



Breaking E-bay audio captcha

Elie Bursztein Steven Bethard
Stanford University



- Breaking an audio captcha
- E-bay audio captcha
- What is next ?

Bad or good user ?



Bad or good user ?



Bad or good user ?

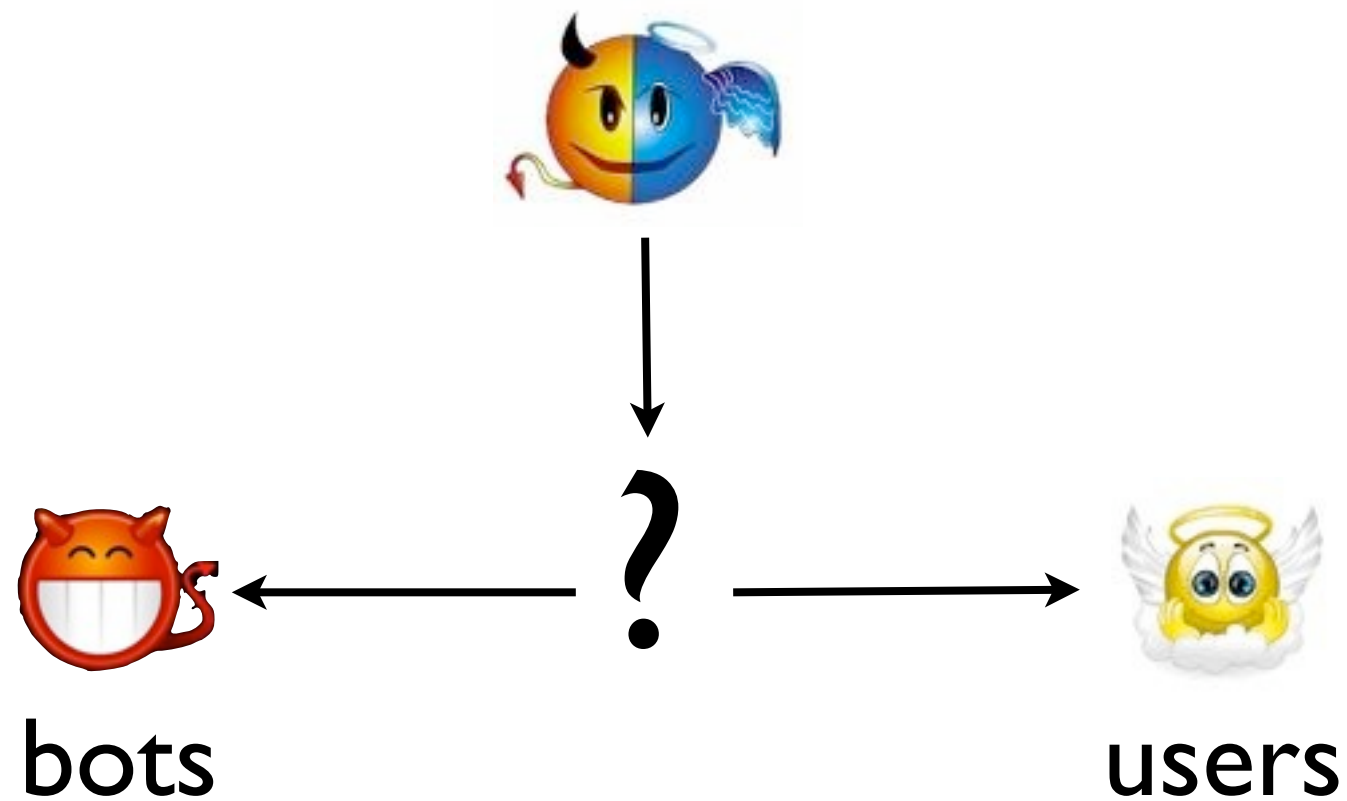


?



users

Bad or good user ?



CAPTCHA



following

finding



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

6ne3

prosper

xmqKi

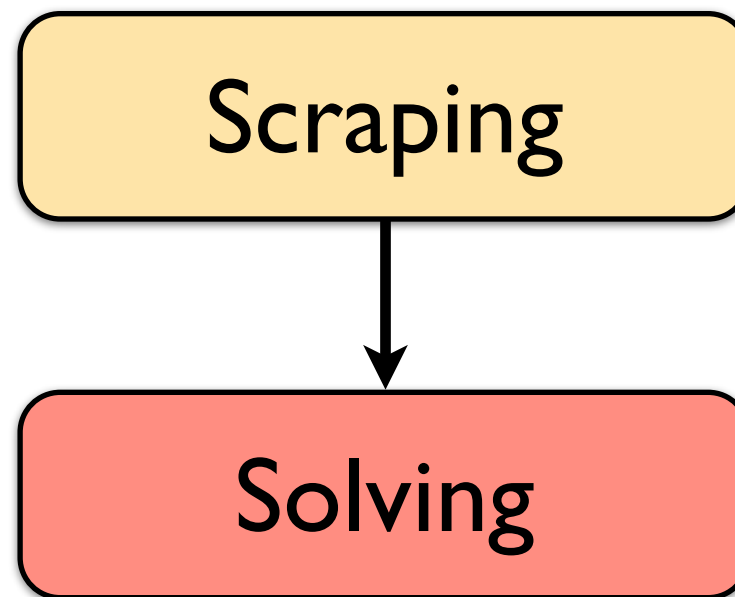
Term coined in 2000 [VBHL]

