ScatterType: A Legible but Hard-to-Segment CAPTCHA

Henry S. Baird and Michael A. Moll and Sui-Yu Wang

Computer Science & Engineering Dept
Lehigh University
19 Memorial Dr West
Bethlehem, PA 18017 USA

E-mail: `baird@cse.lehigh.edu`, {`mam7`|`syw2`}`@lehigh.edu`
URL: `www.cse.lehigh.edu/~baird`

Abstract

The ScatterType CAPTCHA, designed to resist character–segmentation attacks and shown to be highly legible to human readers, is analyzed for vulnerabilities and is offered for experiments in automatic attack. As introduced in [BR05], 'ScatterType' challenges are images of machine-print text whose characters are cut into pieces which then drift apart, in an attempt to frustrate segment-then-recognize computer vision attacks. Analysis of experimental human legibility data has shown that better than 95% correct legibility can be achieved through judicious choice of the pseudorandom generating parameters [BMW05]. That analysis is summarized and discussed here as motivation for a discussion of potential vulnerabilities. An invitation to attack ScatterType is offered.

Keywords: *CAPTCHAs, human interactive proofs, document image analysis, abuse of web sites and services, human/machine discrimination, Turing tests, legibility of text, segmentation, fragmentation, Gestalt perception, automatic attacks on CAPTCHAs*

1 Introduction

In 1997 Andrei Broder and his colleagues at the DEC Systems Research Center developed a scheme to block the abusive automatic submission of URLs to the AltaVista web-site [Bro01,LABB01]. Their approach was to challenge a potential user to read an image of printed text formed specially so that machine vision (OCR) systems could not read it but humans still could. Since that time, also inspired (as Broder's team was) by Alan Turing's 1950 proposal of methods for validating claims of artificial intelligence [Tur50], many such CAPTCHAs—Completely Automated Public Turing tests to tell Computers and Humans Apart—have been developed, including CMU's EZ-Gimpy [BAL00, HB01], PARC's PessimalPrint [CBF01] and BaffleText [CB03], Paypal's CAPTCHA ((`www.paypal.com`)), Microsoft's CAPTCHA [SSB03,CLSC05], and Lehigh's ScatterType [BR05]. As reported in [BR05,CLSC05] attacks on some of these CAPTCHAs have succeeded. We and other researchers (*e.g.* [CLSC05,MM03]) believe that many, perhaps most, of today's CAPTCHAs are vulnerable to custom-tailored preprocessing that segments the words into characters, followed by trainable OCR.

These observations motivated us to investigate CAPTCHAs which are likely to resist character–segmentation attacks. In [BR05] we first described the 'ScatterType' CAPTCHA: the image of each character making up a word is fragmented using horizontal and vertical cuts, then the fragments are forced to drift apart until it is difficult automatically to reassemble them into characters. Experimental data reported in that article showed that subjective ratings of difficulty by human readers were strongly and usefully correlated with illegibility. A systematic exploration of the legibility of ScatterType as a function of the generating parameters, detailed in [BMW05] and summarized in this paper, revealed an operating regime within which human legibility exceeds 95 per cent.

In this paper we describe the design of a new ScatterType trial to identify subregimes with well characterized difficulty (both subjective and objective). We also propose a methodology for offering ScatterType to the research community for experiments in CAPTCHA attack.

[0] Accepted for publication in *Proc., IAPR 8th Int'l Conf. on Document Analysis and Recognition*, Seoul, Korea, August 29 – September 1, 2005. [in press]

ScatterType Parameter	Range used in Trial
Cut Fraction (both x & y)	0.25-0.40
Expansion Fraction (both x & y)	0.10-0.30
Horizontal Scatter Mean	0.0-0.40
Vertical Scatter Mean	0.0-0.20
Scatter Standard Error (both h & v)	0.50
Character Separation	0.0-0.15

Figure 1. ScatterType parameter ranges selected for the first ScatterType human legibility trial.

