exposition time is ca. 70 ms). The prime is a word semantically connected with the task solution; in case of the above example it might be a word "mleko" (milk). It is known from cognitive psychology that this setting enables human to recognize a target word much faster than a stand alone target word. Consequently, human user will solve SemCAPTCHA test instances easier and faster (cf. next section, [5] and [6]).

B. SemCAPTCHA—a system's perspective

SemCAPTCHA is not implemented yet, but the procedures needed for the system are already developed.

SemCAPTCHA works on a word base consisting of 500 animals' names. Names are grouped in categories, e.g. mammals, birds, reptiles. Each word has its own semantic field (stored as semantic network). Semantic field contains words semantically connected with a given animal name. Each connection of words is marked by a label containing information about relation type and relation strength, expressed by a numerical value 1–100 (as sources for semantic fields generation we used IPI PAN—corpus of Polish developed by the Polish Academy of Sciences—and Google). Such architecture enables efficient and automatic generation of test instances.

To generate a test instance system chooses randomly two categories from the word base. Then it picks (also randomly) one word (w_1) from the first category and two words from the second one (w_2, w_3). Then the system picks a prime for w_1, using semantic network stored for w_1. The system randomly chooses possible relation strength with w_1 (e.g. 50–70) and a word that obeys this restriction. Then a distorted picture is generated using w_1, w_2, w_3 and it is preceded by a prime and a mask.

After a test is generated and displayed SemCAPTCHA starts measuring the time. A solution time (an interval between exposition of a picture and a mouse click) is compared with a standard solution time for SemCAPTCHA. On this basis SemCAPTCHA estimates the probability that a user is a human and decides if a test has been passed or not.

Our experiment shows, that for humans solution time varies from 1,2 to 5,5 seconds (cf. next section, [5] and [6]; more thorough research could help verify these limits). This is one of the most characteristic properties of SemCAPTCHA: it not only generates and scores test instances but it also constantly checks solution time, and its verdict depends not only on correctness of a solution but also on time needed for it. In this point SemCAPTCHA differs substantially from widely used OCR-based CAPTCHA systems.

III. SEMCAPTCHA EXPERIMENT

To verify the idea of using linguistic competence and positive semantic priming in SemCAPTCHA system we have carried out an experiment (details on the instruments used and methods of statistical analysis can be found in [5] and are available from the authors).

Our research questions for these issues were:

1) Is the effect of positive semantic priming statistically significant for solution time of SemCAPTCHA test instances?
2) Is the effect of positive semantic priming statistically significant for solution accuracy of SemCAPTCHA test instances?

The experiment consisted of one training task and 10 test instances. A single instance consisted of a picture with 3 Polish words (names of animals). One word was different from the other two in that it was a name of an animal of a different class. For each picture we used one of standard CAPTCHA's method of distortion. We prepared two sets of tasks, A and B, consisting of the same test instances. In an experimental set A each test instance was preceded by a prime, semantically connected with the word which formed the correct solution of a task. A prime was followed by a mask. In a control set B there was no prime. Detailed characteristics of test instances are given in table I.

The sample consisted of 64 students at the Adam Mickiewicz University (19 males, 43 females, 2 no data), who

TABLE I
TASKS CHARACTERISTICS

Task	Prime (ms)	Mask (ms)	Text dist.	Bg. dist.
T1	70	50	G-blur	HSV
T2	60	50	G-blur	RGB
T3	80	50	G-blur	fog
T4	90	50	dispersion	HSV
T5	100	60	dispersion	RGB
T6	60	30	dispersion	fog
T7	70	50	Whirl&Pinch	HSV
T8	70	50	Whirl&Pinch	fog
T9	70	50	Whirl&Pinch	RGB
T10	70	50	newspaper printout	HSV