Breaking the registration process

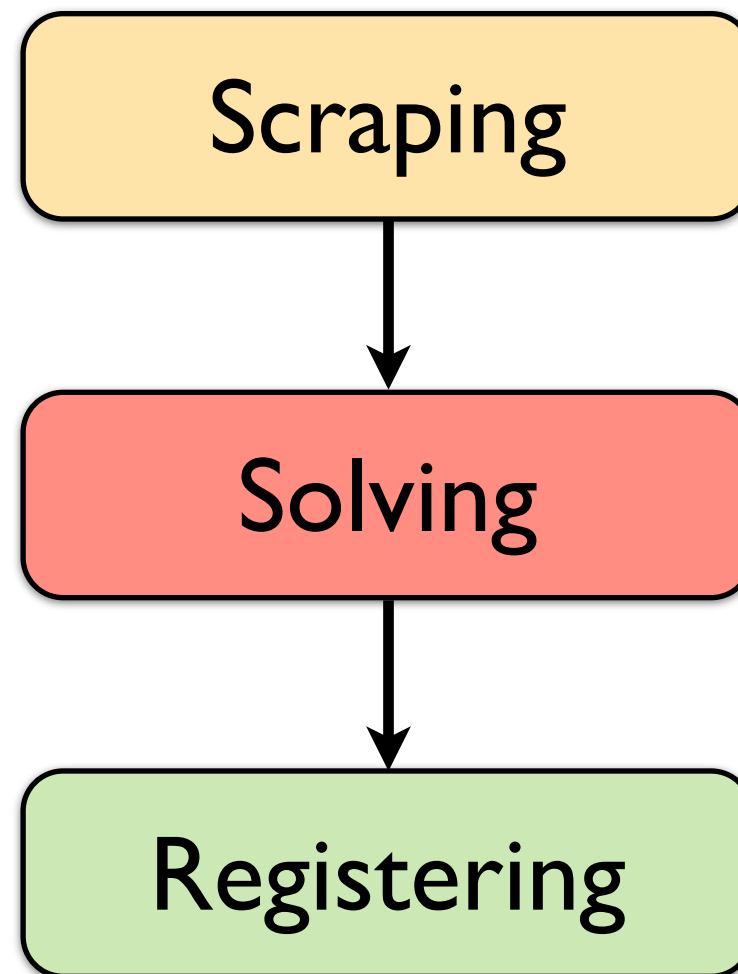


Scraping

Breaking the registration process



Breaking the registration process

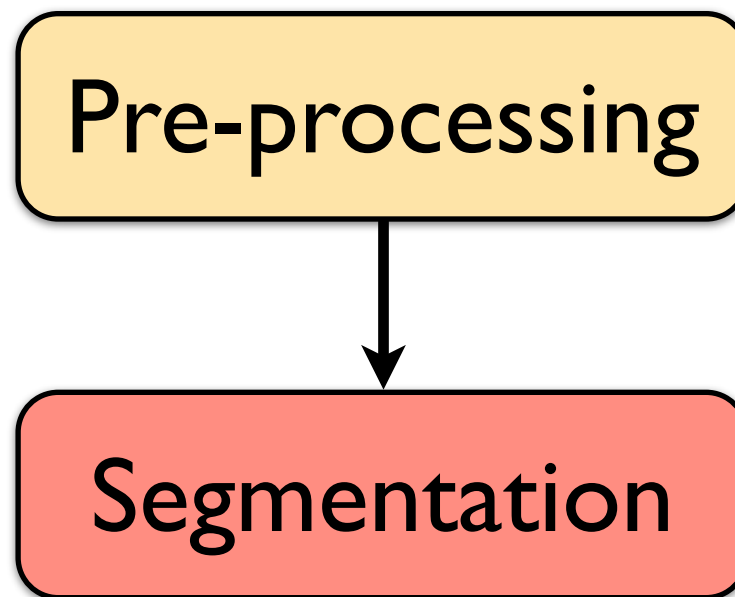


Breaking captcha

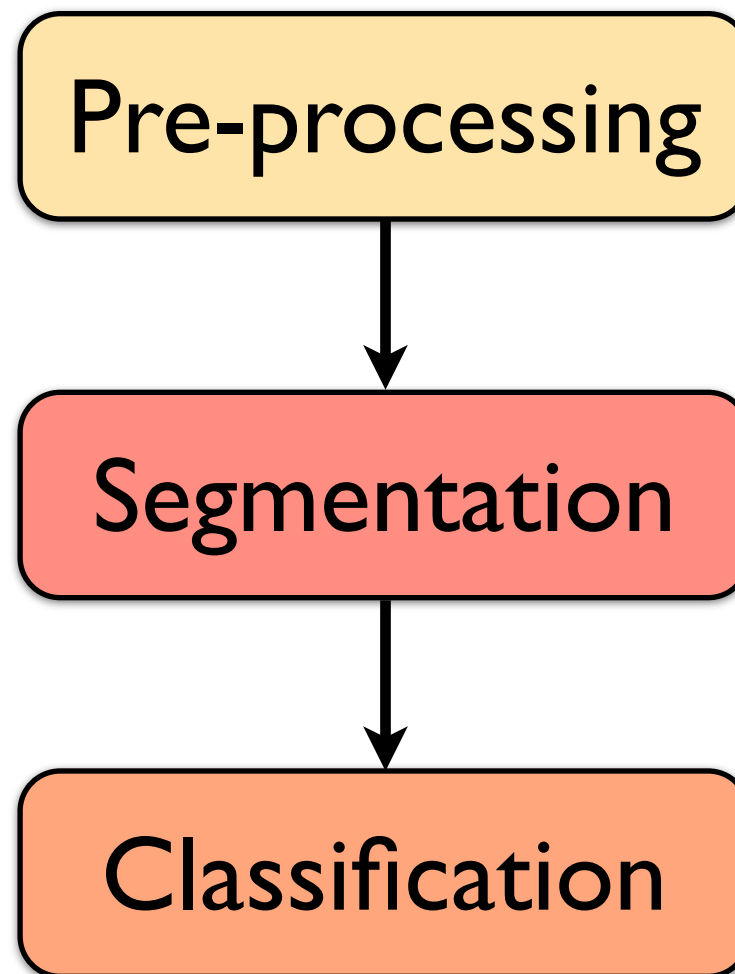


Pre-processing

Breaking captcha



Breaking captcha





- Coverage
- Precision
- Accuracy

$$\text{Accuracy} = \text{Coverage} \times \text{Precision}$$

Catpcha goal = 0.01% accuracy [CLC05]





Breaking E-Bay captcha

The registration page



Choose your user ID and password - All fields are required

Create your eBay user ID
 [Check availability](#)
Use letters or numbers, but not symbols. [Learn more about creating great user IDs.](#)

Create your password

case sensitive. [Learn about secure passwords.](#)

Re-enter your password

Pick a secret question
Select your secret question...


Your secret answer

If you forget your password, we'll verify your identity with your secret question

Date of birth
--Month-- --Day-- Year YYYY
