

ECE 498KL: eCrime and Internet Service Abuse

# CAPTCHAs

Kirill Levchenko

December 4, 2018

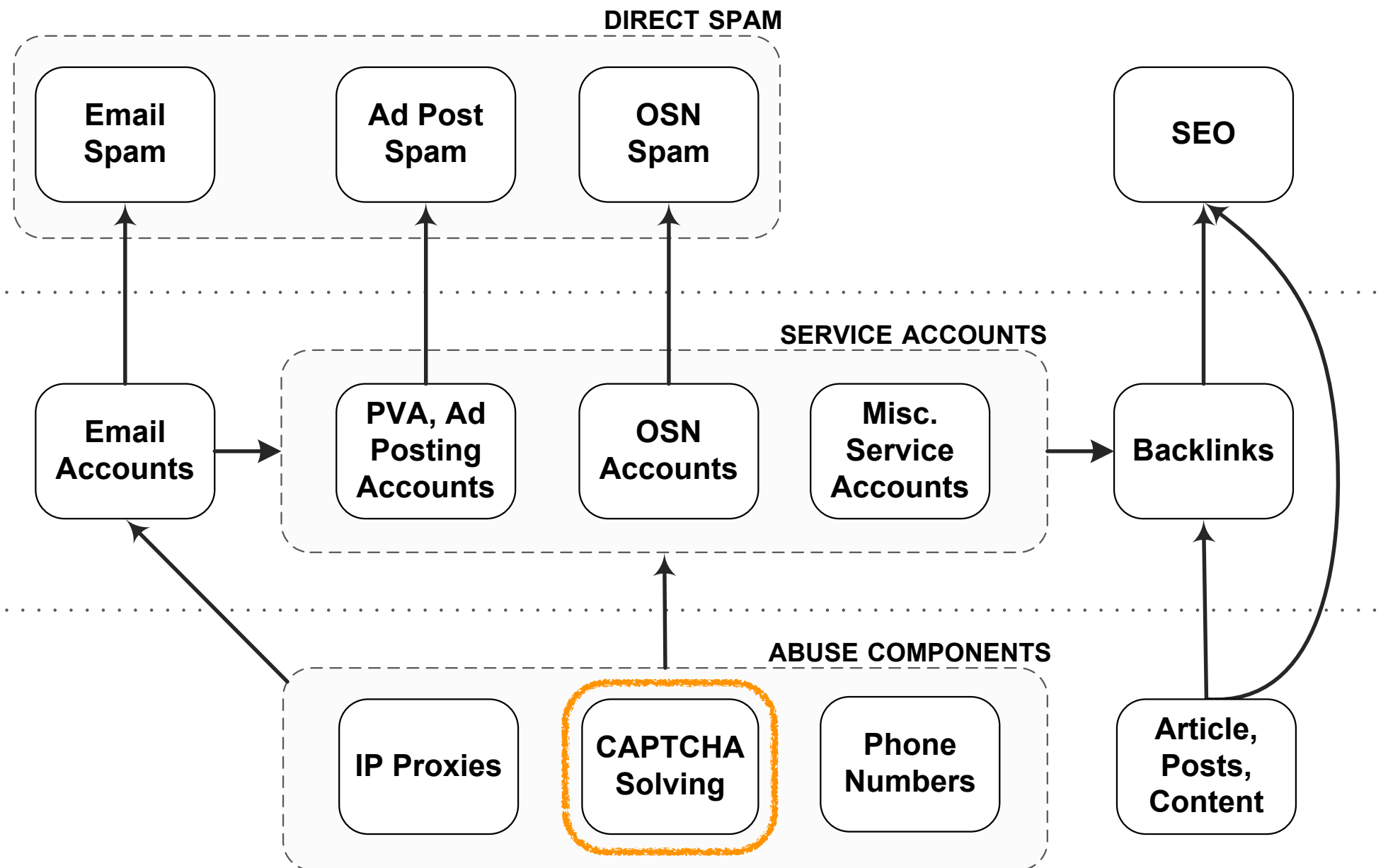
**I** ILLINOIS

Electrical & Computer Engineering

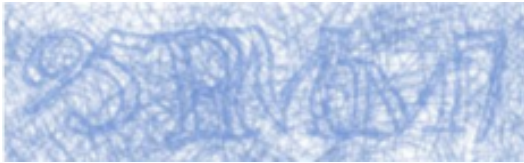
COLLEGE OF ENGINEERING

# Major Commodities

- ❖ **Web traffic** (categorized by topic and country origin)
  - **Topics:** gambling, pharma, adult, etc.
  - **Origin:** US, EU, Asia, Mix, etc.
  - **Sources:** SEO, spam, ads, bots, resale, etc.
  - **Monetization:** affiliate marketing, click fraud, resale
- ❖ **Installs** (categorized by target country)
  - **Target:** US, EU, Asia, Mix, etc.
  - **Sources:** compromise (Web, email, etc.), trojans, resale, etc.
  - **Monetization:** bots (e.g. spam), ransomware, fake AV, etc.