2 Synthesizing ScatterType Challenges

In this section we briefly review the generating parameters (a fuller discussion can be found in [BR05,BMW05]). ScatterType challenges are synthesized by pseudorandomly choosing: (a) a text-string; (b) a typeface; and (c) cutting and scattering parameters. The text strings are English-like nonsense words generated pseudorandomly as described in [CB03]: in the original ScatterType trial, 4000 words were used; in the next trial, about 15,000 will be used. The typefaces may be chosen from a large set: in the original trial, twenty-one fonts were used; in the next trial, at least 100 fonts will be used.

The word is rendered as a bilevel image to which cutting and scattering operations are applied, separately to each character (more precisely, to each character's image within its own 'bounding box'). The parameters controlling this are, briefly:

Cutting Fraction Each character image is cut into rectangular blocks of this size, but random offset.

Expansion Fraction Fragments are moved apart by this distance, held constant across all characters in the string.

Horizontal Scatter Each row of cut fragments is shifted horizontally by a random distance chosen independently for each row. Adjacent rows alternate left and right movements.

Vertical Scatter Each fragment within a row is shifted vertically by this distance, as for horizontal scatter. Adjacent fragments within a row alternate up and down movements.

The resulting character images are combined using:

Character Separation The character images are separated horizontally by this distance.

	Difficulty Level					
	ALL	1	2	3	4	5
Total challenges	4275	610	1056	1105	962	542
% correct answers	52.6	81.3	73.5	56.0	32.8	7.7

Figure 6. Human reading performance as a function of the difficulty level that the subject selected.

3 The First Legibility Trial

Volunteers at Lehigh University and Avaya Labs Research attempted to read ScatterType challenges of the kind shown in Figure 2, using challenges generated using the parameter ranges summarized in Figure 1.

4 Experimental Results from the First Trial

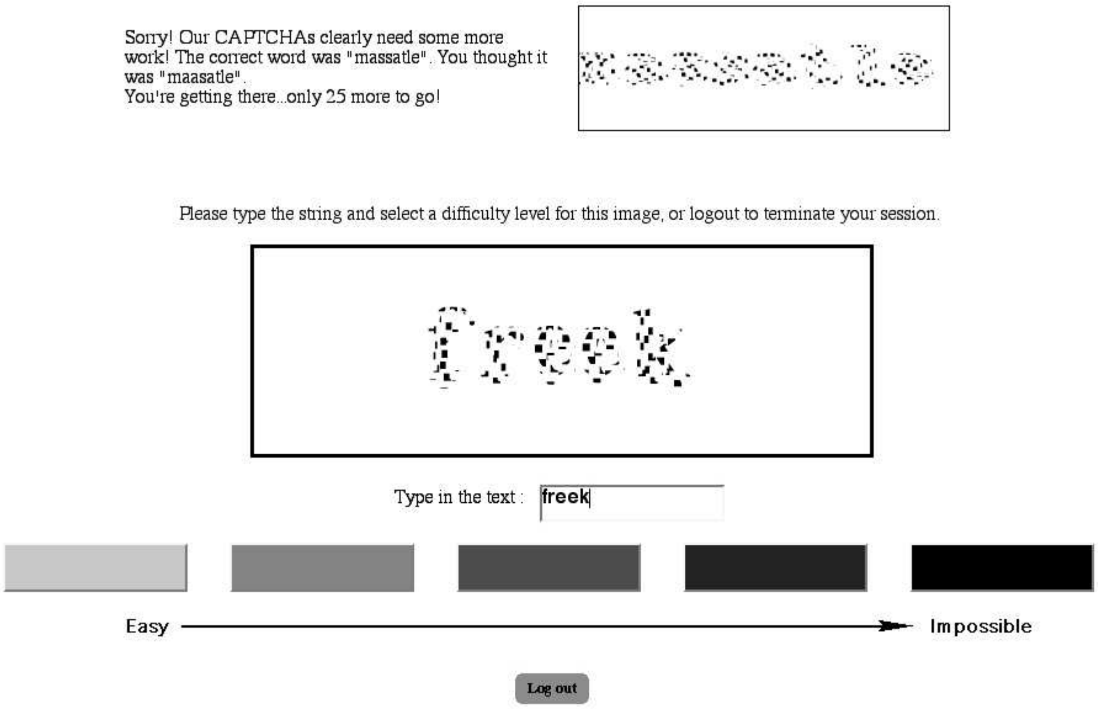
A total of 4275 ScatterType challenges were presented to human subjects: they are illustrated in Figures 3-5, at three subjective levels of difficulty: "Easy," medium difficulty, and "Impossible." Human legibility — percentage of challenges correctly read — is summarized in Figure 6. Overall, human legibility averaged 53%, and exceeded 73% for the two easiest levels. Legibility was strongly correlated with subjective difficulty level, falling off monotonically with increasing subjective difficulty (details in [BR05,BMW05]).