You must be at least 18 years old to use eBay.

The not-so-fine print

For added security, please enter the verification code hidden in the image.



[Refresh the image](#) | [Listen to the verification code](#)

☐ I agree that:

- I accept the [User Agreement](#) and [Privacy Policy](#).
- I may receive communications from eBay and can change my notification preferences in My eBay.
- I'm at least 18 years old.

[Continue](#)

[Continue](#)

• I'm at least 18 years old.
• I may receive communications from eBay and can change my notification preferences in My eBay.
• I accept the User Agreement and Privacy Policy.

Audio Security Measure

Click the "Play audio" link and listen to the verification code. Type the code you hear into the box below, then click the **Submit** button.
If you have trouble hearing the verification code, press the "Play audio" link again.

[Play audio](#)

Enter verification code:

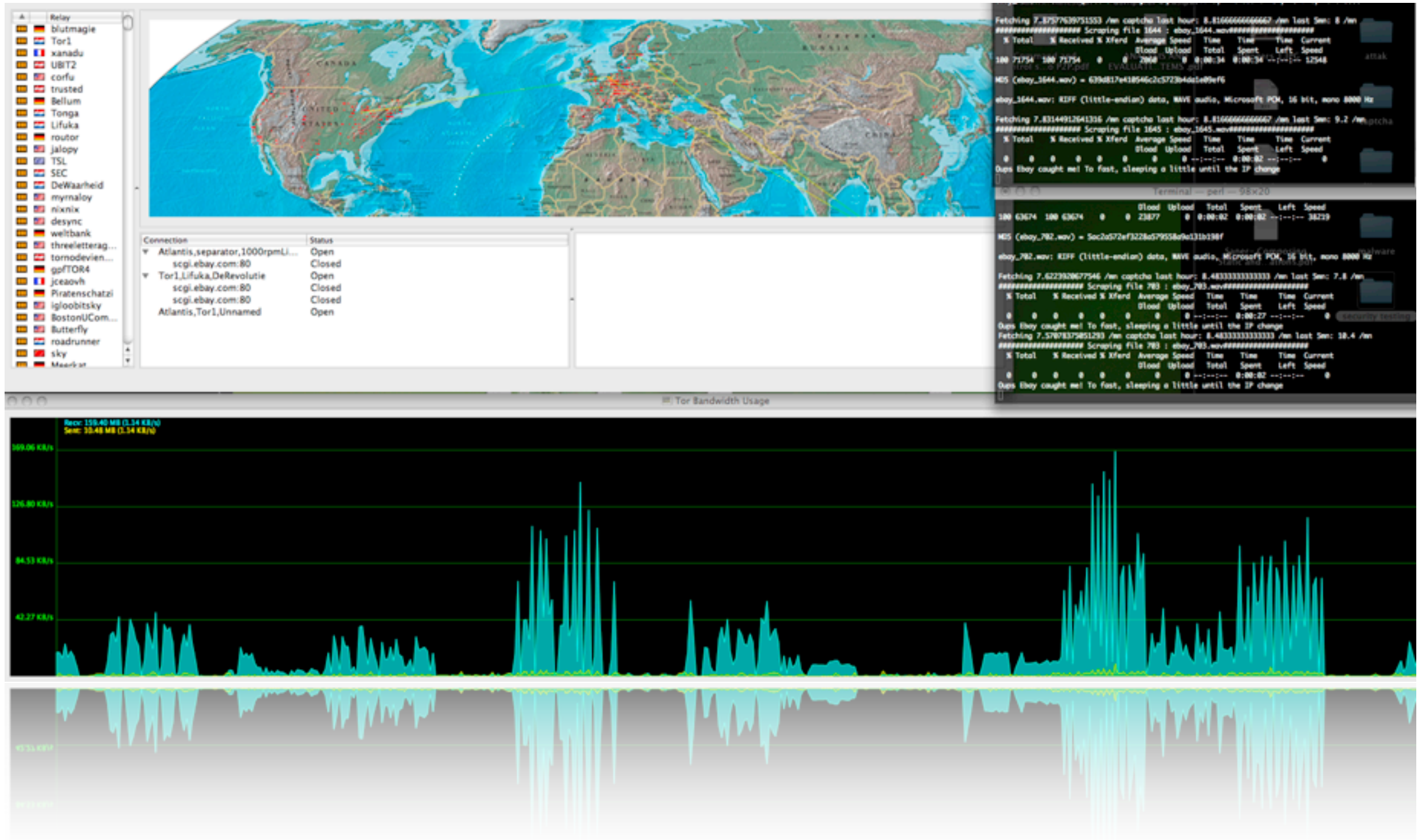
[Submit](#) | [Cancel](#)

