A CAPTCHA in the Rye

Tal Be’ery,
Web Research TL, Imperva

OWASP
The Open Web Application Security Project

Tal Be’ery

Web Security Research Team Leader @Imperva
Holds MSc & BSc degree in CS/EE from TAU
10+ years of experience in IS domain
Facebook “white hat”
Speaker at RSA, BlackHat, AusCERT
Columnist for securityweek.com

Agenda

Introduction to Hacker Intelligence Initiative
Automation on the Web
- Good bots, bad bots

CAPTCHA
- Caveats
- Mitigation

Case study analysis
Summary of recommendations

Hacker Intelligence Initiative

Hacker Intelligence Initiative (HII)

OWASP
The Open Web Application Security Project

The Hacker Intelligence Initiative is focused on understanding how attackers operate in practice

- A different approach from vulnerability research

Data set composition

- ~50 real world applications
- Anonymous Proxies

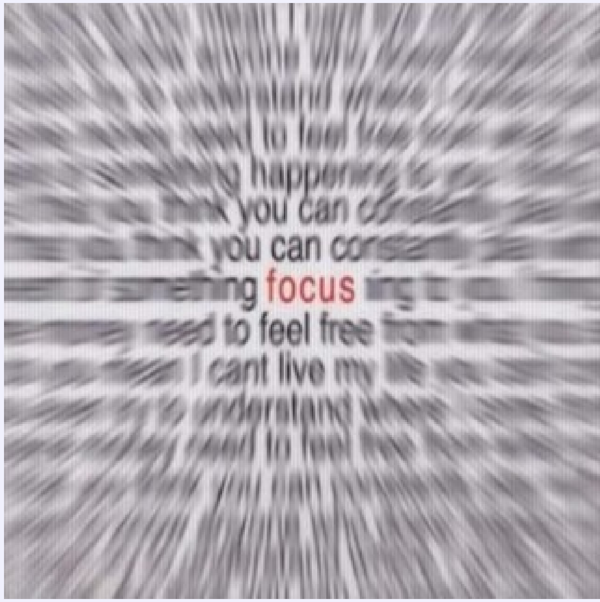
More than 18 months of data

Powerful analysis system

- Combines analytic tools with drill down capabilities

HII - Motivation

- Focus on actual threats
  - Focus on what hackers want, helping good guys prioritize
  - Technical insight into hacker activity
  - Business trends of hacker activity
  - Future directions of hacker activity
- Eliminate uncertainties
  - Active attack sources
  - Explicit attack vectors
  - Spam content
- Devise new defenses based on real data
  - Reduce guess work



HII Reports

Monthly reports based on data collection and analysis

Drill down into specific incidents or attack types

2011 / 2012 reports

- Remote File Inclusion
- Search Engine Poisoning
- The Convergence of Google and Bots
- Anatomy of a SQLi Attack
- Hacker Forums Statistics
- Automated Hacking
- Password Worst Practices
- Dissecting Hacktivist Attacks
- CAPTCHA Analysis

WAAR – Web Application Attack Report

Semi annual

Based on aggregated analysis of 6 / 12 months of data

Motivation

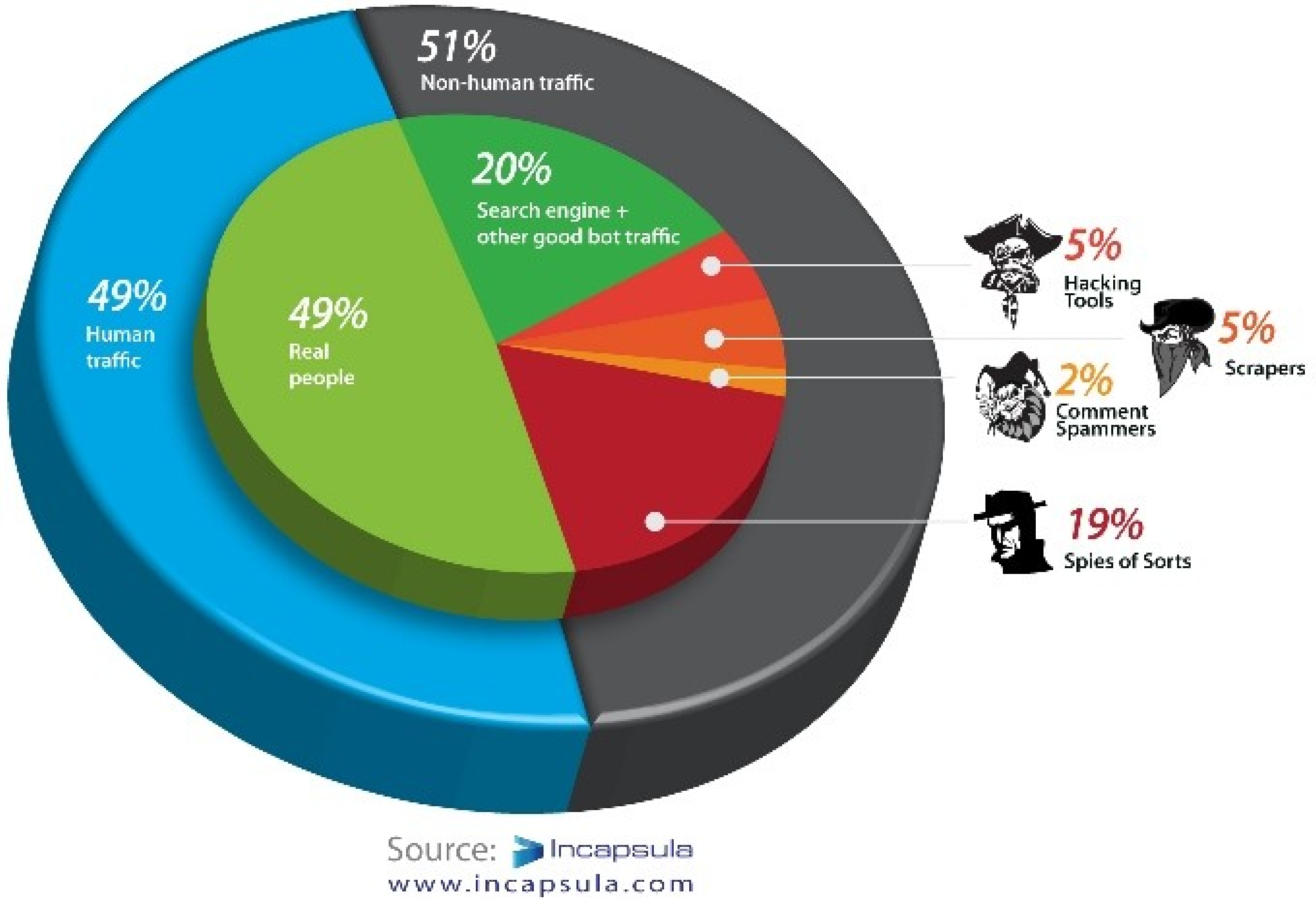
- Pick-up trends
- High level take outs
- Create comparative measurements over time

