Text-based captchas: strengths and weaknesses

Elie Bursztein, Matthieu Martin, John Mitchell
Stanford University
About this Research

- Presenter: Elie Bursztein (http://elie.im)
- Conference: ACM CCS 2011
- Slides and paper freely available from http://ly.tl/p22
- Follow me for more security research
  - Twitter @elie
  - Google+
  - Facebook
Twitter Follower Packages

Please Select One Of Our Targeted Follower Pages

Silver Package

1000 Targeted Followers
- Guaranteed REAL, Targeted People Interested In Your Business
- Added To Your Page Within 25 days
- Targeted To Your Business/Niche
- Select The Country/s Where You Want Your Followers From
- No Automatic Bots/Programs To Get Followers, We Proudly Target 100% Of Your Followers Manually

$49.99
Order Now »

Gold Package

5000 Targeted Followers
- Guaranteed REAL, Targeted People Interested In Your Business
- Added To Your Page Within 40 days
- Targeted To Your Business/Niche
- Select The Country/s Where You Want Your Followers From
- No Automatic Bots/Programs To Get Followers, We Proudly Target 100% Of Your Followers Manually

$139.99
Funny Captchas

Captcha Validation: *

[Image of mathematical equation]
Funny Captchas

Captcha Validation: *

No premium user. Please enter the one that can NOT be created from the unfolded pattern. 29 seconds remain.
Funny Captchas

Captcha Validation: *

Please click on the images that show cats:
Funny Captchas

Please click on the images that show cats:

- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me
- [ ] adopt me

No premium user. Please enter the one that can NOT be created from the unfolded pattern. 29 seconds remain.
The world most-popular captchas

- [Megaupload]
- [Reddit]
- [eBay]
- [Baidu]
- [Authorize]
- [Digg]
- [Recaptcha]
- [NIH]
- [3.2 parks]
- [Google]
- [Captcha.net]
- [Wikipedia]
- [Slashdot]
- [CNN]
- [Skyrock]
- [Blizzard]
How to break text captcha and design secure ones
Based on the break of 13 of the most popular schemes
Outline
• How to break text-captcha?
Outline

• How to break text-captchas?
• Evaluating anti-recognition techniques security
Outline

• How to break text-captchas?
• Evaluating anti-recognition techniques security
• Attacking anti-segmentation techniques
Outline

• How to break text-captchas?
• Evaluating anti-recognition techniques security
• Attacking anti-segmentation techniques
• Real-world captcha security summary
Outline

• How to break text-captchas?
• Evaluating anti-recognition techniques security
• Attacking anti-segmentation techniques
• Real-world captcha security summary
• Decaptcha (our breaker) demo
Outline

• How to break text-captchas?
• Evaluating anti-recognition techniques security
• Attacking anti-segmentation techniques
• Real-world captcha security summary
• Decaptcha (our breaker) demo
• Lessons learned
Breaking captcha

**Divide and Conquer** approach
How to break captchas?

Slashdot captcha
How to break captchas?

Preprocessing

Slashdot captcha
How to break captchas?

Preprocessing

Slashdot captcha
How to break captchas?

Slashdot captcha

Segmentation

http://ly.tl/p22
How to break captchas?

Segmentation → fastest

Slashdot captcha
How to break captchas?

Post-segmentation → fastest

Slashdot captcha
How to break captchas?

Post-segmentation

Slashdot captcha
How to break captchas?

Recognition ➔ fastest

Slashdot captcha
How to break captchas?

Recognition → fastest

Slashdot captcha
How to break captchas?

Post-recognition → fastest

Slashdot captcha
How to break captchas?

Post-recognition

\text{fa s t e s t}

Slashdot captcha
Anti-recognition techniques
Anti-recognition techniques
Anti-recognition techniques

Blurring
Anti-recognition techniques

- Blurring
  - 3tr2bb

- Distortion
  - 0zt99n
## Anti-recognition techniques

<table>
<thead>
<tr>
<th>Anti-recognition Technique</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blurring</td>
<td>3tr2bb</td>
</tr>
<tr>
<td>Distortion</td>
<td>ozt99n</td>
</tr>
<tr>
<td>Rotation</td>
<td>a \ y \ b</td>
</tr>
</tbody>
</table>
Anti-recognition techniques

**Blurring**

3tr2bb

**Distortion**

0zt99n

**Rotation**

⊙ a ⊙ y ⊙ b

**Fonts**

0AGTP22

Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAS strenghts and weaknesses

http://ly.tl/p22
Anti-recognition techniques

Blurring

Distortion

Rotation

Fonts

Charsets
KNN learning rate

% success vs. Training set size for different conditions:
- 09
- AZ09
- azAZ09
- Distortion
- 3 fonts
- 5 fonts
- Angles
Anti-segmentation techniques
Background Confusion
Background Confusion