5 A Highly Legible Regime

Manual exploration of the trial results—using Tin Kam Ho's Mirage data analysis tool http://cm.bell-labs.com/who/tkh/mirage/index.html—revealed that judicious choices of the generating parameters can select levels of varying difficulty, both objective and subjective.

The Mirage scatter plot (Figure 7) with the mean horizontal scatter distance on the x-axis and the mean vertical scatter distance on the y-axis clearly shows a concentration of legible challenges in the lower left hand region near the origin. Good performance results by classifying all instances within an Euclidean distance of 0.25 from the origin as legible.

By limiting the mean scatter distances to less than 0.15, human legibility climbed to 80%. Simultaneously restricting the cut fraction parameter to 0.25 improved legibility, but only slightly. We examined the effect of pruning fonts and characters used in the nonsense words, and concluded that pruning fonts was unlikely to help. (Later, after pruning the worst performing characters, this was validated in



Sorry! Our CAPTCHAs clearly need some more work! The correct word was "massatle". You thought it was "maasatle".
You're getting there...only 25 more to go!

Please type the string and select a difficulty level for this image, or logout to terminate your session.

Type in the text : freek

Figure 2. An example of a ScatterType legibility trial challenge page. The Difficulty Level radio buttons (marked ’Easy’ to ’Impossible) were colored (left to right) Blue, Green, Yellow, Orange, and Red (these colors do not print in this Proceedings). The text at the top of the page refers to the challenge that was answered just before.

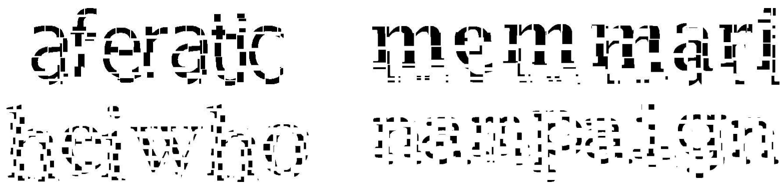


Figure 3. ScatterType challenges rated by subjects as “Easy” (difficulty level 1 out of 5). All of these examples were read correctly: “aferatic,” “memari,” “heiwho,” “nampaign.”

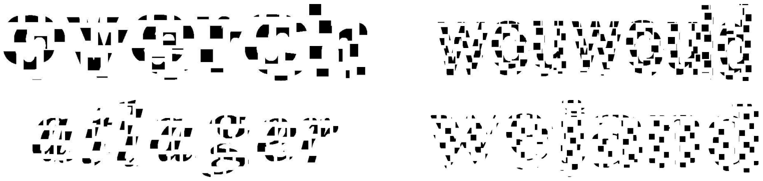


Figure 4. ScatterType challenges rated by subjects as being of medium difficulty (difficulty level 3 out of 5). Only one of these examples was read correctly (correct/attempt): “overch”/“ovorch”, “wouwould”, “atlager”/“adager”, “wejund”/“weland”.

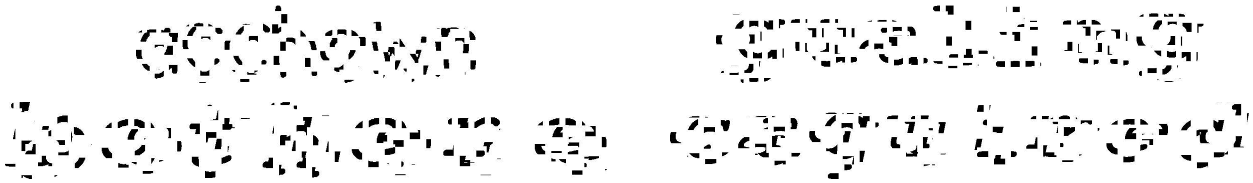


Figure 5. ScatterType challenges rated by subjects as “Impossible” (difficulty level 5 out of 5). None of these examples were read correctly (correct/attempt): “acchown”/“echaeva”, “gualing”/“gealthas”, “bothere”/“beadave”, “caquired”/“engaberse”.