Automation on the Web

2012’s Web: Automation all over the place

OWASP
The Open Web Application Security Project

Human traffic is in the minority



Source:
http://www.incapsula.com/the-incapsula-blog/item/225-what-google-doesnt-show-you-31-of-website-traffic-can-harm-your-business

Good Automation

- Search engines
  - E.g. GoogleBot
- Validators
  - Link checkers
  - CSS/HTML/.. Format validators
  - Friendly vuln scan
- RSS feed readers
  - IE RSS reader
- B2B

Good Bot

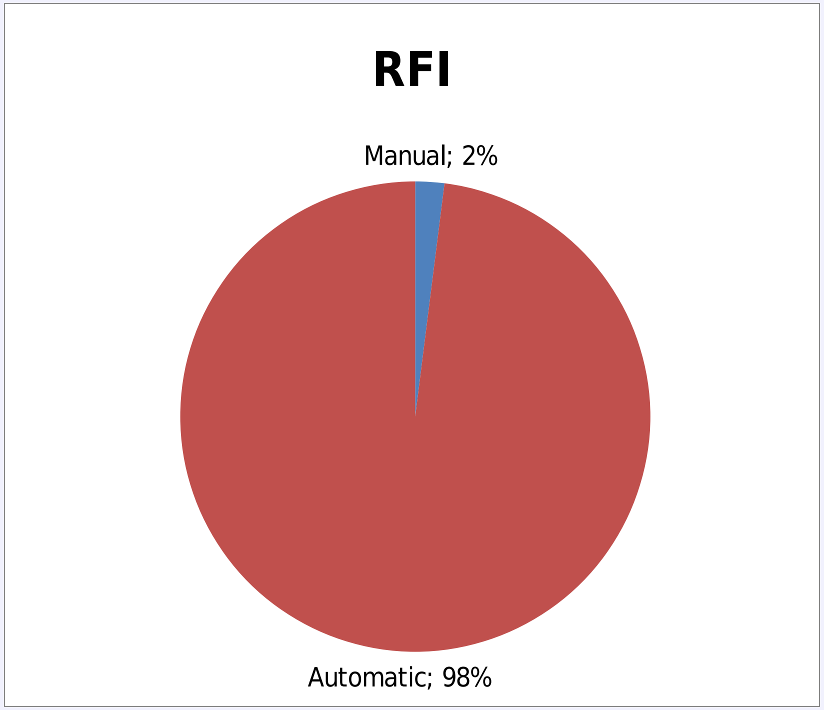
OWASP
The Open Web Application Security Project

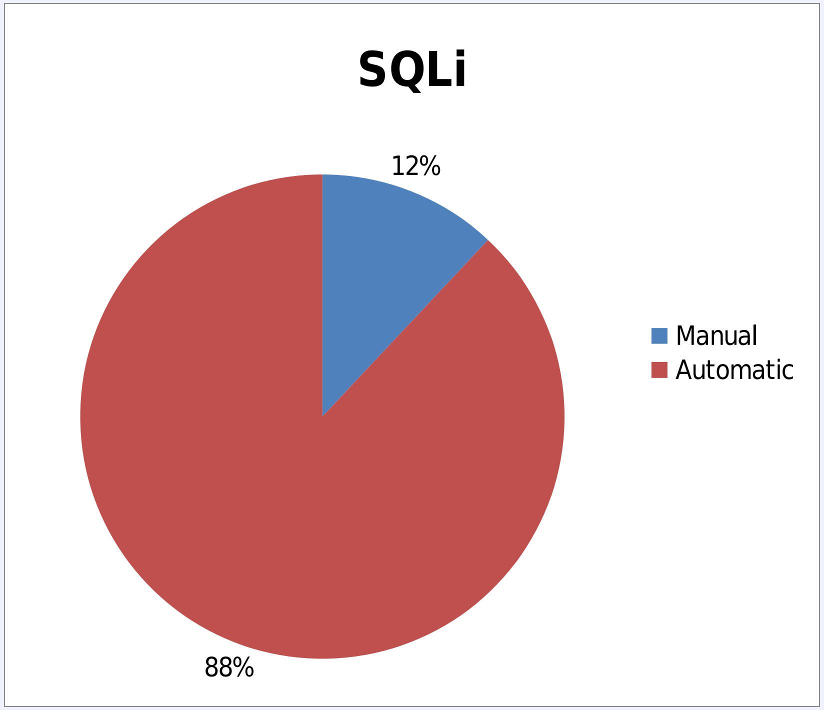
- A good bot is a polite bot
- Introduces itself
  - User agent
- Specifies a method to validate identity
  - Usually by reverse DNS
- Keeps the house rules - adheres to robots.txt
  - Who can crawl
  - What can be crawled
  - Rate of crawling