TABLE II
AVERAGE TIME, ACCURACY AND SUBJECTIVE DIFFICULTY OF TASK SOLUTIONS

Task	Group	N	Average time (sec.)	Accuracy	Difficulty (Average)
T1	A	31	5,5408	17	6,35
	B	33	5,9048	16	6,55
T2	A	31	2,3467	30	2,61
	B	33	2,7859	32	3,56
T3	A	31	1,8594	27	3,26
	B	33	2,7749	31	3,48
T4	A	31	2,7456	21	5,61
	B	33	4,7085	25	6,50
T5	A	31	1,2047	31	3,00
	B	33	3,3308	31	3,63
T6	A	31	1,8863	30	3,03
	B	33	2,8534	32	3,47
T7	A	31	2,5314	21	4,50
	B	33	3,5239	22	5,28
T8	A	31	1,7810	28	3,67
	B	33	3,2051	31	5,25
T9	A	31	1,4193	30	2,67
	B	33	2,6340	32	2,97
T10	A	31	1,5180	23	3,07
	B	33	2,6648	27	3,47

volunteered to participate in the experiment. They all belonged to the largest group of Internet users, i.e. people aged between 21 and 25. Participants were randomly distributed over groups A (experimental group) and B (control group).

The subjects were asked to choose on each picture from three names of animals the name of an animal which differs from other two and point it by a mouse click. Solution time was measured as an interval between exposition of a picture and a click. Time and correctness of a solution were written down automatically by a server. After completion of all ten test instances the subjects were asked to fill a short questionnaire concerning subjective difficulty of each task (a complete set of pictures was presented on the monitor at this stage) and their willingness to solve such tasks while surfing the Internet.

The results enable to formulate a positive answer to our first question and a negative answer to the second one (cf. table II). First and foremost, we observed the effect of positive semantic priming in solving test instances of SemCAPTCHA: there was statistically significant difference in time of solving test instances between the experimental group (A) and the control group (B). Participants from group A solved test instances faster than participants from group B and thus it is possible to differentiate betwen experimental and control group on the basis of the average time of solving test instances. This effect was present in case of eight out of ten test instances (T3 – T10). Lack of positive semantic priming effect in case of the first and second instance can be explained by the need for some practice in solving such tasks.

On the other hand, improvement in time of solving test instances does not affect in a statistically significant way the accuracy of solutions. Participants from the experimental group solved test instances just faster, not more accurate than participants from the control group.

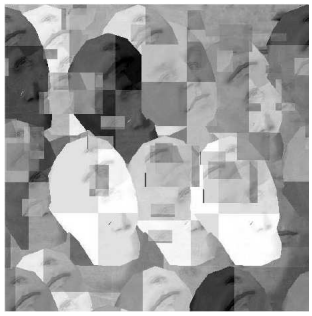


Fig. 2. Example of ARTiFACIAL test [12]

IV. SEMCAPTCHA AND OTHER PROPOSALS

As we noticed above, user friendliness is one of the crucial issues for an effective CAPTCHA systems: for humans they should be as easy as possible. Thus it is interesting to compare our system with other CAPTCHAs on the basis of declared subjective difficulty of test instances and declared willingness to use them in practice. For comparison we have chosen CAPTCHAs for which such data were available.

We mentioned already that in our experiment we asked participants to declare subjective difficulty of test instances (on the scale 1–10, where 1 means the simplest). For each test instance subjective difficulty declared by participants from experimental group was slightly lower than the one declared by participants from the control group (however, only in one instance this difference was statistically significant). We observed high correlation between average declared difficulty and average solution time ($r^2 = 0.71$ for group A). As a consequence, time of solution seems to be a good estimator of task complexity. This observation gives some base for comparing SemCAPTCHA with other CAPTCHA systems on the objective basis of their solution times.