[Previous](#) | [Next](#)

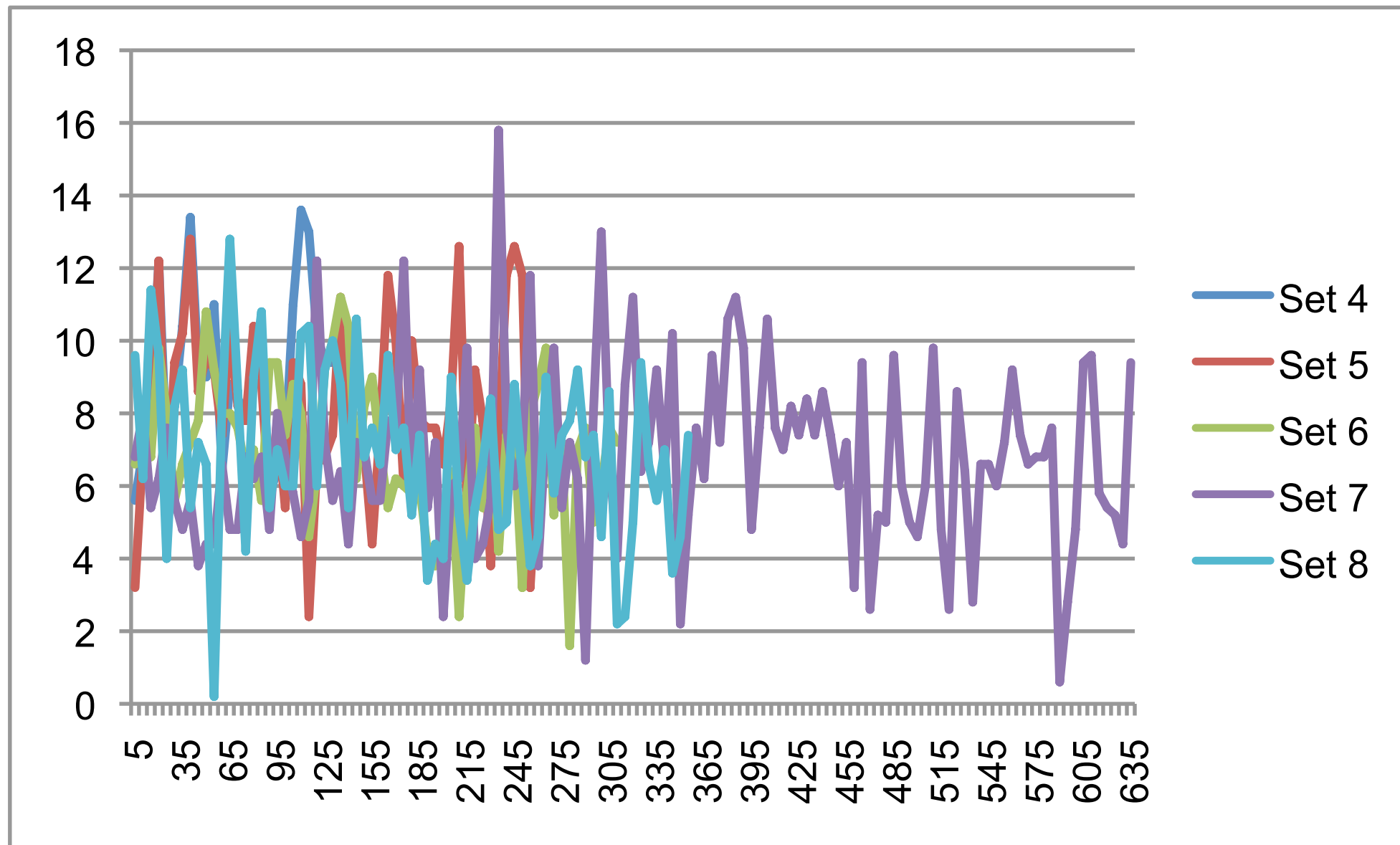


- Allow to download/try multiple time the same captcha
- Limit poorly the number of captcha download

