Human Based Computation

how crowd-sourcing can solve some of the tricky security problems

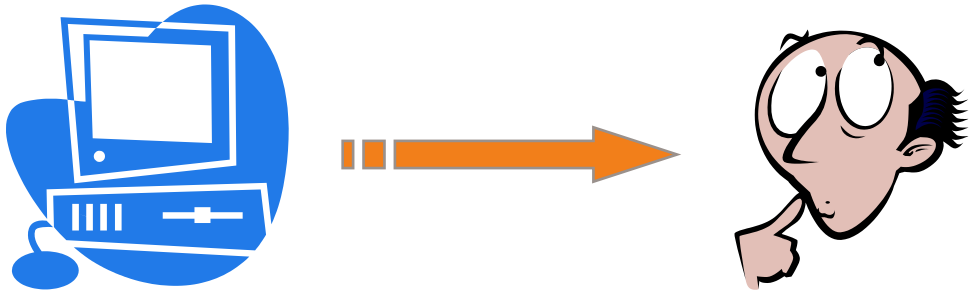
Sumesh Jaiswal

Security Technology and Response Group, Symantec

What is Human Based Computation?

What is HBC?

- HBC is a class of hybrid techniques in which computers outsource certain functions to humans



- Ask a machine to point to a picture of a bird and it usually fails. But even the most dim-witted human can do this easily

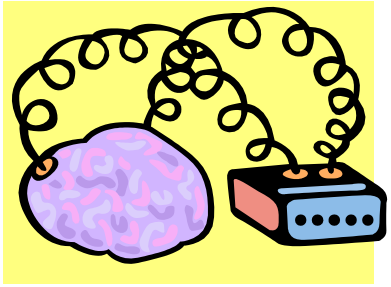
Because for certain tasks the Cortex still beats the CPU

What is HBC?



Tasks like image or voice recognition are trivial for humans, but continue to challenge even the most sophisticated computer programs

HBC allows the computer to take help from humans to solve such challenging parts of the problem



What is HBC?

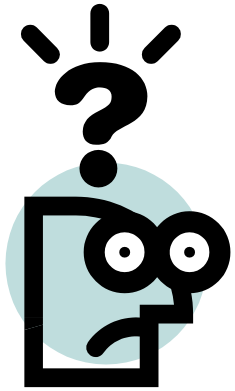
This is HUMAN COMPUTATION

the art of using massive groups of networked human minds to solve problems that computers cannot

symantec™

Confidence in a connected world.

Is HBC used for real?



Is HBC used for real?

CAPTCHA

Completely Automatic Public Turing test to
Tell Computers and Humans Apart

Is HBC used for real?

CAPTCHA

Completely Automatic Public Turing test to
Tell Computers and Humans Apart

- Captchas are automated tests that humans can pass but computers cannot
- They take advantage of human processing power in order to differentiate humans from computers
- Developed by **Luis von Ahn** (CMU) in 2000 to thwart spam-bots

Is HBC used for real?

So did Captcha deter spammers?

NO

CAPTCHA Sweat Shops

Spam companies hire humans to solve captchas **all day long**

$2.50 per hour for each human

720 captchas per hour per human

1/3 cent per email account

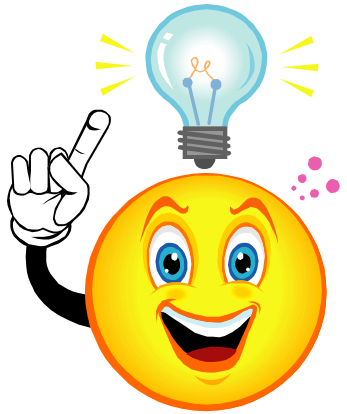
Now THIS is human computation!!

Is HBC used for real?

But atleast there are two consolations

First, it costs them some money

Second, Captchas are actually generating jobs in under-developed countries!