Bad Automation I: Hacking

OWASP
The Open Web Application Security Project

Web application hacking attacks



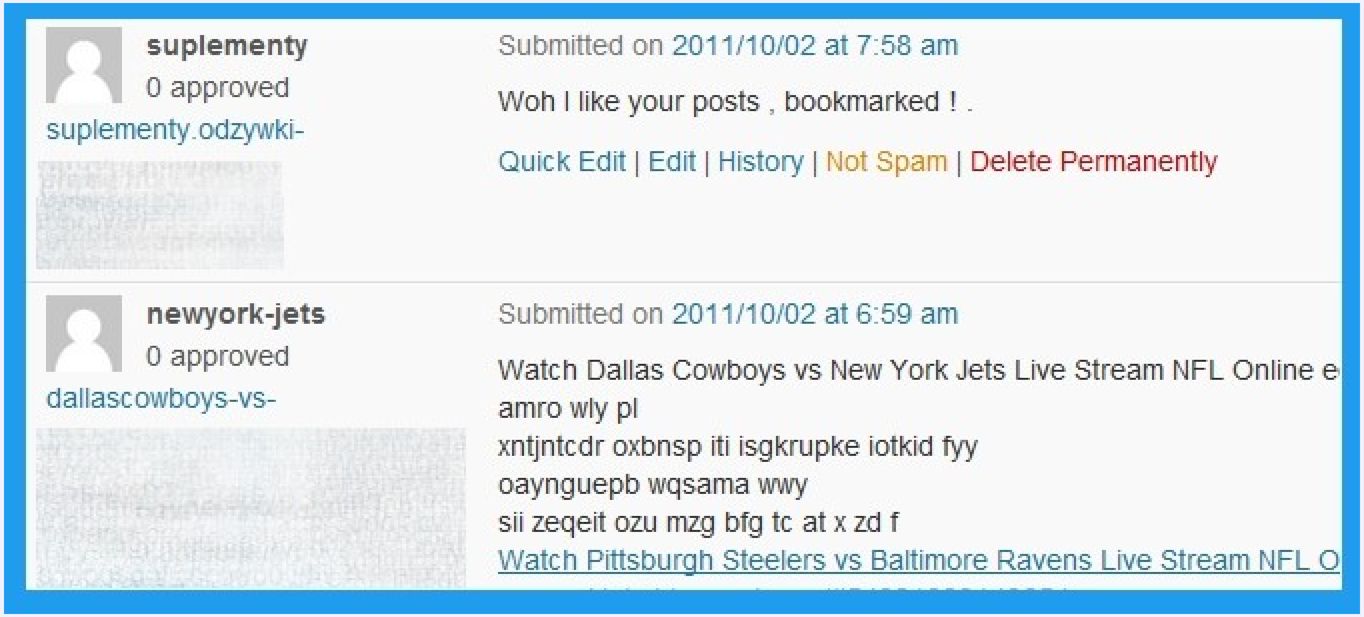


Source:
http://www.imperva.com/docs/HII_Automation_of_Attacks.pdf

Bad Automation II: Comment Spamming

OWASP
The Open Web Application Security Project

Abusing comment functionality to embed spam content



suplementy
0 approved
suplementy.odzywki-
[redacted]

Submitted on 2011/10/02 at 7:58 am

Woh I like your posts , bookmarked ! .

Quick Edit | Edit | History | Not Spam | Delete Permanently

newyork-jets
0 approved
dallascowboys-vs-
[redacted]

Submitted on 2011/10/02 at 6:59 am

Watch Dallas Cowboys vs New York Jets Live Stream NFL Online e
amro wly pl
xntjntcdr oxbnsp iti isgkrupke iotkid fyy
oaynguepb wqsama wwy
sii zeqeit ozu mzg bfg tc at x zd f
Watch Pittsburgh Steelers vs Baltimore Ravens Live Stream NFL O

Bad Automation III: Site Scraping

- Stealing site’s Intellectual Property
  - PII from government sites
  - Price quotes
  - Stealing media (images) from media sites
- We will analyze some “in the wild” examples

CAPTCHA Defined

CAPTCHA Defined

Completely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part

A good CAPTCHA is a test

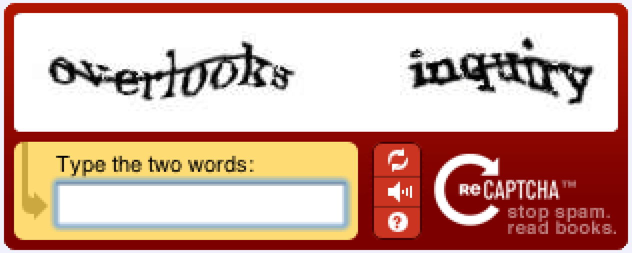
- Easy for humans
- Hard for computers

Can be used to fight automation

CAPTCHA Implementations

Hosted solutions

- ReCAPTCHA - Acquired by Google



Application Add-ons

- PHP CAPTCHA

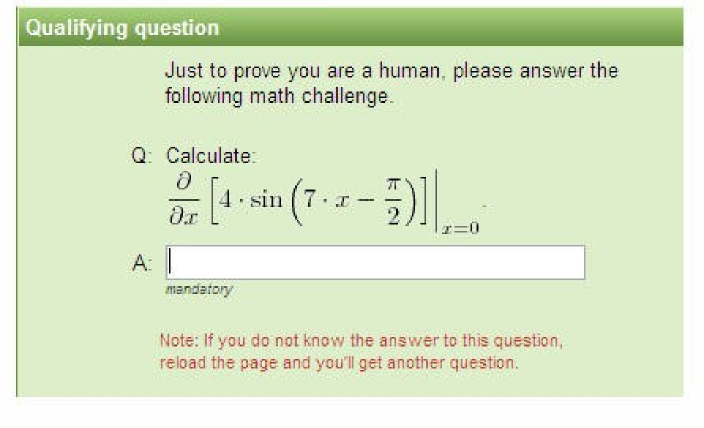
CAPTCHAS Caveats

Bad implementations

- Too easy for computers
- Too hard for humans
- Sometimes both

Can be defeated by “Artificial Artificial” (Artificial^2) Intelligence

- Mechanical turk



Qualifying question

Just to prove you are a human, please answer the following math challenge.