3nc9z
Background Confusion

3nc9z

quxg4h
Anti-recognition taxonomy

Background Confusion

3nc9z
quxg4h
p m y m k u
Anti-recognition taxonomy

Background Confusion

Lines

3nc9z

quxg4h

p my m ku
Anti-recognition taxonomy

Background Confusion

3nc9z

Lines

2p uc p
Background Confusion

Lines
Anti-recognition taxonomy

Background Confusion

3nc9z

Lines

zip

apricot

dramacharm
Anti-recognition taxonomy

Background Confusion

Lines

Collapsing
Anti-recognition taxonomy

Background Confusion

3nc9z
quxg4h
p m y m k u

Lines

2puc0
apricot
dramacharm

Collapsing

984505
Background Confusion

Lines

Collapsing
Anti-recognition taxonomy

Background Confusion

Lines

Collapsing
Breaking World of Warcraft
Breaking World of Warcraft
Breaking World of Warcraft
Breaking World of Warcraft
Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAS strengths and weaknesses

Breaking World of Warcraft
Breaking Captcha.net

Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAS strenghts and weaknesses

http://ly.tl/p22
Breaking Captcha.net
Breaking Captcha.net

Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAS strengths and weaknesses

http://ly.tl/p22
Breaking Wikipedia

dramacharm
dramacharm
Breaking Wikipedia
Breaking Wikipedia
Breaking Wikipedia
Breaking Digg
Breaking Digg
Breaking Digg

Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAS strengths and weaknesses

http://ly.tl/p22
Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAS strengths and weaknesses

http://ly.tl/p22

Breaking Slashdot
Breaking Slashdot

Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAs strengths and weaknesses

http://ly.tl/p22
Breaking Slashdot
Breaking eBay

Text-based CAPTCHAs strengths and weaknesses
Breaking eBay

Text-based CAPTCHAS strengths and weaknesses

Elie Burszttein, Matthieu Martin, John C. Mitchell

http://ly.tl/p22
Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAs strengths and weaknesses

http://ly.tl/p22
Failing to break eBay
Failing to break eBay
Failing to break eBay

Text-based CAPTCHAS strengths and weaknesses

Elie Bursztein, Matthieu Martin, John C. Mitchell

http://ly.tl/p22
Failing to break eBay
Failing to break eBay
Elie Bursztein, Matthieu Martin, John C. Mitchell

Breaking Baidu

Text-based CAPTCHAs strengths and weaknesses

http://ly.tl/p22
Breaking Baidu

Elie Bursztein, Matthieu Martin, John C. Mitchell

Text-based CAPTCHAs strengths and weaknesses

http://ly.tl/p22
Breaking Baidu
Real-world captchas security summary
## Overall results

<table>
<thead>
<tr>
<th>System</th>
<th>Segmentation rate</th>
<th>Solving rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorize</td>
<td>84%</td>
<td>66%</td>
</tr>
<tr>
<td>Baidu</td>
<td>98%</td>
<td>5%</td>
</tr>
<tr>
<td>Blizzard</td>
<td>75%</td>
<td>70%</td>
</tr>
<tr>
<td>Captcha.net</td>
<td>96%</td>
<td>73%</td>
</tr>
<tr>
<td>CNN</td>
<td>50%</td>
<td>16%</td>
</tr>
<tr>
<td>Digg</td>
<td>86%</td>
<td>20%</td>
</tr>
<tr>
<td>eBay</td>
<td>95%</td>
<td>43%</td>
</tr>
<tr>
<td>Google</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>MegaUpload</td>
<td>n/a</td>
<td>93%</td>
</tr>
<tr>
<td>NIH</td>
<td>87%</td>
<td>72%</td>
</tr>
<tr>
<td>Recaptha</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Reddit</td>
<td>71%</td>
<td>42%</td>
</tr>
<tr>
<td>Skyrock</td>
<td>30%</td>
<td>2%</td>
</tr>
<tr>
<td>Slashdot</td>
<td>52%</td>
<td>35%</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>57%</td>
<td>25%</td>
</tr>
</tbody>
</table>
Lessons learned
Building a breaker guidelines

• Immediate visual feedback
• Visual debugging
• Algorithm independence
• Exposing algorithm parameters
Decaptcha main interface
Core principles

- Randomize the length
- Randomize the character size
- Wave the captcha
Anti-recognition principles

- Use anti-recognition as a means of strengthening captcha security
- Don’t use a complex charset
  - Bad for human (see our research on this)
  - Useless for security
The Robustness of Google Captchas

- New heuristic to break the easy version of Google / Recaptcha
- Published online in May 2011
- Use letters shape as a side-channel
- Conclusion reduce your charset (not t or s...)

The Robustness of Google CAPTCHAs
Ahmad S El Ahmad, Jeff Yan, Mohamad Tayara
Anti-segmentation principles

• Use collapsing or lines
• Be careful in the implementation
• Create alternative schemes
Future

• Generic breaker for weak captchas
• Use higher-order features
  • to remove lines
  • Breaking collapsed captchas
Questions?

Captcha research
http://elie.im/tag/captcha

Follow-me on Twitter
@elie