Some statistics about HBC

9 billion human-hours of Solitaire were played in 2003

- we talk about wasted CPU cycles
- what about wasted human cycles

Compared to that

- it took 7 million hours to build the **Empire State Building**
- which is just 6.8 hours of playing solitaire

Human computation

- making good use of these wasted human cycles

Making good use of human processing power

Human Processing Power

- Labelling images with words

Input: an arbitrary image

Output: a set of keywords that appropriately describe the image

Man

Indiana Jones

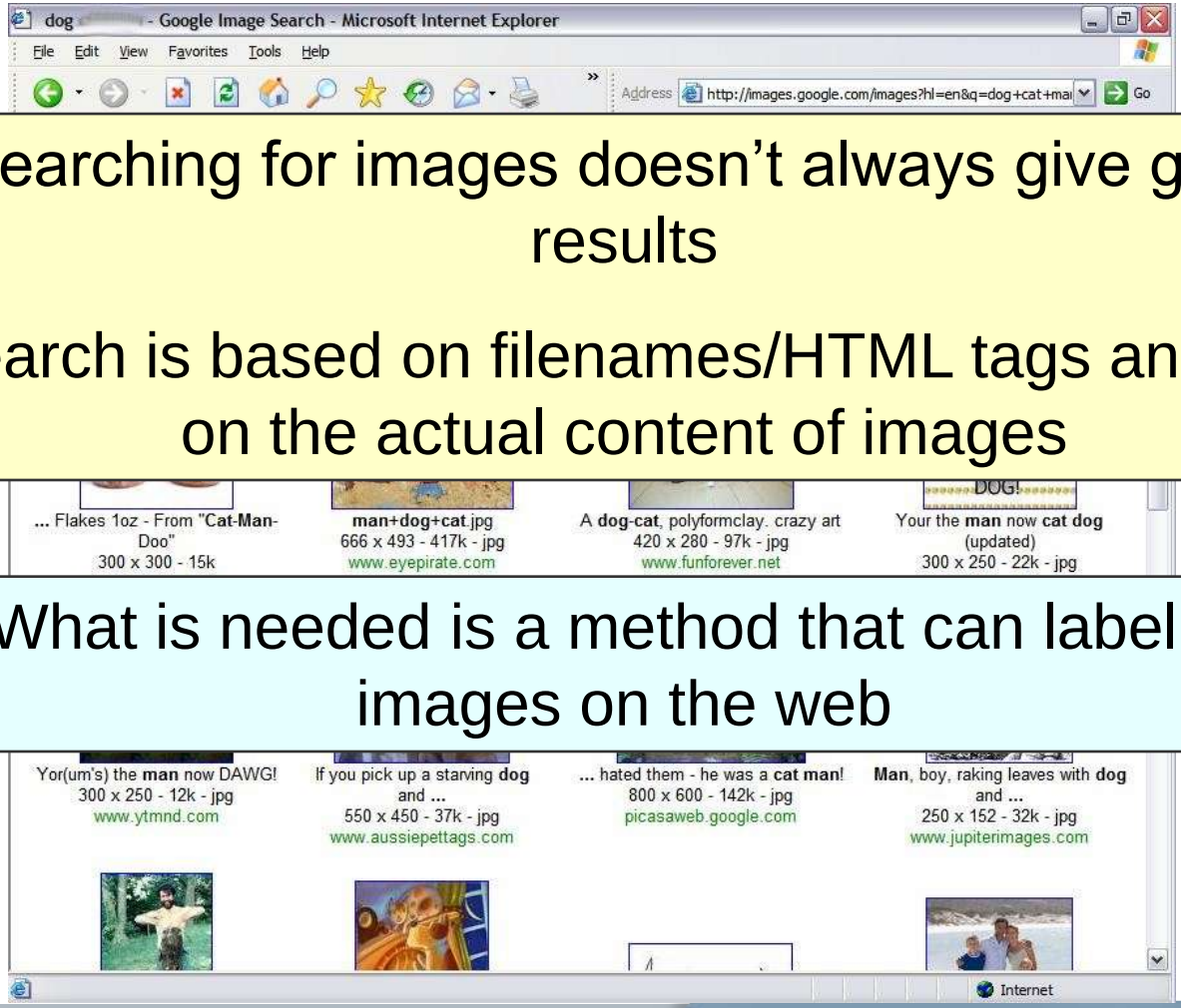
Harrison Ford

Fugitive

STILL AN OPEN PROBLEM

Human Processing Power

- Searching for images



Human Processing Power

- **How HBC can help**
  - A method to label all images on the web – fast & cheap
  - USING HUMANS **CLEVERLY**
  - Make people want to label images for free
  - How do we do that?
  - Create an extremely enjoyable online multiplayer game
  - **The ESP Game** (by Luis von Ahn)
  - As people play, as a side effect of the game they label images

The ESP Game

- The ESP Game has two characteristics
  - labels are always accurate
  - labelling is extremely fast

- If put up on a popular gaming site, all images on google search can be labelled in a matter of weeks!