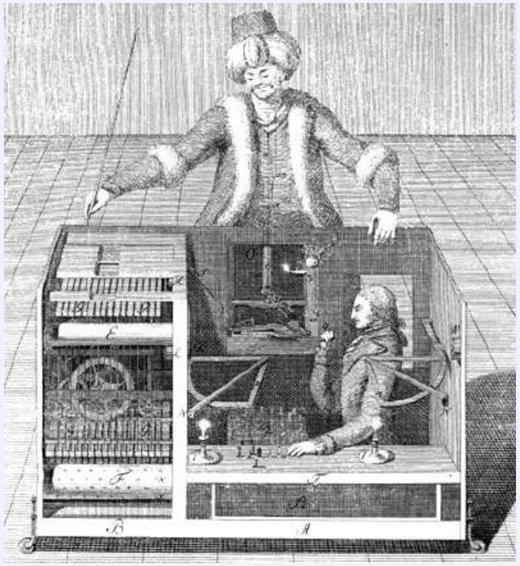
Q: Calculate:

$$\frac{\partial}{\partial x}\left[4\cdot\sin\left(7\cdot x-\frac{\pi}{2}\right)\right]\Bigg|_{x=0}.$$

A:

mandatory

Note: If you do not know the answer to this question, reload the page and you'll get another question.



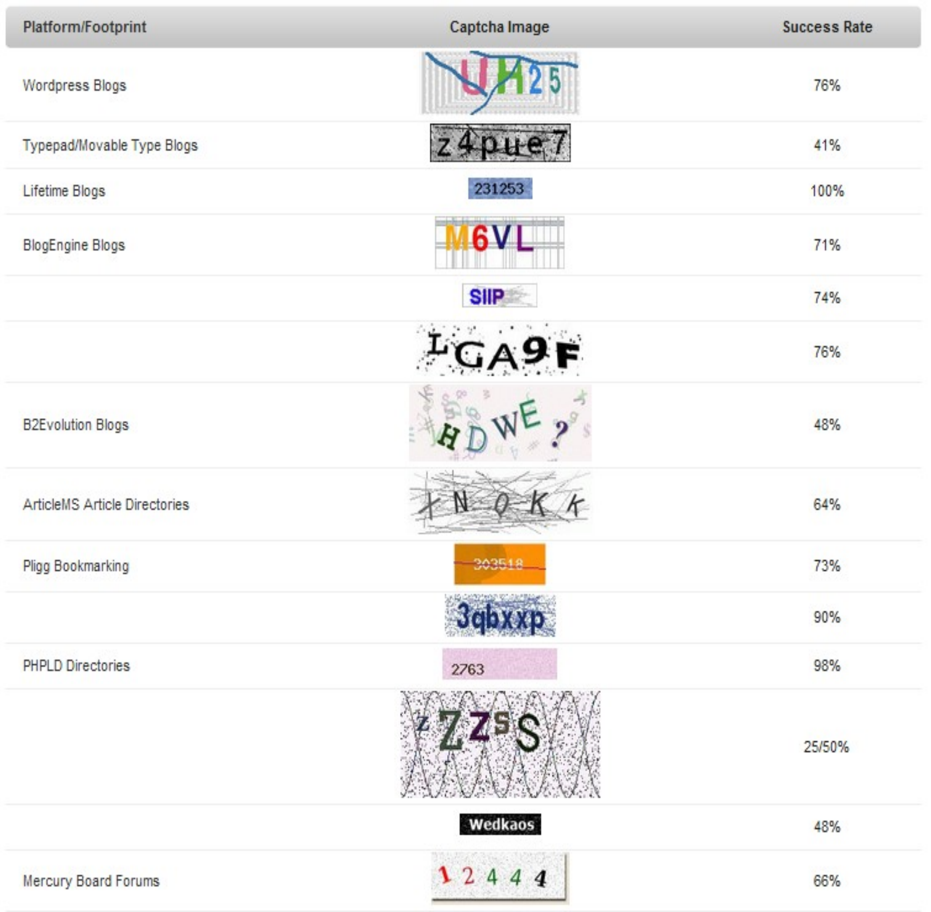
Source:
http://www.johnmwillis.com/other/top-10-worst-captchas/

Bad CAPTCHA: Easy for Computers

Many are based on the character recognition problem

Can be broke with OCR based tool

CAPTCHA Sniper tool



Platform/Footprint	Captcha Image	Success Rate
Wordpress Blogs	UH25	76%
Typepad/Movable Type Blogs	z4pue7	41%
Lifetime Blogs	231253	100%
BlogEngine Blogs	M6VL	71%
	SIIP	74%
	LGA9F	76%
B2Evolution Blogs	HDWE?	48%
ArticleMS Article Directories	XNQKk	64%
Pligg Bookmarking	303518	73%
	3qbxxp	90%
PHPLD Directories	2763	98%
	zZZsS	25/50%
	Wedkaos	48%
Mercury Board Forums	12444	66%

Bad CAPTCHA: Easy for Computers

Low entropy

Example “what’s the animal in the picture”

10,000 animal pictures

Attackers can

- Solve each picture once and bypass CAPTCHA forever
- Guess thousands of times until they get it right
  - Computers don’t get bored in the process

Known to happen with many “Audio CAPTCHA”