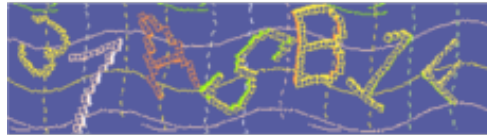


# CAPTCHAs



pursued

valle



Milwaukee- them

redcoats President

Select all squares that match the label:  
Sarah Connor.

If there are none, click skip.



SKIP

# CAPTCHA: Using Hard AI Problems for Security

Luis von Ahn<sup>1</sup>, Manuel Blum<sup>1</sup>, Nicholas J. Hopper<sup>1</sup>, and John Langford<sup>2</sup>

<sup>1</sup> Computer Science Dept., Carnegie Mellon University, Pittsburgh PA 15213, USA

<sup>2</sup> IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA

**Abstract.** We introduce CAPTCHA, an automated test that humans can pass, but current computer programs can't pass: any program that has high success over a CAPTCHA can be used to solve an unsolved Artificial Intelligence (AI) problem. We provide several novel constructions of CAPTCHAS. Since CAPTCHAS have many applications in practical security, our approach introduces a new class of hard problems that can be exploited for security purposes. Much like research in cryptography has had a positive impact on algorithms for factoring and discrete log, we hope that the use of hard AI problems for security purposes allows us to advance the field of Artificial Intelligence. We introduce two families of AI problems that can be used to construct CAPTCHAS and we show that solutions to such problems can be used for steganographic communication. CAPTCHAS based on these AI problem families, then, imply a win-win situation: either the problems remain unsolved and there is a way to differentiate humans from computers, or the problems are solved and there is a way to communicate covertly on some channels.

# 1 Introduction

A CAPTCHA is a program that can generate and grade tests that: (A) most humans can pass, but (B) current computer programs can't pass. Such a program can be used to differentiate humans from computers and has many applications for practical security, including (but not limited to):

- **Online Polls.** In November 1999, *slashdot.com* released an online poll asking which was the best graduate school in computer science (a dangerous question to ask over the web!). As is the case with most online polls, IP addresses of voters were recorded in order to prevent single users from voting more than once. However, students at Carnegie Mellon found a way to stuff the ballots by using programs that voted for CMU thousands of times. CMU's score started growing rapidly. The next day, students at MIT wrote their own voting program and the poll became a contest between voting "bots". MIT finished with 21,156 votes, Carnegie Mellon with 21,032 and every other school with less than 1,000. Can the result of any online poll be trusted? Not unless the poll requires that only humans can vote.

A CAPTCHA is a program that can generate and grade tests that: (A) most humans can pass, but (B) current computer programs can't pass. Such a program can be used to differentiate humans from computers and has many applications for practical security, including (but not limited to):





## CAPTCHA solving service

- ✓ Cheapest price on the market  
Starting from 0.5USD per 1000 images,  
depending on your daily upload volume
- ✓ Pay as you go

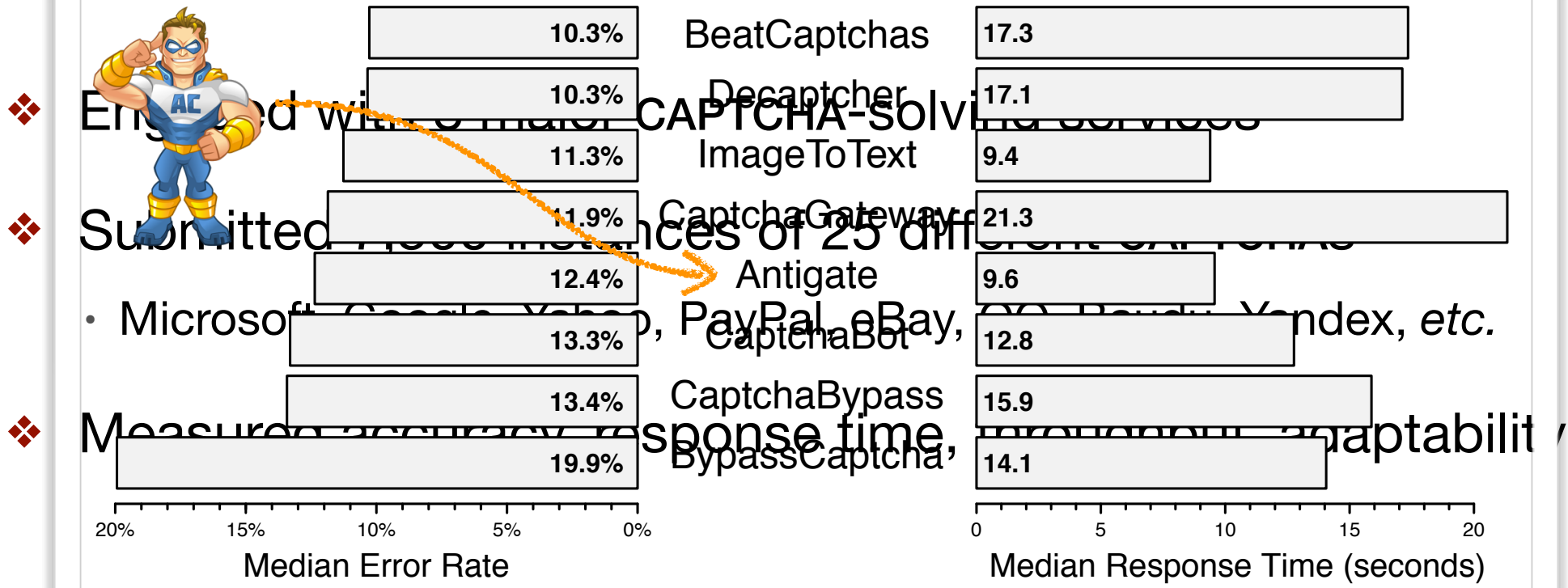
[registration](#)