Fetching performance



Fetching performance



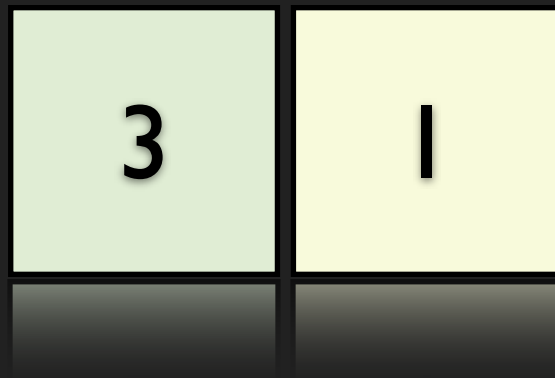
E-Bay captcha



3

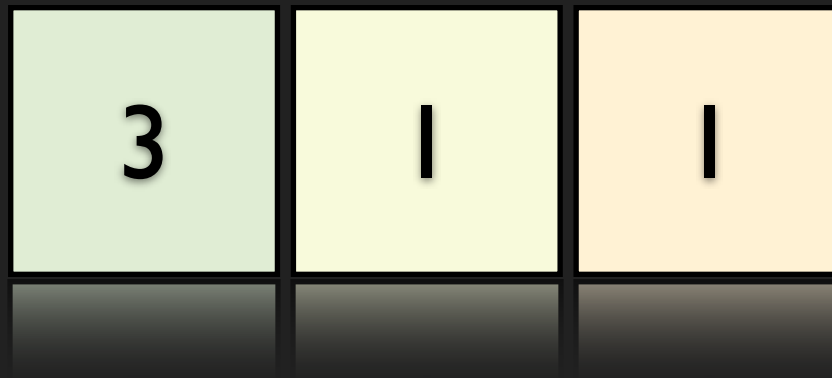
E-Bay captcha





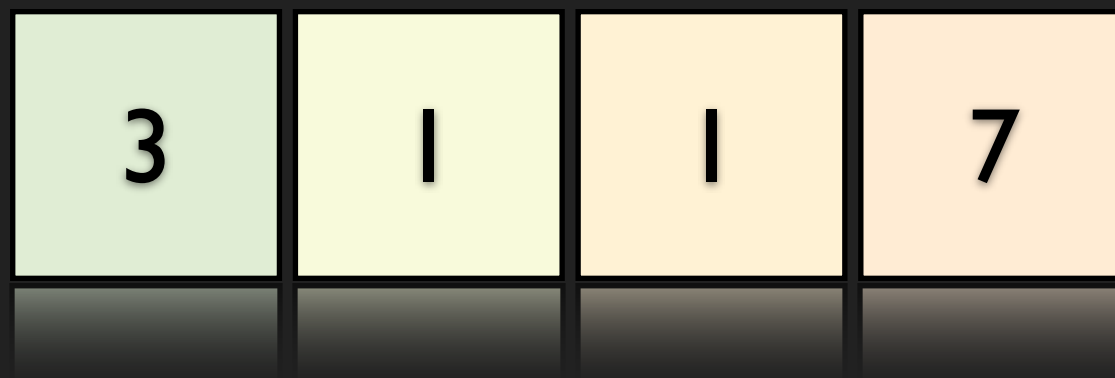
E-Bay captcha





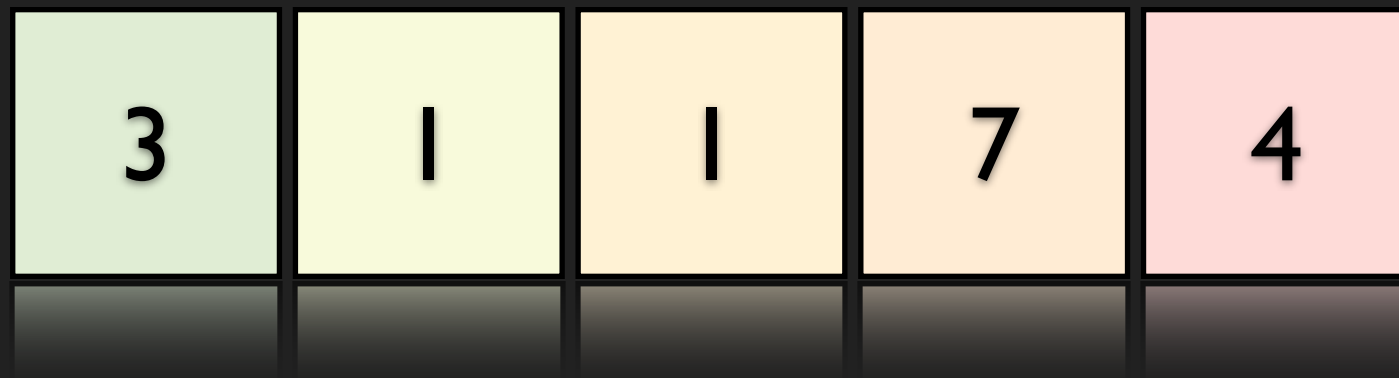
E-Bay captcha





E-Bay captcha





E-Bay captcha





E-Bay captcha



Captcha implementation



- Fixed number of digit
- Multiple downloads with different voices



- We downloaded **26 000** captchas overall
 - 8x 2000 samples (training set)
 - 50 x 200 samples (evaluation set)

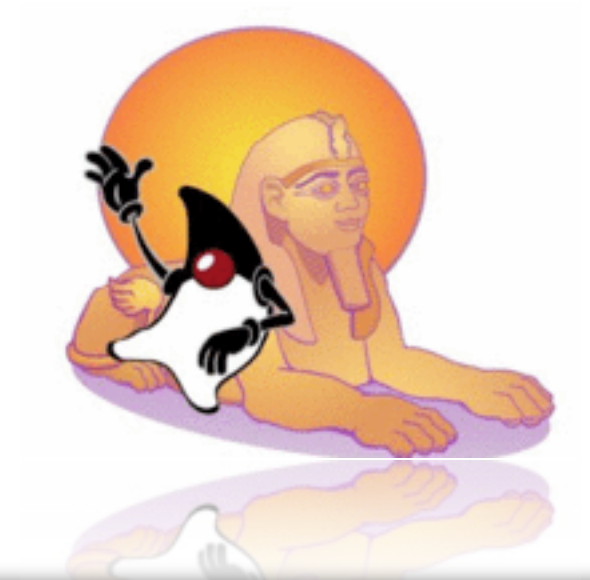
Building a breaker



- Two options
 - Use state of art speech recognizer
 - Build a specialized classifier



- State of art speech recognizer
- Available models
 - HUB4 : 64 000 words model
 - Isolated digits
 - Connected digits



Attempt 1 result



Model	Acc	Prec	Cov
TIDGITS	0.3%	9.6%	3.6%
HUB4	1%	28.9%	3.6%

Decaptcha: The audio captcher !