Bad CAPTCHA: Easy for Computers

Attacker can force the specific CAPTCHA test

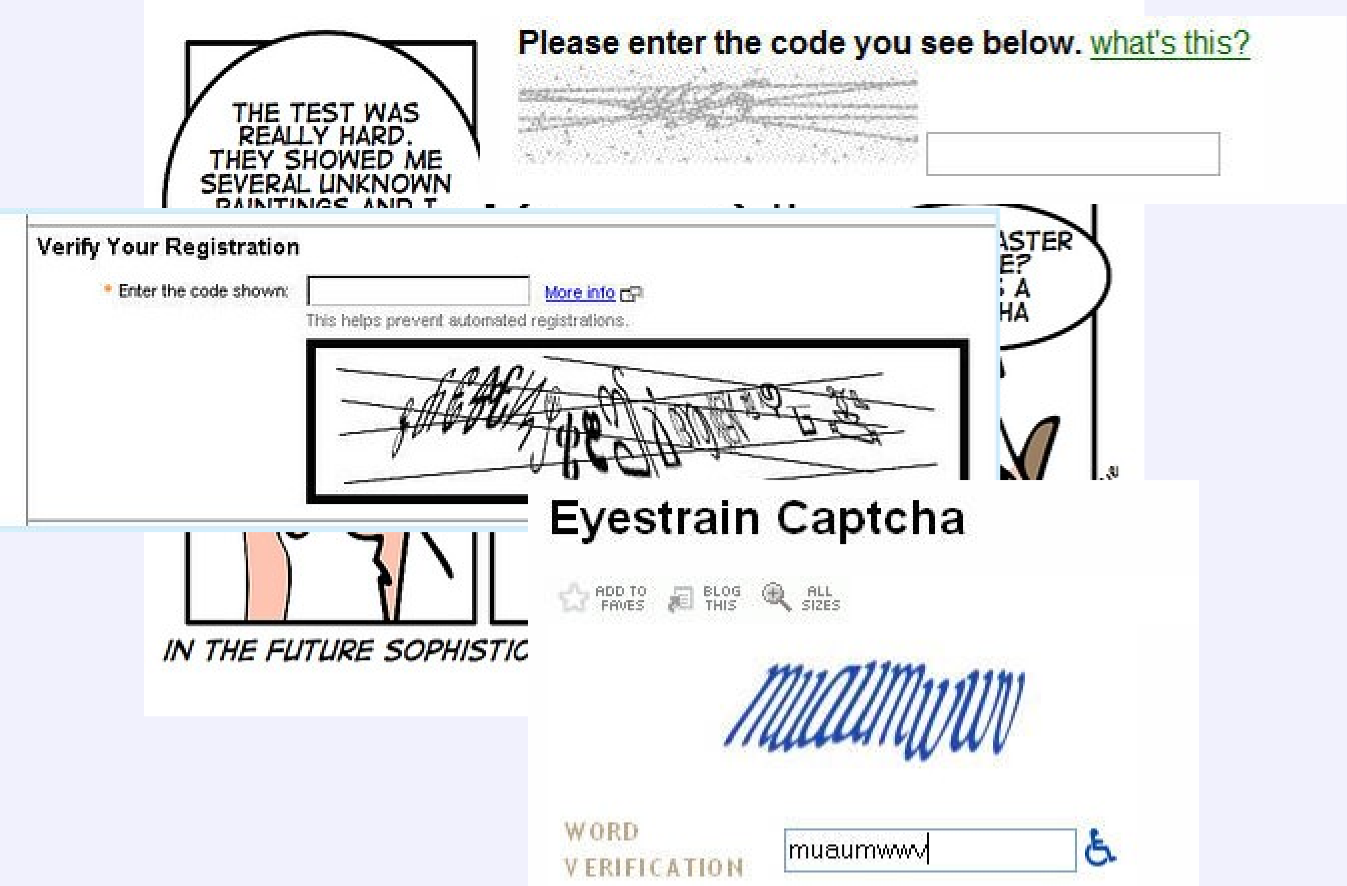
The servers validates the answer based on some value passed by the client

/captcha.jsp?test_id=1234&answer=cat

Attackers can solve a single test once and bypass CAPTCHA forever

Bypassing CAPTCHA: Artificial Artificial

Can be very annoying


THE TEST WAS REALLY HARD. THEY SHOWED ME SEVERAL UNKNOWN PAINTINGS AND I
ASTER
E?
A
HA
IN THE FUTURE SOPHISTIC

Please enter the code you see below. what's this?

Verify Your Registration

* Enter the code shown: More info

This helps prevent automated registrations.

Eyestrain Captcha

ADD TO FAVES
BLOG THIS
ALL SIZES

WORD VERIFICATION muaumwwv

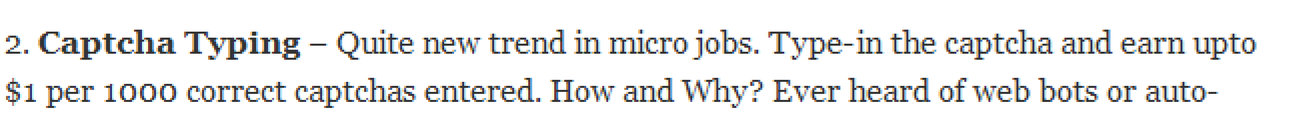
Source:
http://ordinarygaming.blogspot.co.il/2012/08/dear-captcha-i-hate-you-and-wish-you.html

Bypassing CAPTCHA: Artificial^2 intelligence

Convincing humans to solve CAPTCHA

Money

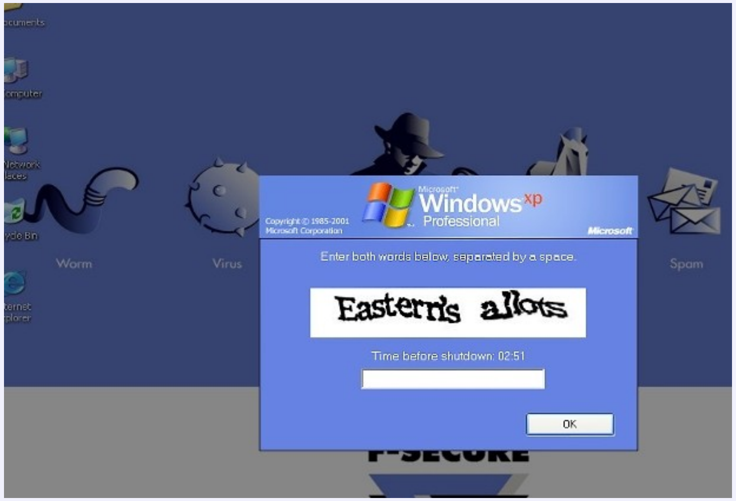
- Paying for micro jobs



2. **Captcha Typing** – Quite new trend in micro jobs. Type-in the captcha and earn upto $1 per 1000 correct captchas entered. How and Why? Ever heard of web bots or auto-