✓ Cheapest price on the market  
Starting from 0.5USD per 1000 images,  
depending on your daily upload volume

Your app uploads a captcha  
to our server



# Re: CAPTCHAs



# Automated CAPTCHA Solving

- ❖ Can CAPTCHAs be solved automatically?
- ❖ Hard to automatically solve all possible CAPTCHAs
- ❖ May be possible to build solver for specific family

# Automated CAPTCHA Solving

- ❖ Xrumer 5.0.0 released in Oct 2008 with solvers for broad range of CAPTCHAs used in forums/blogs



**Xrumer 5.0** palladium

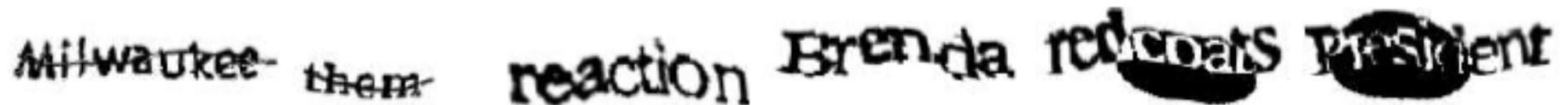
**Powerfull SEO software**

**Buy NOW or try FREE Demo**

**Only \$540**

The advertisement features a blue and black color scheme. At the top, the product name 'Xrumer 5.0' is displayed in white, with a blue logo to the left and the word 'palladium' in a smaller font to the right. Below this, the text 'Powerfull SEO software' is centered in white. Further down, the call to action 'Buy NOW or try FREE Demo' is written in white, and at the bottom left, the price 'Only \$540' is shown in white.

# Automated CAPTCHA Solving



Milwaukee them reaction Brenda redcoats President

Early 2008

Dec 16, 2009

Dec 24, 2009

- ❖ Reported to the public on Dec 15, 2009
  - Approx. 30% accuracy against old reCaptcha and 18% against current (at the time) reCaptcha
- ❖ Dec 16, 2009 automated solving rolled into popular Decaptcher.com service (at 25% normal price)
- ❖ Dec 24, 2009 reCaptcha changed to modern version (blobs): *Solver no longer effective.*

# Automated CAPTCHA Solving


- ❖ ***Solvers are fragile:***  
Easy to change CAPTCHA to break current generation of solvers that are tuned for specific family
- ❖ Cost of developing automated solver is high
  - Requires highly skilled labor and time
- ❖ How many does an automated solver have to solve to break even?

# CAPTCHA solving service



- ✓ **Cheapest price on the market**  
Starting from 0.5USD per 1000 images, depending on your daily upload volume
- ✓ **Pay as you go**  
Pay-per-captcha payment basis. Minimum refill is 1 USD, no recurring charges
- ✓ **99.99% uptime since 2007**  
Vast amount of workers and premium infrastructure allows us to provide highly reliable 24/7/365 service

[registration](#)

 [Client area](#)

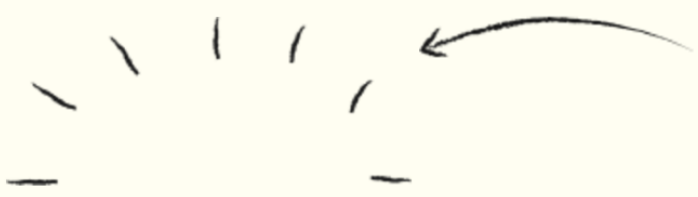
1

Your app uploads a captcha to our server



1

Your app uploads a captcha to our server



Easy [API](#) integration



Unlimited number of simultaneous uploads



High-speed request processing

2

We assign a worker for your captcha



100% of captchas are solved by human workers from around the world. This is why by using our service you help thousands of people to feed themselves and their families. [Check out some of their stories here.](#)