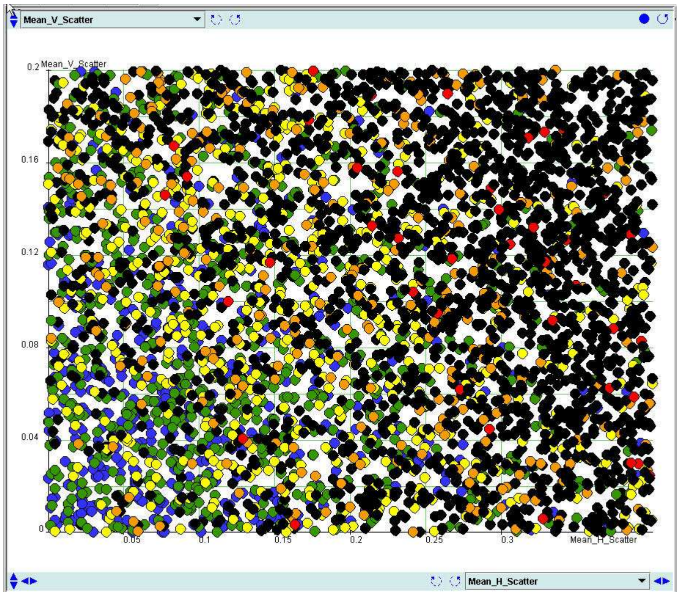


Figure 7. Mirage scatter plot of the Mean Horizontal Scatter (X-axis) versus Mean Vertical Scatter (Y-axis) parameters. Legible samples clustered strongly near the (0,0) origin. Black indicates a reading mistake: for legible samples, the colors red, orange, yellow, green, and blue (in this Proceedings these are visible only as shades of grey as in Figure 2) indicate five subjective difficulty levels from “impossible” to “easy”.

trials that included it with removing the worst performing characters, we tried this once more, and suffered a *loss* of legibility of four per cent). In the preliminary analysis in [BR05], when the five characters with the highest ”confusability”(’q’, ’c’, ’i’, ’o’, and ’u’) were removed brought us rapidly to above 90%. Combined with our new restrictions, we achieved a legibility of almost 93%. From this analysis we inferred that restricting mean scatter distances and pruning the worst performing characters are strongly positively correlated with legibility, while using larger cut fraction can be useful when used in combination with other features. Removing poorly performing fonts however seem to offer little benefit.

Among a large number of such policies, reported in detail in [BMW05], one achieved legibility of 97.5%, but for only 78 instances. This required pruning five characters and limiting the scatter distances to 0.1. To summarize, we have shown that through removing a small subset of easily confusable characters and restricting the range of two parameters, legibility can be raised to above 95%.

6 A Negative Result on Image Complexity

We also tried to predict legibility of ScatterType challenges using features that can be automatically extracted from the images. Tests on the ’Perimetric Image Complexity’ metric that worked well on BaffleText images [CB02] failed to predict legibility of ScatterType challenges.

7 Discussion and Future Work

A systematic analysis of the first ScatterType human legibility trial data has identified an operating regime — a combination of restrictions placed on generating parameters and pruning of the character set — which achieves legibility better than 95%. Within that regime we can pseudorandomly generate many millions of distinct ScatterType challenges.

We are designing a second legibility trial to replenish the data set and investigate how well can we automatically control difficulty levels, both objective and subjective.

ScatterType’s vulnerabilities to automatic attack seem to be, principally:

1. most of the fragments of characters will ‘fit’ perfectly if correctly joined—this might allow a kind of “jigsaw puzzle” attack;

2. the variable-length character n-gram model used to generate the nonsense words is of course more constraining than random strings—perhaps this model can be reverse-engineered or approximated and so constrain the search employed by attackers; and

3. our policy of using a single typeface within a word may allow automatic font inference.