Extortion

Shutdown Malware

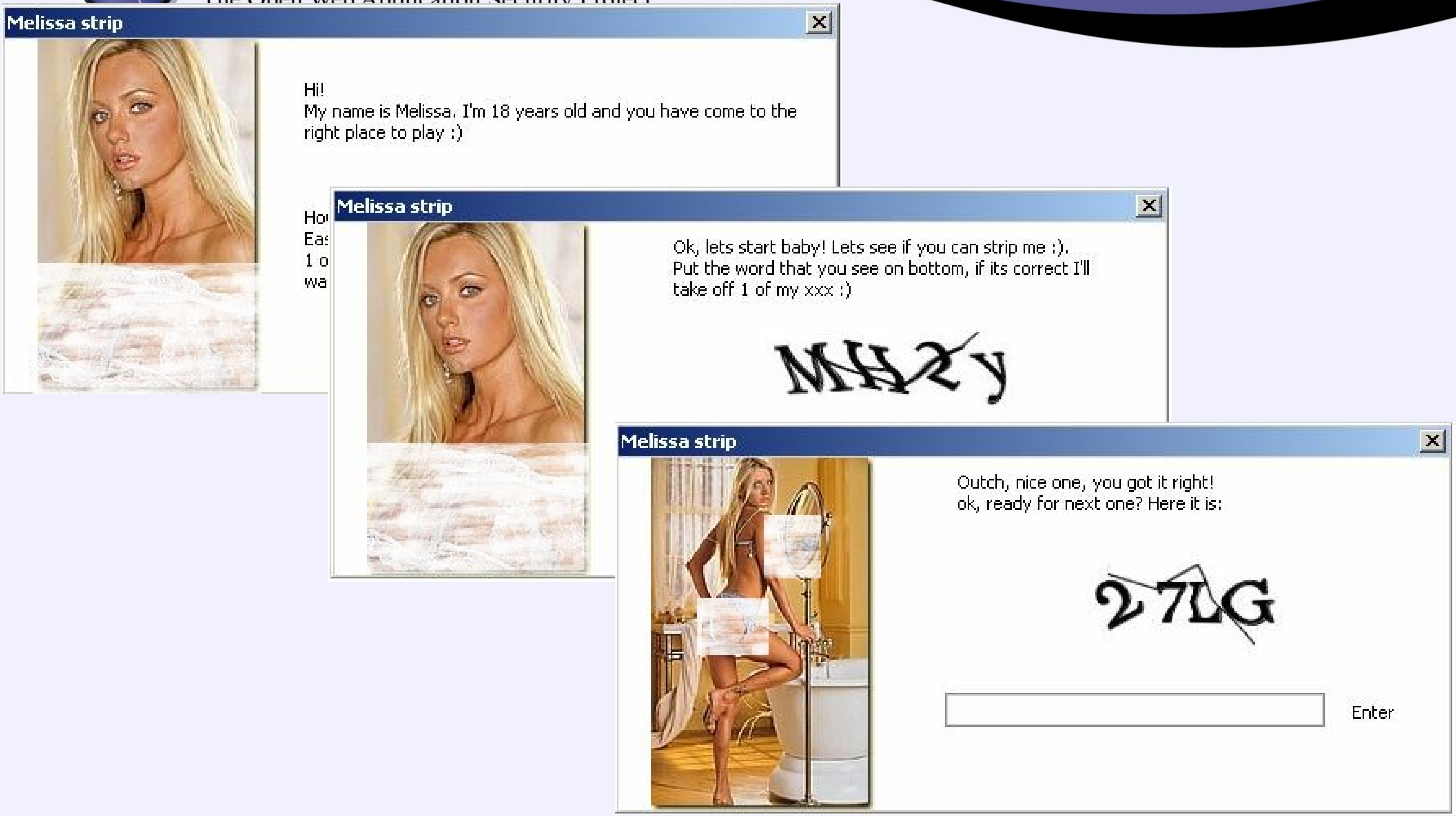


Enter both words below, separated by a space.

Time before shutdown: 02:51

Worm
Virus
Spam
Microsoft Windows xp Professional
Copyright © 1985-2001 Microsoft Corporation
Microsoft
Easterns allots
OK
F-SECURE

Bypassing CAPTCHA: Playing Strip CAPTCHA



Hi!
My name is Melissa. I'm 18 years old and you have come to the right place to play :)

Ho
Eas
1 o
wa

Ok, lets start baby! Lets see if you can strip me :).
Put the word that you see on bottom, if its correct I'll take off 1 of my xxx :)

Outch, nice one, you got it right!
ok, ready for next one? Here it is:

Source:
http://www.ghacks.net/2007/12/07/melissa-strip-captcha-breaker-trojan/

Detecting Automation: Additional Automats Tests I

OWASP
The Open Web Application Security Project

Adding defense dimensions

- Augmenting CAPTCHA with other anti-automation measures

Detecting Automation: Additional Automats Tests II

The combination of tests makes bypassing harder

- Tests cannot be solved by merely exporting to humans

Invisible tests don’t change User Experience

Detecting Automation: Passive

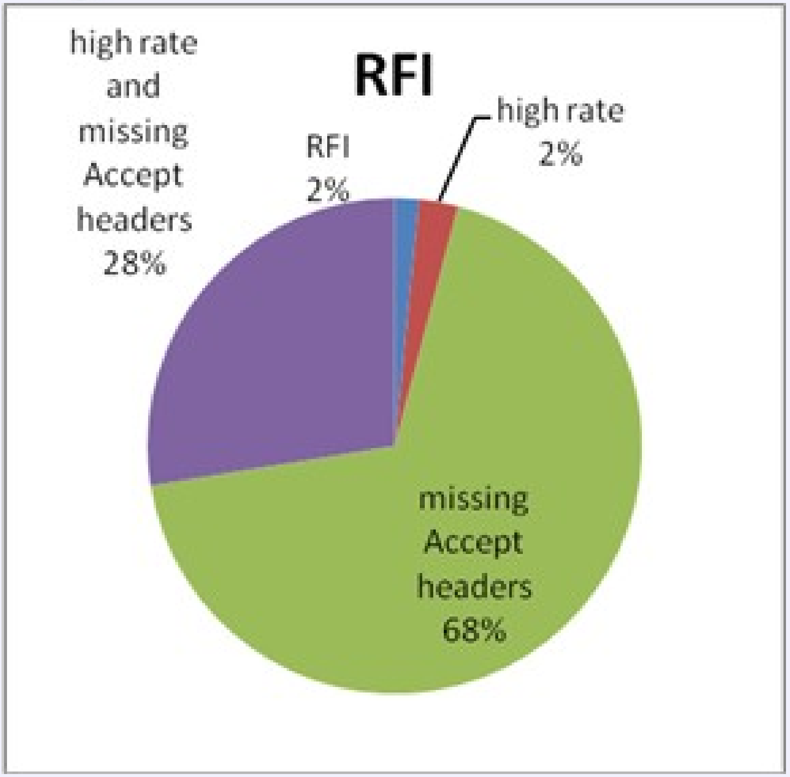
OWASP
The Open Web Application Security Project