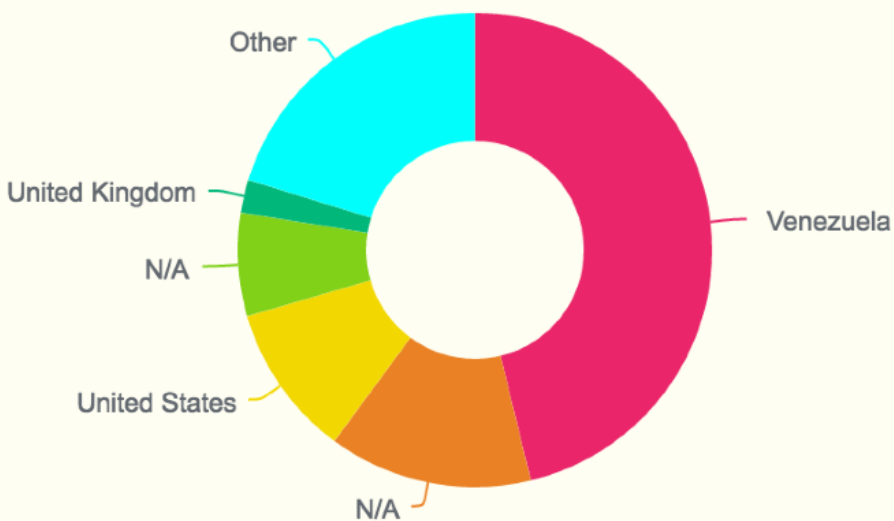
Easy [API](#) integration

Unlimited number of simultaneous uploads

High-speed request processing

2

## We assign a worker for your captcha



100% of captchas are solved by human workers from around the world. This is why by using our service you help thousands of people to feed themselves and their families.

[Check out some of their stories here.](#)

An average worker makes about \$100 per month which is a very good salary in such countries like India, Pakistan, Vietnam and others. With your help they now have a choice between working in polluted industries and working in front of a computer.



# captcha

100% of captchas are solved by human workers from around the world. This is why by using our service you help