Of course every CAPTCHA including ScatterType must be tested systematically using the best available OCR engines, and should be offered to the research community for attack by experimental machine vision methods. At this conference we will invite such attacks by the research community. For up-to-date instructions concerning the new trial and its ‘attackers entrance’, see `http://arcturus.cse.lehigh.edu/CAPTCHAs`.

8 Acknowledgments

We are grateful first to Terry Riopka, who implemented the first version of ScatterType and ran its first human legibility trial. We are also grateful to Patrice Simard, Dan Lopresti, Jarret Raim, and Jon Bentley for insightful advice. And we warmly thank the nearly 60 people who volunteered their time in the legibility trial.

References

[BAL00] M. Blum, L. A. von Ahn, and J. Langford, *The CAPTCHA Project*, “Completely Automatic Public Turing Test to tell Computers and Humans Apart,” `www.captcha.net`, Dept. of Computer Science, Carnegie-Mellon Univ., and personal communications, November, 2000.

[BMW05] H. S. Baird, M. A. Moll, & S-Y Wang, “A Highly Legible CAPTCHA that Resists Segmentation Attacks,” in H. S. Baird & D. P. Lopresti (Eds), *Proc., 2nd Int’l Workshop on Human Interactive Proofs (HIP2005)*, LNCS Vol. No. 3517, Springer (Berlin), pp. 27–41, May 2005.

[BR05] H. S. Baird and T. Riopka, “ScatterType: a Reading CAPTCHA Resistant to Segmentation Attack,” *Proc., IS&T/SPIE Document Recognition & Retrieval XII Conf,*, San Jose, CA, January 16–20, 2005.

[Bro01] AltaVista’s “Add-URL” site: `altavista.com/sites/addurl/newurl`, protected by the earliest known CAPTCHA.

[CB03] M. Chew and H. S. Baird, “BaffleText: a Human Interactive Proof,” Proc., 10th SPIE/IS&T Document Recognition and Retrieval Conf. (DRR2003), Santa Clara, CA, January 23–24, 2003.

[CBF01] A. L. Coates, H. S. Baird, and R. Fateman, “Pessimal Print: a Reverse Turing Test,” Proc., IAPR 6th Intl. Conf. on Document Analysis and Recognition, Seattle, WA, September 10-13, 2001, pp. 1154-1158.

[CLSC05] K. Chellapilla, K. Larson, P. Y. Simard, & M. Czerwinski, “Building Segmentation Based Human-Friendly Human Interactive Proofs (HIPs),” in H. S. Baird & D. P. Lopresti (Eds), *Proc., 2nd Int’l Workshop on Human Interactive Proofs (HIP2005)*, LNCS Vol. No. 3517, Springer (Berlin), pp. 1–26, May 2005.

[HB01] N. J. Hopper and M. Blum, “Secure Human Identification Protocols,” In: C. Boyd (Ed.) Advances in Crypotology, Proceedings of Asiacrypt 2001, LNCS 2248, pp.52 -66, Springer-Verlag Berlin, 2001

[LABB01] M. D. Lillibridge, M. Abadi, K. Bharat, and A. Z. Broder, “Method for Selectively Restricting Access to Computer Systems,” U.S. Patent No. 6,195,698, Issued February 27, 2001.

[MM03] G. Mori and J. Malik, "Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA," Proc., IEEE CS Society Conf. on Computer Vision and Pattern Recognition (CVPR'03), Madison, WI, June 16-22, 2003.

[SSB03] P. Y. Simard, R. Szeliski, J. Benaloh, J. Couvreur, I. Calinov, "Using Character Recognition and Segmentation to Tell Computer from Humans," Proc., IAPR Int'l Conf. on Document Analysis and Recognition, Edinburgh, Scotland, August 4–6, 2003.

[Tur50] A. Turing, "Computing Machinery and Intelligence," *Mind*, Vol. 59(236), pp. 433–460, 1950.