Captchæcker — Automating Usability-Security Evaluation of Textual CAPTCHAs

Maliha Nazir, Yousra Javed, Muhammad Murtaza Khan, Syed Ali Khayam

School of electrical engineering and computer science
National university of sciences and technology, Islamabad, Pakistan

Shujun Li

Department of Computer & Information Science
University of Konstanz, Germany

INTRODUCTION

CAPTCHAs are now deployed ubiquitously on the Internet to combat automated malicious programs. A major problem with many deployed textual CAPTCHA schemes is that they are either too weak in terms of security or unacceptable in terms of usability. This tradeoff can be more easily balanced if we can quantitatively evaluate the security and usability of a given textual CAPTCHA in an automated manner.

The main goal of the Captchæcker reported here is to automate the process of quantitatively evaluating the hardness of different kinds of textual CAPTCHAs as judged by an average user, thus providing a quantitative metric of usability and an indirect metric of security. We defined several geometric indicators to differentiate between easy and hard CAPTCHAs. User studies were done to collect subjective data for training and testing a neural network based CAPTCHAs classifier. The performance of the classifier was verified by a separate testing set and the average successful differentiation accuracy was found to exceed 80%.

Security-Usability Dilemma (I)

High Security ⇒ Low Usability

Figure 1(a): Three hard but unusable CAPTCHA images

Security-Usability Dilemma (II)

High Usability ⇒ Low Security

Figure 1(b): Three easy but unsecure CAPTCHA images

PROPOSED STRENGTH INDICATORS

Compactness-Length (CL)

→ Compact-Length (CL) = Compactness / CAPTCHA text width

→ Compactness (Cn) = Crowdedness [1] = Perimeter^2 / Area [2]

→ CAPTCHA text width (Cw) = Length of text in the CAPTCHA

→ Harder CAPTCHAs tend to have higher CL values.

EULER-Thickness (ET)

→ Euler-Thickness (ET) = Euler's Number / Erosion Steps

→ Euler's Number (EN) = Number of connected components — Number of holes

→ Erosion Steps (ES) = Number of steps to morphologically erode the CAPTCHA

→ Harder CAPTCHAs tend to have lower ET values.

CLASSIFICATION: TRAINING AND TESTING

Model: Supervised learning based binary classification: (CL, ET) ⇒ "hard" or "easy" (1 or 0); NN based learning.

Training Set: 20 users, 50 CAPTCHAs **Testing Set**: five new users, 38 new CAPTCHAs ⇒ **80% accuracy**

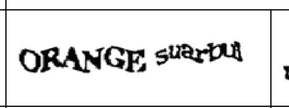
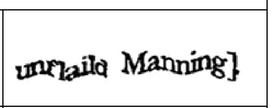
Google CAPTCHA		Google reCAPTCHA	
			
CL=7.43, ET=−1.14	CL=11.7, ET=−6.1	CL=4.8, ET=−0.44	CL=5.12, ET=−1.8
Microsoft CAPTCHA		**Yahoo! CAPTCHA**	
			
CL=5.57, ET=−0.25	CL=8.82, ET=−0.42	CL=13.1, ET=−1.16	CL=6.43, ET=−0.625

Figure 2: Four major textual CAPTCHA schemes were tested.

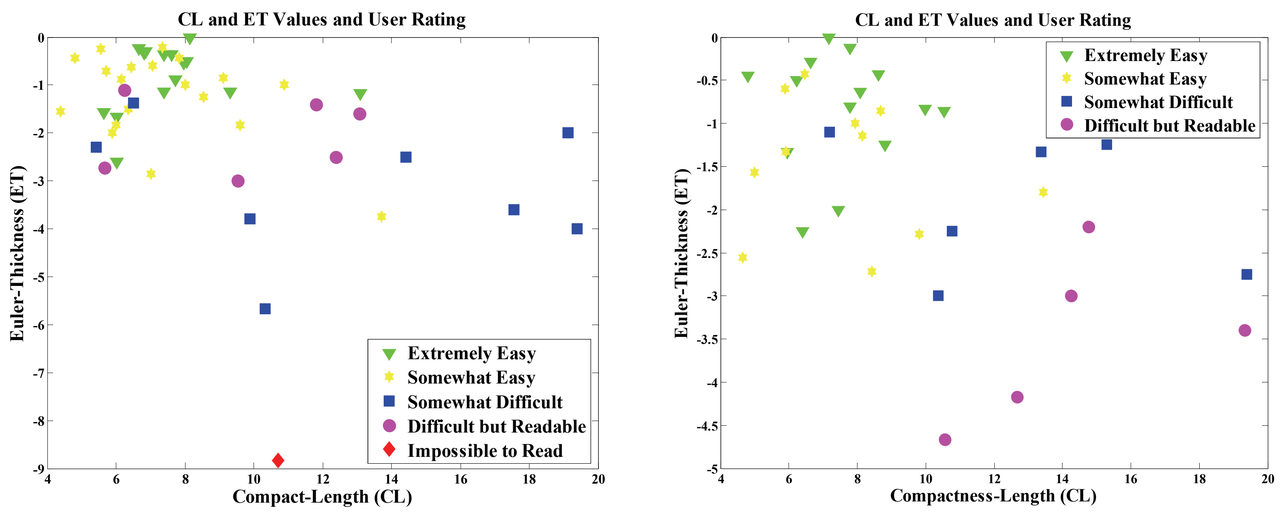


Figure 3: Data for training/testing the classifier: training set (left) and testing set (right).

CONCLUSION AND FUTURE WORK

→ Automated usability-security evaluation of CAPTCHAs is indeed possible!

→ A combination of two simple geometric indicators (CL + ET) allows us to predict hardness of textual CAPTCHAs with an accuracy of 80%.

→ Future work: 1) more subjective data (= more users); 2) more geometric indicators; 3) more CAPTCHAs; 4) publicly available subjective database and source code; 5) ...

REFERENCES

[1] 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* (CCS'08), pp. 543-554, ACM, 2008.

[2] Hermann Kremer and Eric W. Weisstein, "Isoperimetric Quotient," from MathWorld – a Wolfram web resource, http://mathworld.wolfram.com/IsoperimetricQuotient.html

SOUPS 2011 (7th Symposium On Usable Privacy and Security), Carnegie Mellon University campus, Pittsburgh, PA, USA, July 20, 2011