th  
C  
A  
a  
V  
cl  
in

Priority: 0

Balance: 0.009

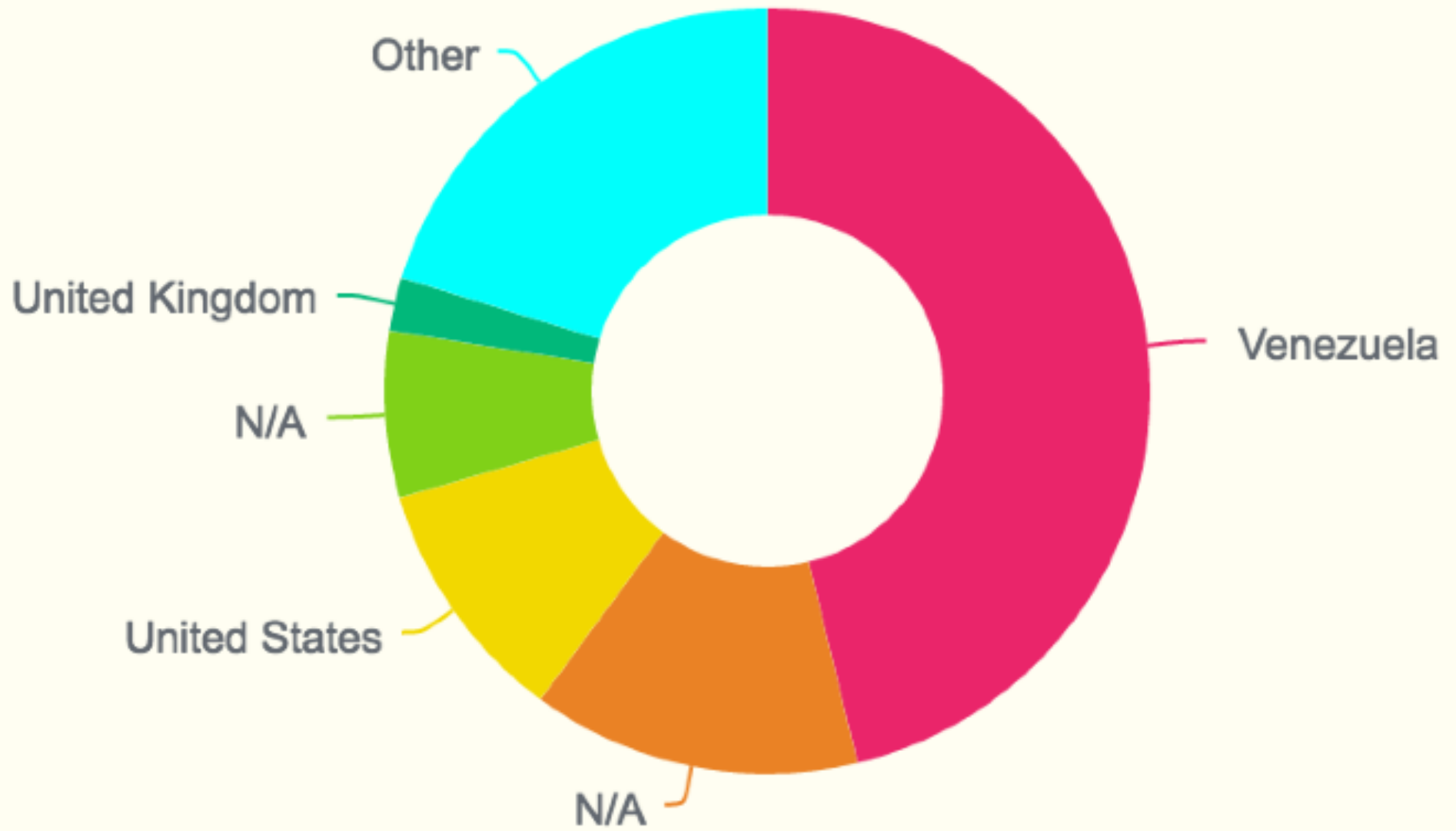


HUMAN

Send

Countdown: 27

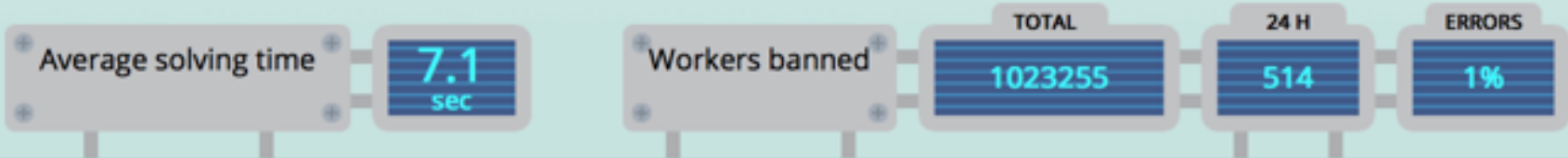
ca



10  
an  
th  
[Ch](#)  
An  
a  
Vi  
ch  
in

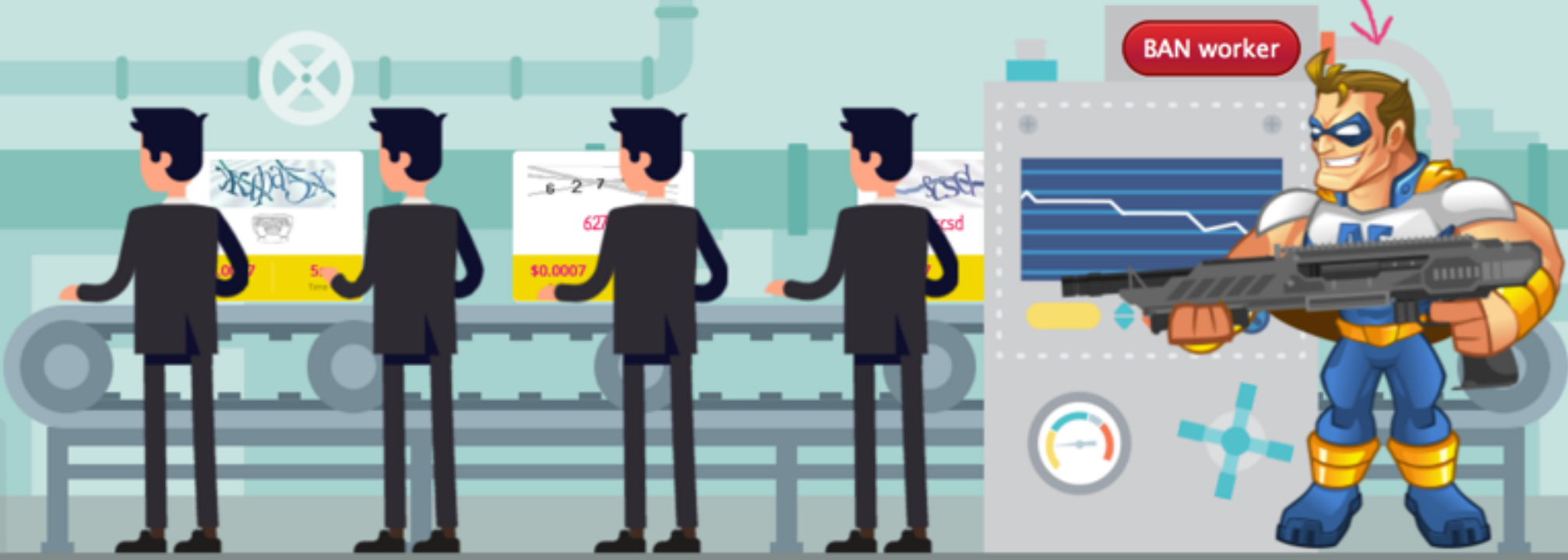
3

Worker types answer and we send it to your app



4

Our advanced quality control system monitors worker's entries and quickly eliminates cheaters.