One of the alternatives for OCR-based CAPTCHA is ARTiFACIAL. It is based on ability to recognize faces. Motivation for this system is similar to ours—make use of higher levels of human data processing. ARTiFACIAL test consists of one picture containing background (with randomly choosen facial features) and a face (exemplary tets instance is presented in figure 2). The task is to find and point six points on such picture (left and right corner of: left eye, right eye and mouth). As could be expected, ARTIFiCIAL is really

TABLE III
AVERAGE TIME (IN SEC.) OF ARTIFACIAL TEST SOLUTION

task	1	2	3	4	5	6	7	8	9	10
time	22	15	16	13	12	11	12	12	11	12

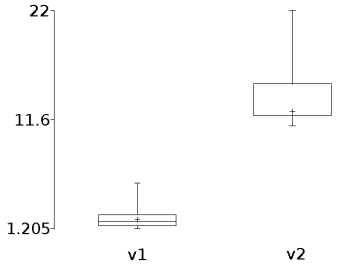


Fig. 3. Average time of solution for SemCAPTCHA (v1) and ARTiFACIAL (v2)

hard for machines, but is it simple enough for human users? ARTiFACIAL authors carried out an experiment on this issue. It consisted of 10 ARTiFACIAL test instances. The sample consisted of 34 subjects (accountants, administrative staff, architects, executives, receptionists, researchers, software developers, support engineers, and patent attorneys). Average solution times are presented in table III (cf. [12, p. 500]).

The mechanics of ARTiFACIAL and SemCAPTCHA are quite similar and it can be claimed that ARTiFACIAL's underlying problem is not more difficult than SemCAPTCHA's one. Thus, if we use the solution time as an estimator of task complexity for human users we may say that ARTiFACIAL is a really complex CAPTCHA system. Average solution time for all tasks is 14 seconds. SemCAPTCHA seems to be much easier, since the average solution time is 2.3 seconds (cf. figure 3).

On the basis of declared willingness to use them in practice we can compare SemCAPTCHA to a simple visual CAPTCHA system—BaffleText. In [4, p. 7] there are given results of a short questionnaire which was ment to investigate BaffleText users feelings about this system. It has been filled by 18 out of 33 subjects (Palo Alto Research Center employees):

1) 16,7 % reported they would be willing to solve a BaffleText every time they sent email;
2) 38,9 % reported they would be willing, if it reduced spam tenfold;
3) 94,4 % reported they would be willing, if it meant those sites had more trustworthy recommendations data;
4) 100 % reported they would be willing to solve one every time they registered for an e-mail account.

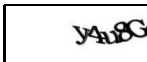
In our experiment we asked subjects to answer the same questions (61 out of 64 did this):

1) 15,6 % reported they would be willing to solve a SemCAPTCHA every time they sent email;
2) 43,8 % reported they would be willing, if it reduced spam tenfold;
3) 65,6 % reported they would be willing, if it meant those sites had more trustworthy recommendations data;
4) 34,4 % reported they would be willing to solve one every time they registered for an e-mail account.

TABLE IV
OCR TESTS FOR SEMCAPTCHA

GOCR		Asprise OCR		ABBYY FR	
words	letters	words	letters	words	letters
0 %	4,11 %	0 %	6,16 %	13,33 %	13,01 %

TABLE V
EXEMPLARY TASKS OF CAPTCHAS USED BY YAHOO!, WP.PL AND GAZETA.PL

		
Yahoo!	wp.pl	gazeta.pl

We think that this results are very promising for SemCAPTCHA. One possible explanation of low results for third and fourth question is that our subjects were students. They might be not so keen in web security issues as PARC employees.

We have also performed some OCR tests, to see how hard are SemCAPTCHA tests for OCR programs. SemCAPTCHA uses slightly distorted pictures, so we intended to compare them with OCR-based CAPTCHAs currently used on popular portals. We tested our experimental test instances against three OCR programs: GOCR, Asprise OCR and ABBYY Fine Reader 9.0 PE. Results (percentage of correctly recognized words and symbols) are presented in table 4.

For comparison we also performed OCR tests (against the same three programs) for other popular visual CAPTCHAs: the ones used by Yahoo!, wp.pl and gazeta.pl (10 instances for each). These CAPTCHAs do not use regular words, but only strings of symbols (letters and numbers). Exemplary tasks are presented in table V.