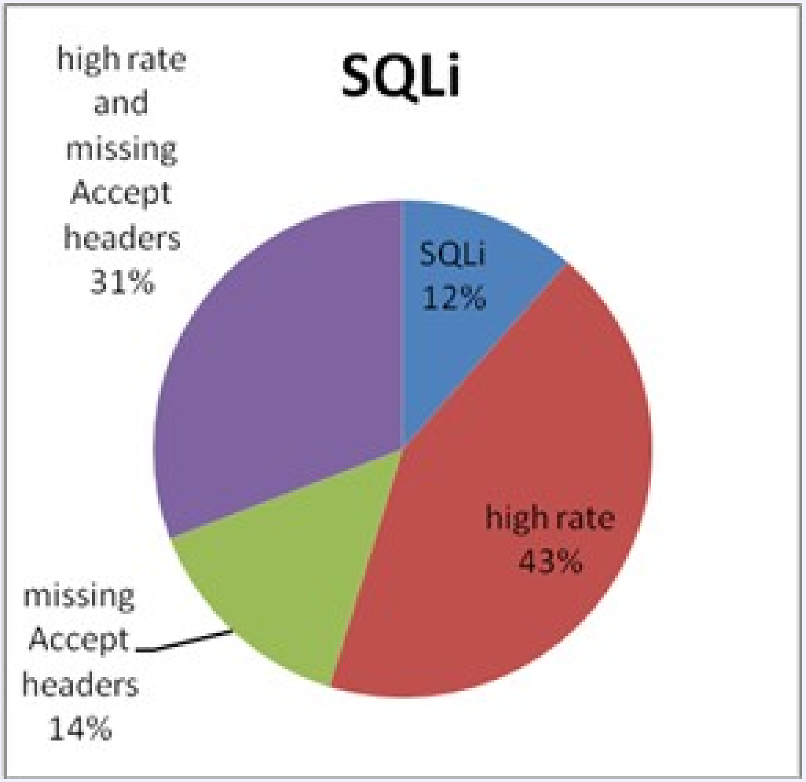
Passive Methods
- Watch network traffic “as-is”
- Non intrusive, do not affect user experience

Traffic Shape Indicators
- We measure suspicious requests (rather than ALL requests)
- Measured attributes
    - Rate
    - Rate change (ramp-up speed)
    - Volume
- Difficult to measure in an inherently noisy source (NAT)

Request Shape Indicators
- Missing headers
- Mismatch between headers and location

Detecting Automation: Passive





Detecting Automation: Active

Introduce changes into the server response

- Test client's reaction to changes
- May affect user experience – use with care
- Verify type of user agent

Browsers support Javascript and an appropriate DOM

- Client is expected to complete some computation
- Application / GW can validate the computed value

Browsers comply with HTML tags (IMG, IFRAME)

- Client is expected to access resource referenced by embedded tags
- Failure to access the resources implies that client is an automated script

Detecting Automation: Wisdom of the Crowds

Detected automation feeds into building fingerprints of tools and reputation data for sources

Leveraged when data is collected within a community

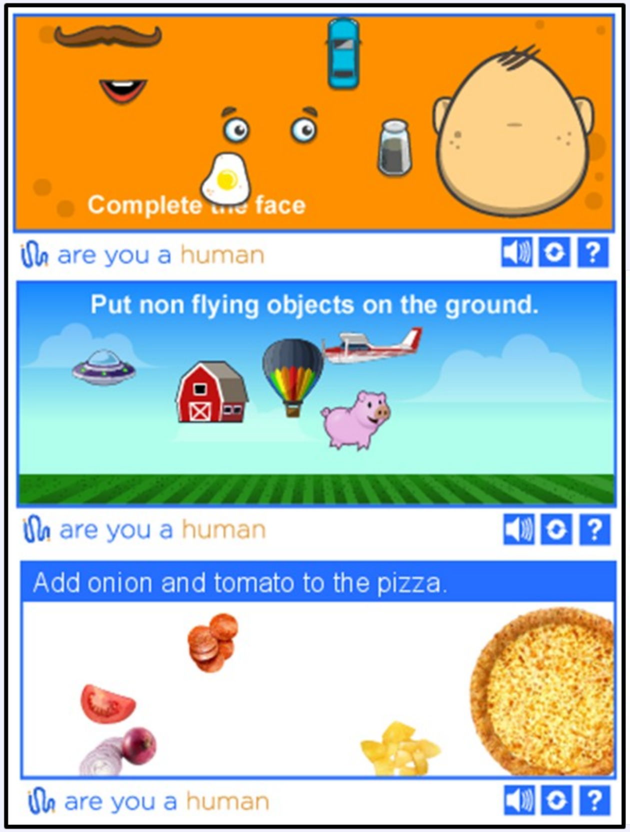
Recent regulatory changes endorse the concept of community

Drop requests matching fingerprints or coming from ill reputed sources

Detecting Automation: CAPTCHA 2.0

OWASP
The Open Web Application Security Project

Gamification - CAPTCHAs that are more fun for humans but hard for computers


Complete the face
are you a human
Put non flying objects on the ground.
are you a human
Add onion and tomato to the pizza.
are you a human