## And by the way, we're good as hell in solving Javascript captchas!

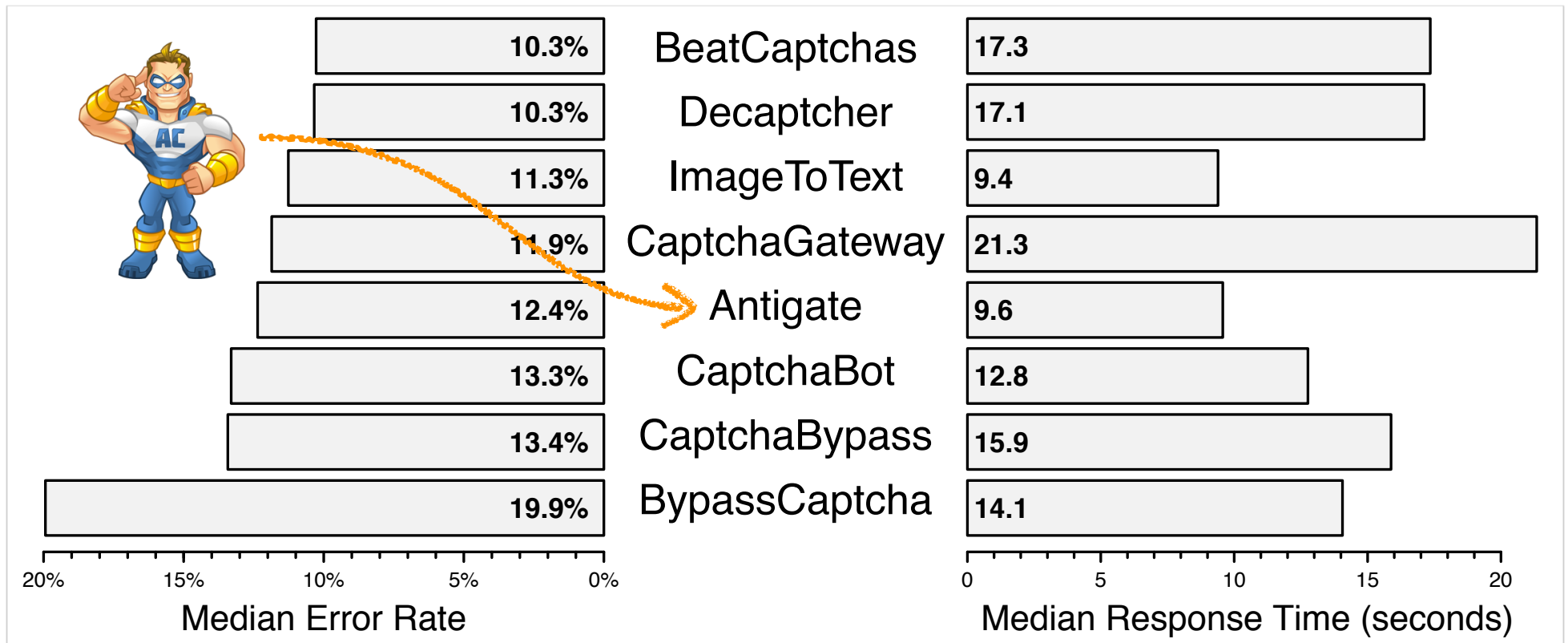
- Costs for Recaptcha: from 1.8 USD per 1000 solutions.
- Costs for Funcaptcha: from 1.8 USD per 1000 solutions.
- You don't need to emulate browser and run javascripts.
- You send us "sitekey" or "public key" value.
- We give you "g-recaptcha-response" and you simply submit form with it.