The ESP Game – how does it work?

The ESP Game

Two-player online game

Partners don't know each other and cannot communicate

Goal of the game

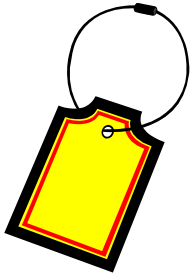
type the exact same word

Only thing in common is

an image

The ESP Game – how does it work?

This word that the two players agree on is usually a very good label for the image because it comes from two independent sources



The ESP Game – how does it work?

The ESP Game

PLAYER 1

Guessing: **BOY**

Guessing: **WHITE SHIRT**

Guessing: **KID**

SUCCESS: You agree on **BOY**

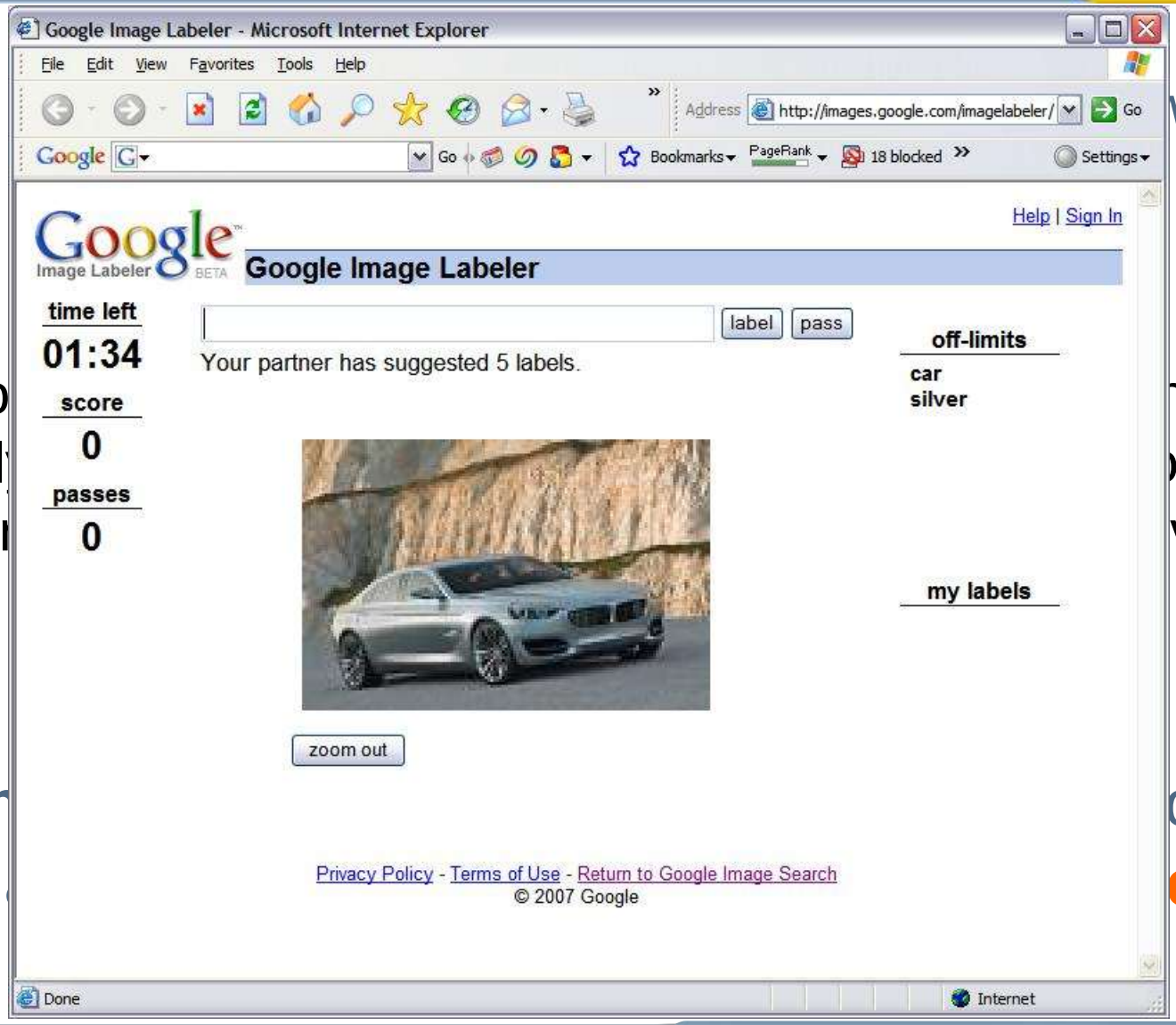
PLAYER 2

Guessing: **HAT**

Guessing: **BOY**

SUCCESS: You agree on **BOY**

The ESP Game – how does it work?