For CAPTCHA used by Yahoo! (considered as hard) GOCR recognised 2.82% signs; Asprise OCR 1.41% and ABBYY FR 19.72%. As for wp.pl results were following: GOCR 52.94%, Asprise OCR 16.67%, ABBYY FR 5%. And for gazeta.pl: GOCR 45%, Asprise OCR 0%, ABBYY FR 47.06%. All results are presented in figure 4.

All tested CAPTCHAs are based on an OCR problem. SemCAPTCHA results are comparable with the others (and it

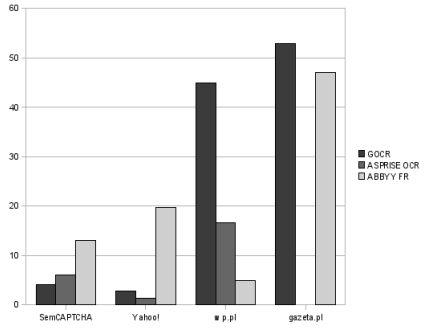


Fig. 4. OCR tests results (in % of recognised symbols)

should be stressed that recognising words in SemCAPTCHA task is only a first step towards solution; cf. section II). Thus we may conclude, that OCR-hardness of SemCAPTCHA is set high enough, i.e. it is at least as hard for machines as CAPTCHAs currently used and still quite easy for human users.

V. Conclusions

SemCAPTCHA, based on a combination of an OCR problem, some linguistic task and positive semantic priming, seems to be a promising system for telling humans and computers apart. On the one hand, engagement of higher level human data processing makes it harder for machines than currently used visual CAPTCHAs. On the other hand, it is not as complex for human users as other alternatives to current systems. SemCAPTCHA has a simple and open algorithm, is easy for humans and can be designed for any language.

References

[1] Ahn L., Blum M., Hopper N. J., Langford J. CAPTCHA: Using Hard AI Problems For Security. Retrieved October 11, 2007 from http://www.captcha.net.

[2] Ahn L., Blum M., Langford J. Telling Humans and Computers Apart Automatically. How Lazy Cryptographers do AI. Retrieved October 11, 2007 from http://www.captcha.net.

[3] Chan N. Sound oriented CAPTCHA. Retrieved October 11, 2007 from http://www.captcha.net.

[4] Chew M, Baird H. S. (2003). BaffleText: a Human Interactive Proof. Proceedings of the SPIE/IS&T Document Recognition and Retrieval Conf. X. Santa Clara, CA.

[5] Łupkowski P., Urbański M. (2006). Positive semantic priming as an optimalization tool for automated user authorization systems. Research report, Institute of Psychology, Adam Mickiewicz University (in Polish).

[6] Łupkowski P., Urbański M. (2008). SemCAPTCHA. Telling Computers and Humans Apart by Means of Linguistic Competence and Positive Semantic Priming, In L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, J. Zurada (Eds.), Computational Intelligence: Methods and Applications (pp. 525–531). Academic Publishing House EXIT.

[7] Shirali-Shahreza M., Shirali-Shahreza J. (2007). Question-Based CAPTCHA, Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)—Volume 04 (pp. 54–58). IEEE Computer Society, Washington DC.

[8] Mori G., Malik, J. (2003). Recognizing objects in adversarial clutter: breaking a visual CAPTCHA. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2003. Retrieved October 11, 2007 from http://www.cs.sfu.ca/$\sim$mori/research

[9] Naor M. (1996). Verification of a human in the loop or Identification via The Turing Test. http://www.wisdom.weizmann.ac.il/$\sim$naor/PAPERS/human.ps

[10] Neely J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), Basic processes in reading (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates.

[11] Plaut D. C. (1995). Semantic and Associative Priming in a Distributed Attractor Network. In Proceedings of the 17th Annual Conference of the Cognitive Science Society (pp. 37–42). Hillsdale, NJ: Lawrence Erlbaum Associates.

[12] Rui Y., Liu Z. (2004). ARTiFACIAL: Automated Reverse Turing test using FACIAL features. *Multimedia System* **9**: 493–502.