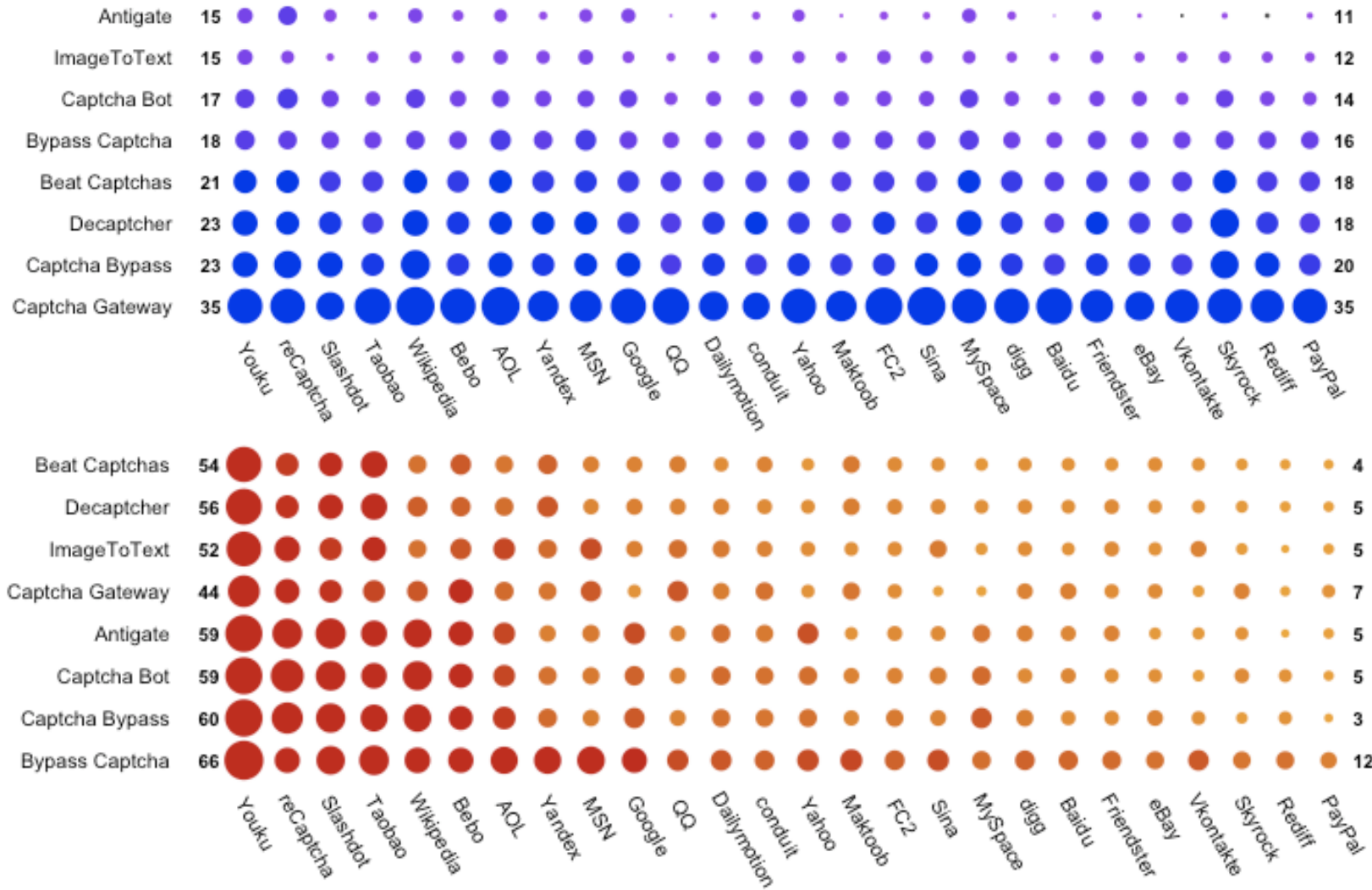
# CAPTCHA Study

- **Sign up as service customers (8 services)**
  - ♦ Pay for CAPTCHAs to be solved (26 distinct commercial CAPTCHAs, plus custom-designed challenges)
    - » Every 5 minutes for every service..
  - ♦ Use to probe behavior of service under different conditions
- **Sign up as laborers (2 “job sites” matched with service)**
  - ♦ Kolotibablo → Antigat
  - ♦ Pixprofit → Decaptcher
  - ♦ Monitor which CAPTCHAs asked to solve (our own CAPTCHAs “tagged” to allow easy identification)
- Let run for months...

# Accuracy and Latency



# CAPTCHA Types



Response Time

Error Rate

# Capacity

- Incrementally increased load (32-1536 threads)
  - ◊ Each thread submitted new CAPTCHA after old one solved
- Unable to max Antigate out (41 CAPTCHAs/sec)
- Other services
  - ◊ Decaptcher, Captchabot: max at 14-15/sec
  - ◊ BeatCatpchas: max at 8/sec
  - ◊ BypassCaptchas: max at 4/sec
- If we assume 10-13sec/CAPTCHA (and no other users)
  - ◊ Antigate has **at least** 400-500 workers
  - ◊ Decaptcher/Captchabot: 140-200 workers
  - ◊ BeatCaptchas: 80-100 workers
  - ◊ BypassCaptchas: 40-50 workers