Case Study Analysis

Case Study: Attacked Site

South American government tax agency

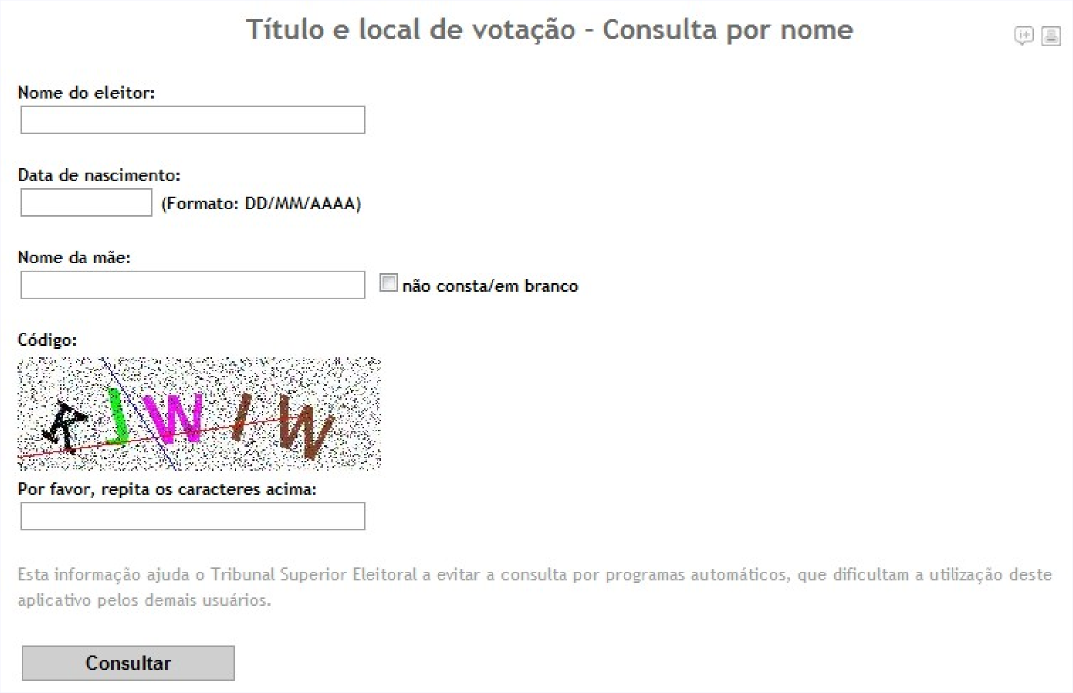
Displays tax statement per company unique ID

Protected against automation with CAPTCHA

Having the whole database offline would allow attackers to run arbitrary additional queries on the database to get financial information

Case Study: The CAPTCHA

5 random letters
Fairly easy to OCR



Título e local de votação - Consulta por nome

Nome do eleitor:

Data de nascimento:
(Formato: DD/MM/AAAA)

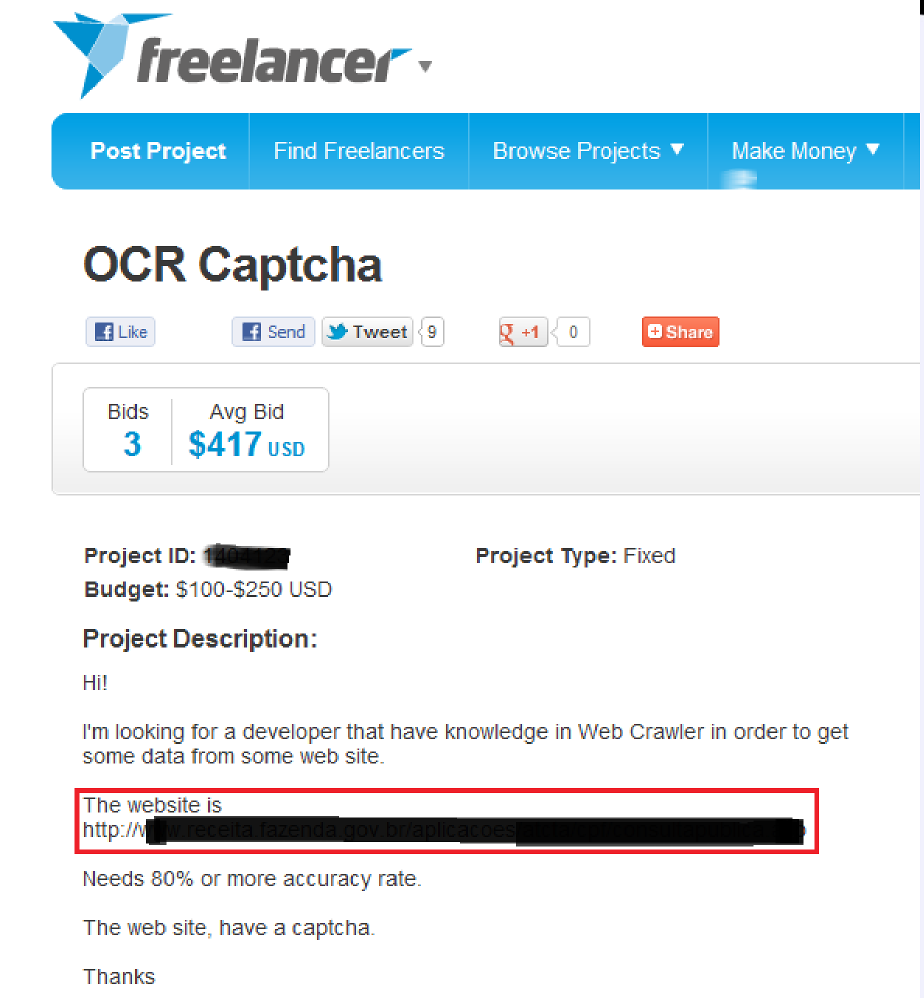
Nome da mãe:
□ não consta/em branco

Código:

Por favor, repita os caracteres acima:

Esta informação ajuda o Tribunal Superior Eleitoral a evitar a consulta por programas automáticos, que dificultam a utilização deste aplicativo pelos demais usuários.

Case Study: CAPTCHA by Passing



OCR Captcha

Bids 3

Avg Bid $417 USD

Project ID: [redacted]

Project Type: Fixed

Budget: $100-$250 USD

Project Description:

Hi!

I'm looking for a developer that have knowledge in Web Crawler in order to get some data from some web site.

The website is
http://[redacted]

Needs 80% or more accuracy rate.

The web site, have a captcha.

Thanks

Attack Characteristics

OWASP
The Open Web Application Security Project

Few attempts per CAPTCHA until solved
Over TOR for anonymity
Requests lacked proper headers

- User agent of known browsers
- But Accept headers were missing

CAPTCHA solving requests sent in a very high rate

Summary & Recommendations

Summary

OWASP
The Open Web Application Security Project

Automation is a major phenomena – used by both good and bad guys

CAPTCHA is a popular anti-automation tool but has caveats, and hackers are abusing them

Augment CAPTCHA with other anti-automation measures – traffic shape, traffic rate

Use community based anti-automation reputation service

Questions?