Decaptcha online version

About decaptcha

Decaptcha is a tool designed to break audio CAPTCHAs (we call this action decapting :)). It does so by performing discrete Fourier transformations and analyzing energy spikes. The current version can break E-bay captchas with 80% accuracy (Based on 10000 tests).

Team

The following people contributed to this project :

- Elie Burszein (Stanford University)
- Steven Bethard (Stanford University)
- Celine Fabry

Try it yourself

If you don't believe us that decaptcha is that accurate, which is totally understandable, then what about testing it ? Just submit below an E-Bay audio captcha and shortly you will have it decaptched.

Select the captcha you want to decaptch:

Select the captcha you want to decaptch:


If you don't believe us that decaptcha is that accurate, which is totally understandable, then what about testing it ? Just submit below an E-Bay audio captcha and shortly you will have it decaptched.



Attempt 2 : build our own classifier

Buggy / Ugly C code !

The audio captcher !



captcha online version

About decaptcha

Decaptcha is tool designed to break audio CAPTCHAs (we call this action decapting :)). It does so by performing discret Fourier transformations and analyzing energy spikes. The current version can break E-bay captchas with 80% accuracy (Based on 10000 tests).

Team

The following people contributed to this project :

- Elie Burszein (Stanford University)
- Steven Bethard (Stanford University)
- Celine Fabry

Try it yourself

If you don't believe us that decaptcha is that accurate, which is totally understable, then what about testing it ? Just submit below an E-Bay audio captcha and shortly you will have it decaptched.

Select the captcha you want to decaptch:

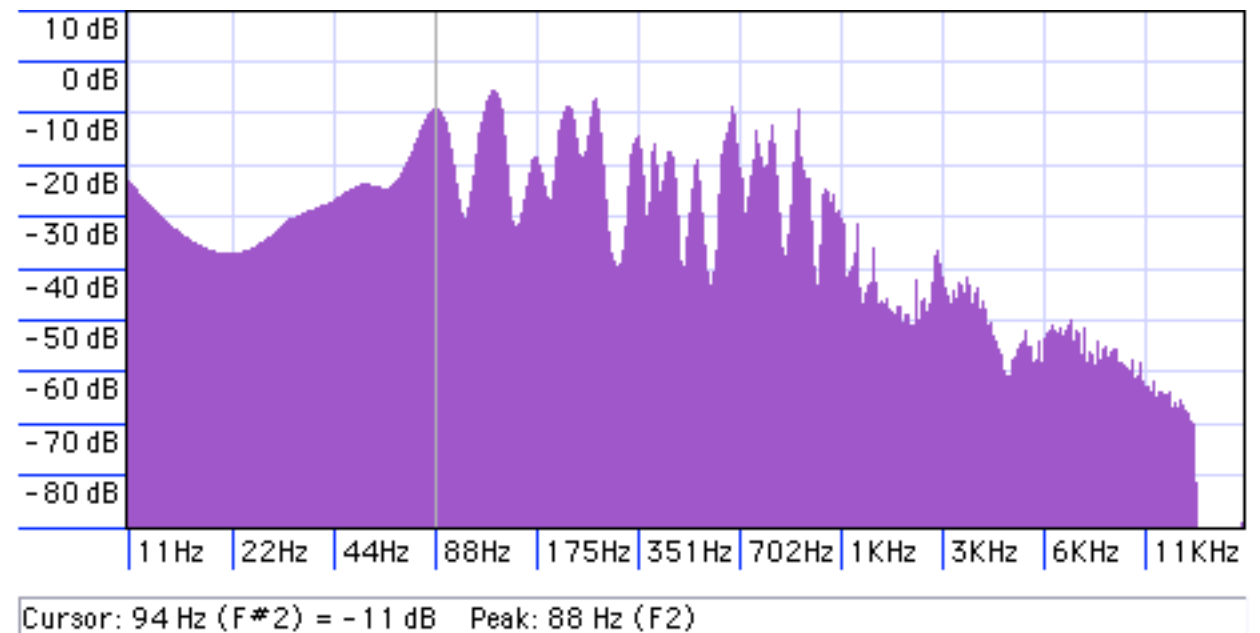
Attempt 2 : build our own classifier



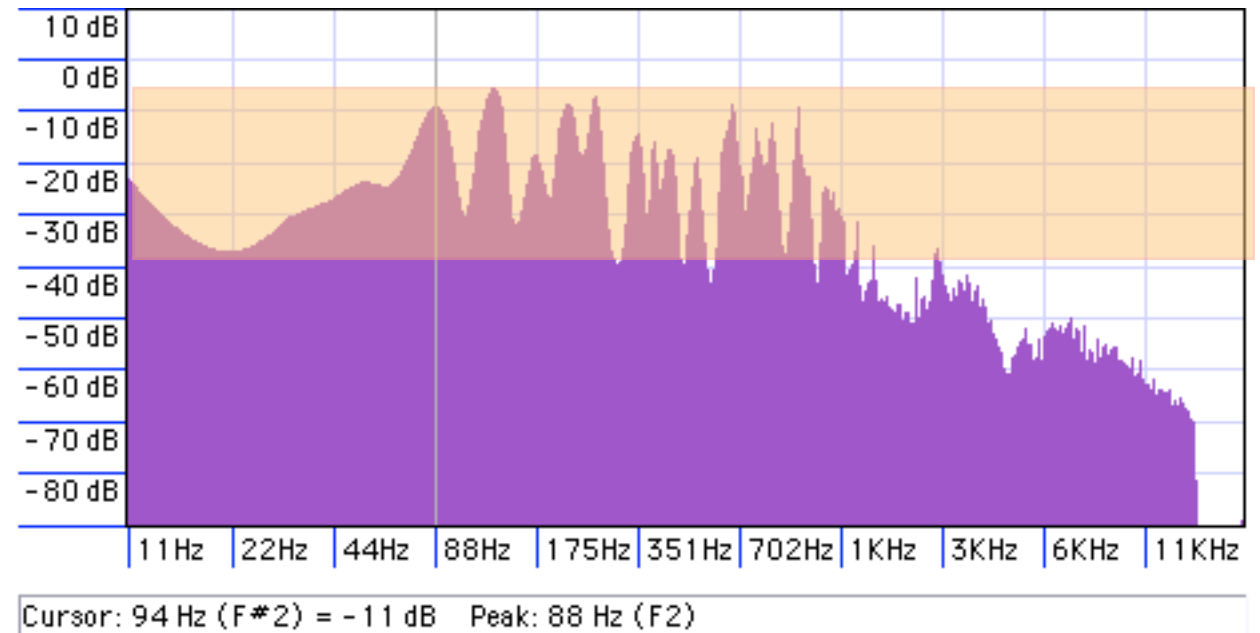
How Decaptcha works



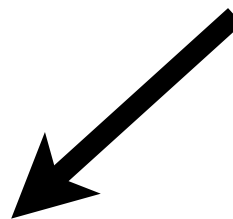
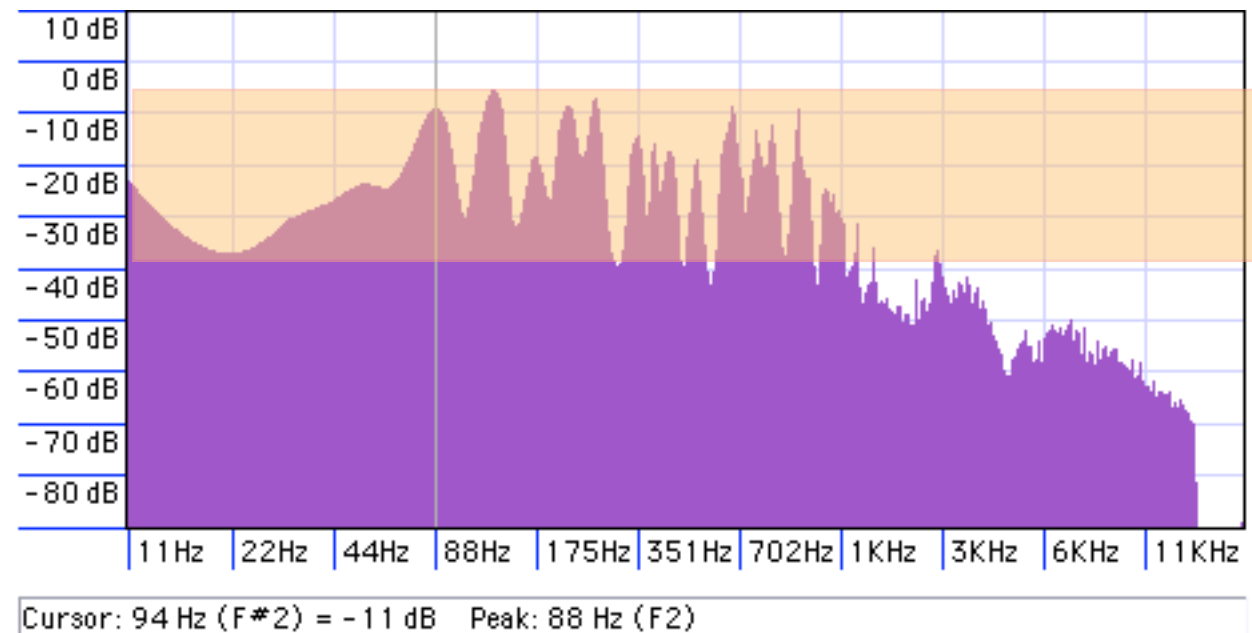
How Decaptcha works



How Decaptcha works



How Decaptcha works



0:
26645 67191 119061 89244 54427 65733 123753 34290 228898 99908 201288 241704 187527
187323 355112 131920 67751 152565 129169 143524 76310 76563 202572 143785 225241
711726 817683 1044637
146120 479933 350394 942188 209828 270356 279296 120812 268639 140728 98185 223023
269431 277125 236198 246403 132344 289054 102437 105509 212886 295669 197393 318710
431137 785228 843192 3190283
24248 399464 830038 880644 111876 760679 469023 78097 297356 146268 298207 257063
556285 606488 352779 159524 229851 164489 102580 201637 81613 112075 249878 226535
477520 866948 902598 4626601
148464 208997 592865 1087574 74635 509526 382653 238383 202327 243430 268997 606916
301822 358115 208761 111799 119561 142713 172740 202756 76463 66374 146260 274962
246128 982006 1654120 5740142
45956 258133 1005985 1098416 57227 828851 1339151 314516 513288 179292 550734 519032
799286 231594 498671 342832 202360 197459 185155 315956 209138 234810 160295 300425
329821 1353146 2571746 4564688
81712 161237 417391 506261 185897 119133 338142 140221 337865 401714 465674 350169
270428 365631 253015 139859 124894 180852 114128 147683 100545 138302 202106 381131
388385 747495 1390303 3786484

Result



Model	Acc	Prec	Cov
3x100	53.4%	65.7%	81.3%
3x500	59.6%	73.3%	81.3%
6x100	65.6%	80.7%	81.3%
8x1000	75.1%	92.3%	81.3%

More captcha or more samples ?