# Demographics

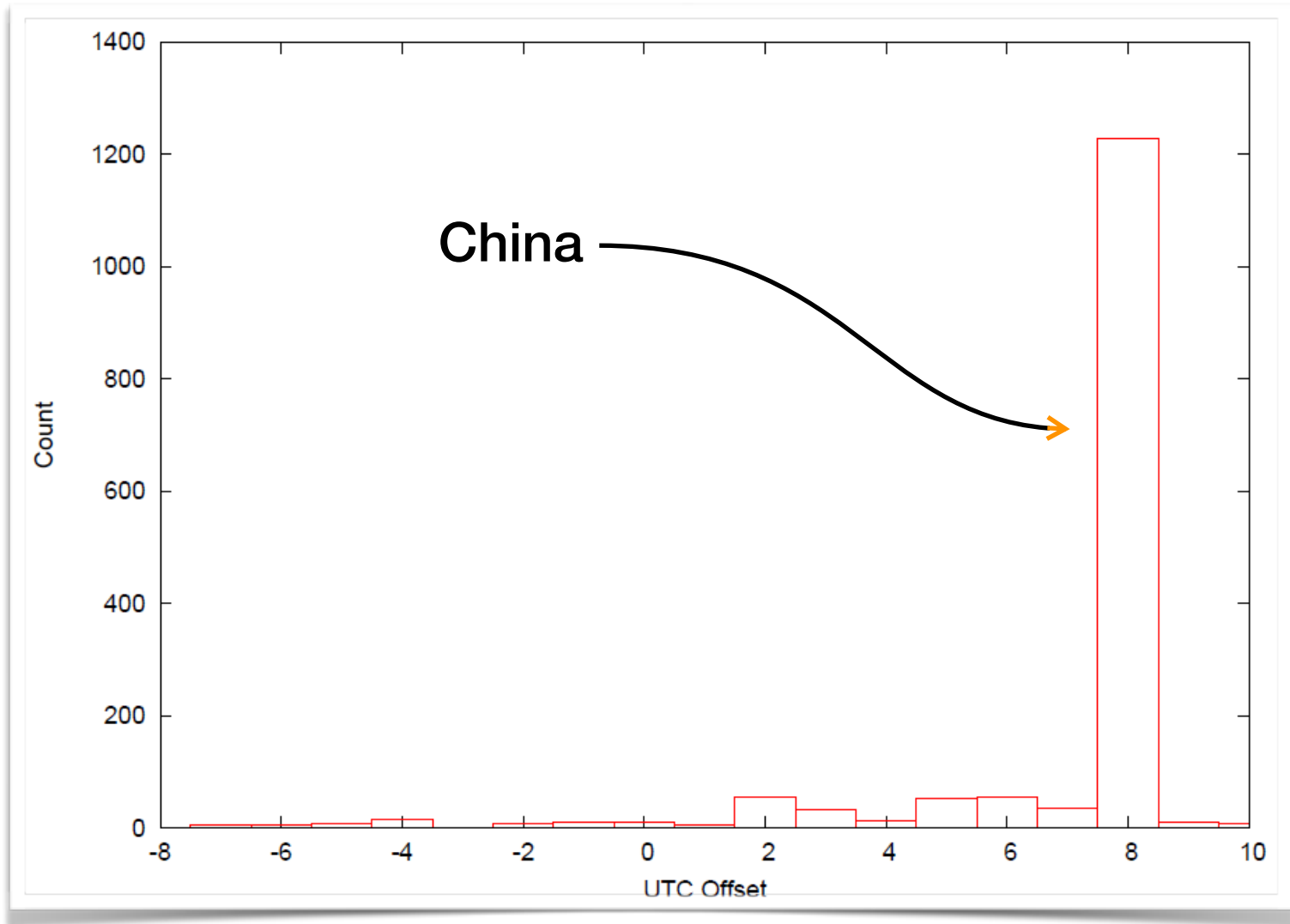
- Which labor markets are driving CAPTCHA-solving?
- Idea: get CAPTCHA solvers to reveal information about their country or location
- Two approaches
  - ◆ Language CAPTCHA: asks for word to digit translation in 20+ languages
  - ◆ Local time CAPTCHA: asks for current time in 14 languages

Введите показанные цифры

четыре	ноль	восемь	девять
4	?	?	?



# Time Zone



# Adaptability

Find all cats • 请找所有猫 • Найдите кошек • बिल्लियाँ ढूँढिए



A



B



C



D



E



F



G



H



I



J



K



L

# CAPTCHA Reality

- ❖ **Wanted:** Prevent automated access to services
- ❖ **Assumption:** Making users solve CAPTCHAS prevents automated abuse
- ❖ **Found:** Attackers uses *humans* to solve CAPTCHAS
  - Rest of abuse workflow remains automated

# CAPTCHA Insight

- ❖ **Are CAPTCHAs broken?**
- ❖ **No? CAPTCHAs hard to solve in the fully general case**
  - But a *concrete family* can be solved automatically
  - Easy for CAPTCHA producer to modify family and break solver
  - Automated solvers “waste of time” (according to one service operator)
- ❖ **Yes? Do *not* prevent automated abuse!**
  - CAPTCHAs solved by human labor for automated tools

# CAPTCHA Insight

- ❖ **Insight:** CAPTCHAs introduce additional **cost** to attacker
  - 1/10 cent to bypass a CAPTCHA
- ❖ **Insight:** Deters rational profit-motivated attacker when (cost of solving captcha) > (expected revenue)
- ❖ **Eliminates nuisance attacks**
  - Attacker business model must support added cost