Is HBC of any use in the IT security domain?

Is HBC useful in IT security?

- We routinely deals with problems that cannot be completely solved by computers

- Some of these could be easily solved using HBC
  - **Identification of phished websites**
  - **Spam classification**
  - **Keyboard typo analysis**
  - **Content classification**

Identification of phished websites

Phished website identification

- **The problem:**
  - Automated techniques used to identify phishing websites are not accurate enough
  - Based on logo comparison, visual similarity of webpage layout / images (histogram)

At best these techniques tell you a suspected phishing website

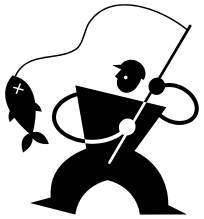
You need a human to confirm it

Phished website identification

- **How HBC can help**
  - Create an enjoyable **online multiplayer game**
  - Show players screenshots of suspected phishing sites
  - Design the game to get players to vote

Go Fish!

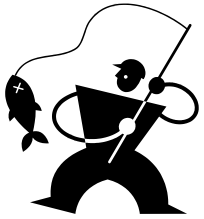
the anti-phishing game



Phished website identification

Go Fish!

the anti-phishing game



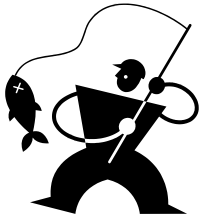
Do these webpages belong to the same company?

YES **NO**

Phished website identification

Go Fish!

the anti-phishing game



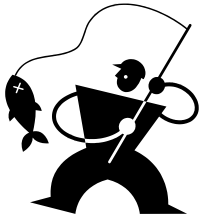
Do both these logos belong to **Citibank**?

YES **NO**

Phished website identification

Go Fish!

the anti-phishing game



MULTIPLAYER VERSION

PLAYER 1

PLAYER 2

What does this image remind you of?

Guessing: **Honda City**

Guessing: **Citibank**

SUCCESS: You agree on **Citibank**

Guessing: **Citibank**

SUCCESS: You agree on **Citibank**

Phished website identification

- Phishing site confirmed if atleast **N** players have agreed on it
- What about cheating?
  - Countered by randomly giving test images in the game
- There is a probabilistic guarantee that if **N** players have successfully identified all test images then they also correctly identify the real image

if **X** = probability of a false positive given that the player successfully identified all test images

Overall probability of a false positive = $\mathbf{X^N}$

Spam identification

Spam identification

- **The problem:**
  - Antispam software is excellent at detecting spam that has been around for a while
  - Not very effective for new (day zero) spam because it works mostly on signatures
  - **Image spam** is even more difficult to identify
  - And **VOIP** / **audio** and **video** spams are not identified at all

At best, existing techniques can tell you that it is suspected spam

You need a human to confirm it

Spam identification

- **How HBC can help**
  - Create an enjoyable **online multiplayer game**
  - Show players samples of suspected spam
  - Design the game to get players to vote

SpamBam

the anti-spam game

Spam identification

SpamBam
the anti-spam game

PLAYER 1 PLAYER 2

What does this image remind you of?

