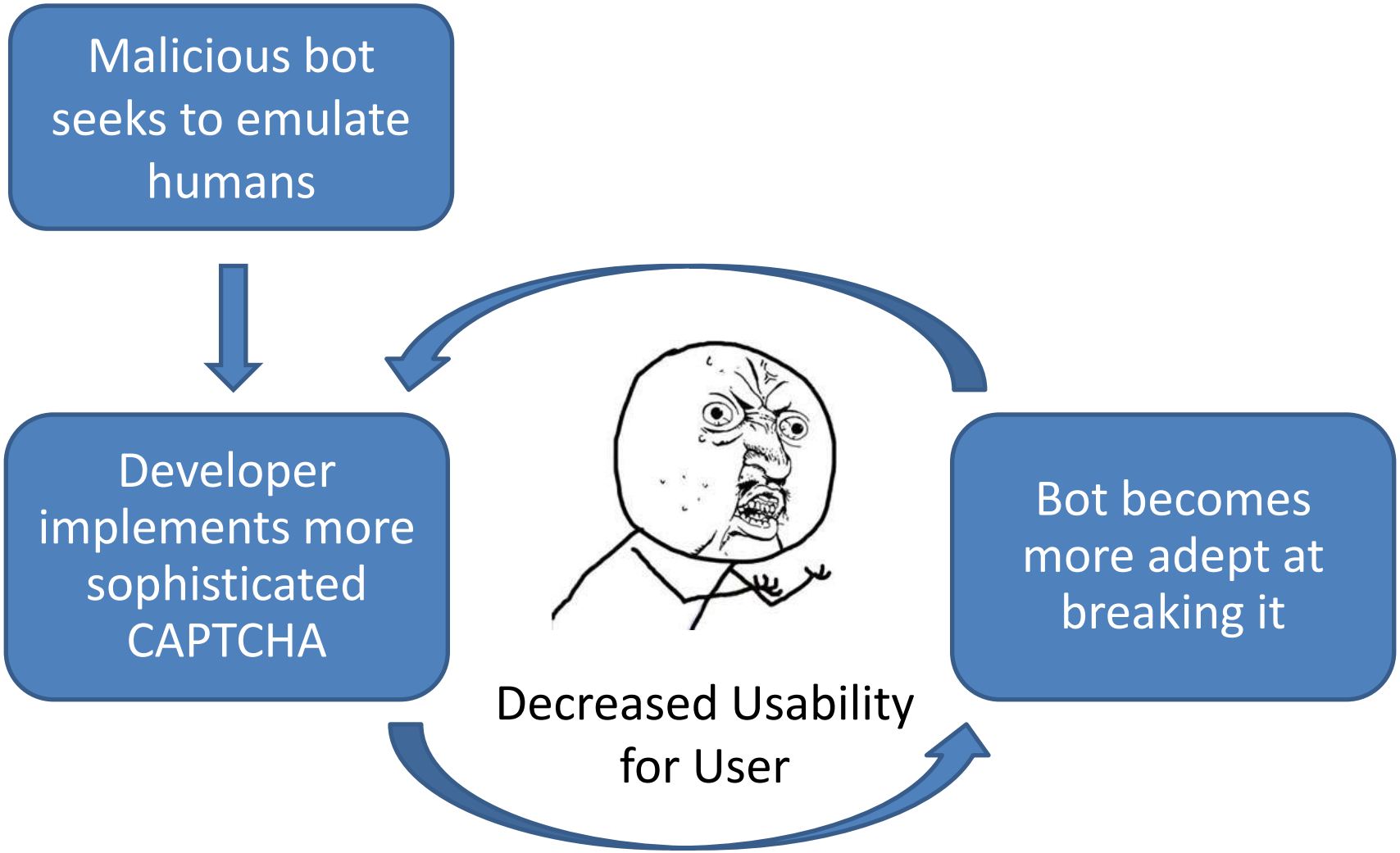
# Distinguishing Between Humans and Robots on the Web

Richard Abrich
Matthew Thorpe
Valentin Berbenetz

# Arms Race


Malicious bot seeks to emulate humans
Developer implements more sophisticated CAPTCHA
Bot becomes more adept at breaking it
Decreased Usability for User


http://captchafail.tumblr.com

# Techniques

- Many different techniques available
  - Active
    - Completely Automated, Semi-Automated
  - Passive
- Inventing a new technique is hard
  - Instead, evaluate current techniques and propose a combination

# Techniques

| | Active | Passive |
|---|---|---|
| CAPTCHA | Text | Biometric Data Analysis |
| CAPTCHA | Audio | Honeypot Form Elements |
| CAPTCHA | Video | JavaScript Detection |
| CAPTCHA | Confirmation Page | |
| CAPTCHA | Unusual Form Interaction | |
| SAPTCHA | Word | |
| SAPTCHA | Image | |

# Evaluation

| Source | Criteria | | Weight ($W_i$) |
|---|---|---|---|
| Literature | Effectiveness | $X_1$ | 5 |
| Website Implementation | Library availability | $X_2$ | 4 |
| | Automation | $X_3$ | 3 |
| User Survey | Time to solve | $X_4$ | 2 |
| | Difficulty | $X_5$ | 1 |
| | % Cancellations | $X_6$ | 1 |
| | % Correct | Z | - |

$Y$ - Final Score

$k$ - Technique

$i$ - Criteria

$$Y_k = Z \sum_{i=1}^{6} W_i \frac{X_i}{\max_k X_{i,k}}$$

# Demonstration

http://thecaptchaexperiment.com

# Results

| Rank | Technique | Score |
|---|---|---|
| 1 | Video Captcha | 12.5 |
| 2 | Biometric Data | 12.1 |
| 3 | Text Captcha | 11.7 |
| 4 | Unusual Form Interaction | 9.3 |
| 5 | Image Saptcha | 8.9 |
| 6 | Honeypot Form Elements | 8.8 |
| 7 | JavaScript Detection | 8.0 |
| 8 | Word Saptcha | 7.0 |
| 9 | Confirmation Page | 6.8 |
| 10 | Audio Captcha | 3.7 |

# Proposed Solution

- Combine Active and Passive techniques
  - Active: something you **know**
  - Passive: something you **are**

- Active: **Video CAPTCHA**
  - Best overall technique
  - Sub-optimal effectiveness

- Passive: **Biometric Data**
  - Highest effectiveness

# Proposed Solution

| Criteria | Video CAPTCHA | Biometric Data | Combined |
|---|---|---|---|
| Effectiveness | 82.5 | 95 | 99.13* |
| Library availability | 3 | 1 | 2 |
| Automation | 3 | 3 | 3 |
| Time required | 12.28 | - | 12.28 |
| Difficulty | 2.18 | - | 2.18 |
| Cancellation rate | 15.21 | - | 15.21 |
| Success rate | 97.21 | 100 | 97.21 |
| **Final Score** | **12.51** | **12.08** | **13.20** |

* Combined Effectiveness = 1 – (1 – 0.825) * (1 – 0.95)

# Conclusion

- All existing techniques still have major flaws
  - Video CAPTCHA best overall technique
- Effectiveness of combination of techniques requires further study

http://thecaptchaexperiment.com