	Accuracy	Precision
10x50	44%	55%
25x20	76%	95%
50x10	72%	90%



What's next






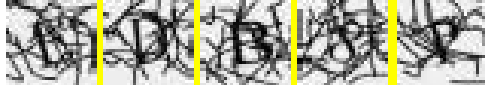


- What is **difficult** to computer ?
- What is **easy** for human ?
- Difference between **visual** and **audio** captcha ?

Visual captcha Hardness



- What is hard for a computer ?
- Decomposition is hard
- Recognition is easy

Characters under typical distortions	Recognition rate
	~100%
	96+%
	100%
	98%
	~100%
	95+%

	82+%
--	------

Decomposition hardness



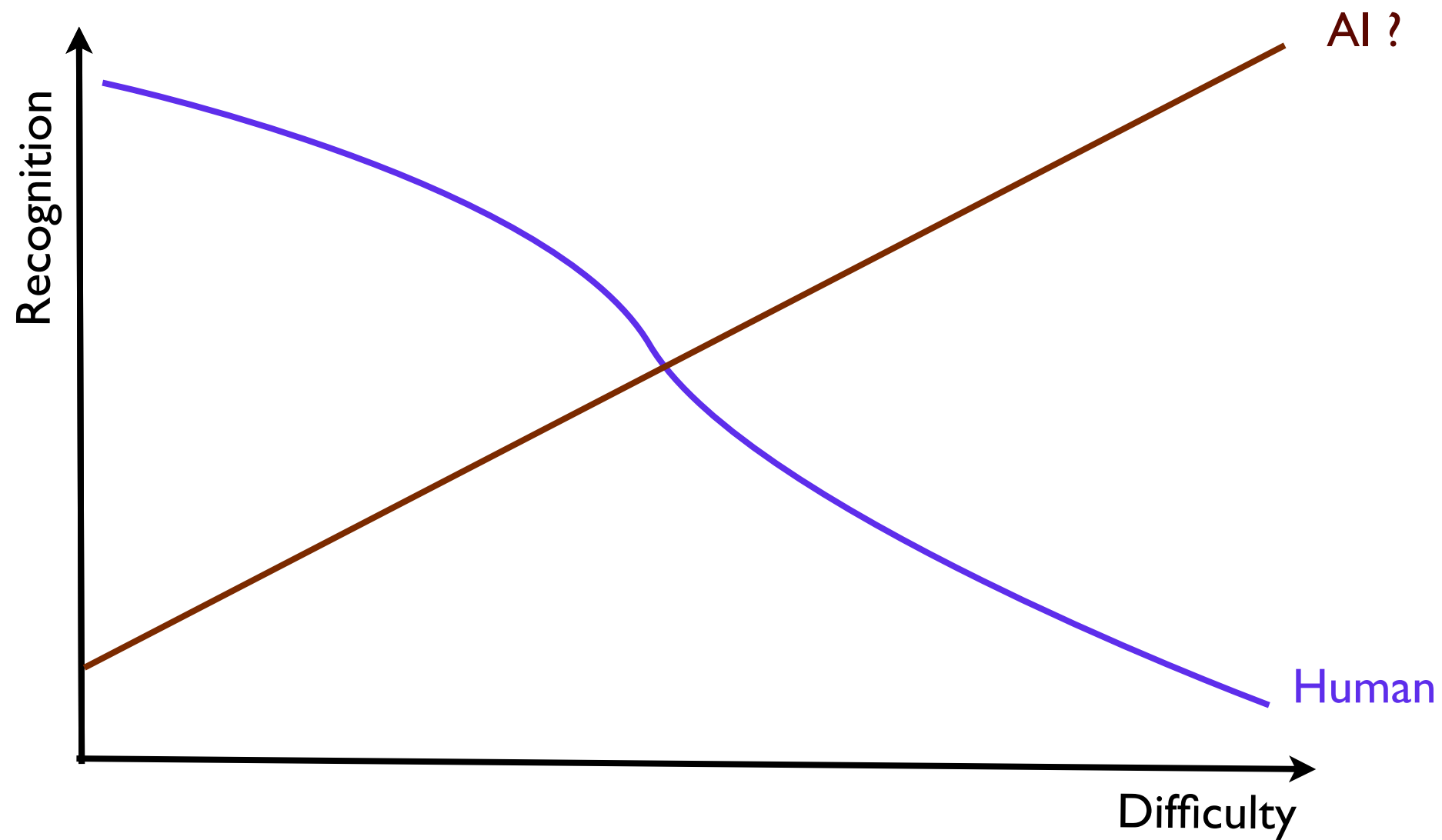
- Separation should be harder for some type
 - Separated digits
 - Letters
 - Numbers
 - Words
 - Continuous speech

Recognition hardness



- Introduce noise
- Introduce distortion
- Larger corpus

What about user ?



Visual vs Audio



- It is all about **data pool** :
 - **Infinite** for visual captcha
 - **finite** for audio captcha



- Things we are currently doing
 - Building a generic attack framework
 - Defining a security metric
 - Studying usability
 - Breaking more captcha scheme :)
- Things we hope to do
 - Find a new way to do audio captcha
 - Going after visual captcha

Thanks / questions



- Thanks to
 - Jerry Louis and the E-bay security team
 - Celine Fabry
- Questions ?

If you want to try decaptcha

<http://www.dontrythisathome.com/decaptcha/>