Guessing: **my wife**

Guessing: **spam**

Guessing: **spam**

SUCCESS: You agree on **spam**

SUCCESS: You agree on **spam**

Spam identification

- Works equally well with
  - text spam
  - image spam
  - VOIP spam
  - video spam
- Highly effective against Day Zero spam

Especially useful for **parasitic spam** (piggy-back spam) which is extremely difficult to differentiate from genuine mail

Typo Analysis

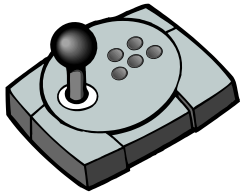
Typo analysis

- **The problem:**
  - Typo-analysis data is required to prevent typo-squatting
  - This requires people to type out common URLs and then measure errors that occurred during typing
  - Needs to be done across diverse sets of people and keyboard layouts

Collecting massive amounts of such data is not easy

Typo analysis

- **How HBC can help**
  - Create an enjoyable **online multiplayer game**
  - Show players words that they need to type
  - Design the game to collect error statistics



KeyBang!

the typo-analysis game

Typo analysis

KeyBang!

the typo-analysis game

Goal of the game: type the word that appears

Speed is important, **Typos** are not

Typo analysis

- Typo-data collected is analyzed
- Probabilities of various typos are computed
- These probabilities are used for ranking URL typos

Pr(Skip Letters) = Ps(a), ….Ps (z), Ps (a | x)….Ps (f | s)
Pr(Double letters) = Pd(a), ….Pd (z), Pd (a | x)….Pd (f | s)
Pr(Reverse letters) = Pr(a | x), ….Pr (x | a)…..
Pr(Missed) = Pm(w | s), Pm (d | s), Pm (a | s), Pm (x | s), ...
Pr(Skip | Missed)

……..

Using these probabilities, rank of a typo can be calculated as:
*Rank(symnetc) = Pr(Skip) * Ps(a)* Pr(Missed | Skip) * Pm(t | e) * C*

Content Classification

Content Classification

- **The problem:**
  - Machine-learning based classifiers need a training dataset
  - This dataset needs to be a huge pre-classified corpus of documents
  - Depending on the number of categories it could run into several hundred thousand documents

Collecting documents is easy

Pre-classifying them manually is painful

Content classification

- **How HBC can help**
  - Create an enjoyable **online multiplayer game**
  - Show players content that they need to classify
  - Design the game such that players choose the correct category

incognito

the classification game

Content classification

incognito

the classification game

Congratulations on an outstanding quarter, our fourth in a row. Your focus on driving revenue growth, while keeping a watchful eye on cost, was clearly the difference maker in our quarter. As a result, we over-performed on all four of our key financial metrics – revenue, earnings per share, deferred revenue, and cash flow from operations. On a non-GAAP basis, revenue for the quarter grew 15 percent, reaching $1.53 billion, and earnings per share of $0.33 were 27 percent higher than last year. These results, coupled with the strength of our pipeline for the March quarter, have provided the confidence for us to raise our outlook for the full year to almost $6 billon.

What does the above text remind you of?

OR

Content classification

symantec.

incognito

the classification game

The Rockets are 4-0 since losing Yao Ming for the season with a stress fracture in his foot and have won 20 of their last 21 games, and own the longest winning streak this season. Houston is still clawing for a playoff spot in the tough Western conference and plays the division rival Mavericks in Dallas on Thursday night.

"It's great to have your name down in history," Rafer Alston said. "We are having some great regular-season success right now, but we know the big picture is getting it done in the postseason. You enjoy it for the moment, but you have to continue to prepare for every game."

The Rockets led by 17 at halftime, but two runs by Indiana in the third quarter got the Pacers within seven. Houston scored 11 of the last 13 points of the quarter to stretch the lead to 16 and the Pacers wouldn't threaten again.

What does the above text remind you of?

OR

Content classification

- Category confirmed if atleast **N** players have agreed on it
- Works well for sub-categories

Works especially well for content that changes fast (sports, news, etc)

Conclusion

- **Human Based Computation** can be used to effectively solve problems that computers cannot solve yet
- Game designs can be further improved to make them attractive to players
- HBC problems can be posted to Amazon's **Mechanical Turk**

References

- **Luis von Ahn** (Carnegie Melon University)

  http://www.cs.cmu.edu/~biglou/research.html

- **Google Image Labeler**

  http://images.google.com/imagelabeler

symantec™

Confidence in a connected world.

from THE HITCHHIKER'S GUIDE TO THE GALAXY

Eddie the Computer: *I am pleased to inform you that two thermo-nuclear missiles are now headed towards our spaceship... if you don't mind, I am going to take action*

Arthur Dent: *COMPUTER DO SOMETHING!!*

Eddie the Computer: *Sure thing fella! Switching to manual control... good luck!*

Questions?

Thank You!

Sumesh Jaiswal

Sumesh_Jaiswal@symantec.com

+91 20 66157539