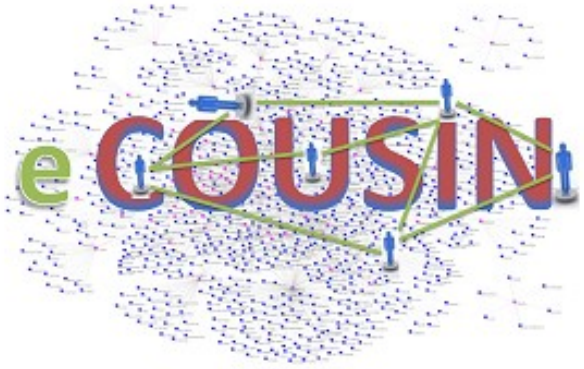


eCOUSIN: enhanced Content Distribution with Social Information

Rubén Cuevas
rcuevas@it.uc3m.es
Universidad Carlos III de Madrid

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° 318398.

Outline

- eCOUSIN overview
  - Consortium
  - Motivation
  - Goal
  - Main Contributions
- Counteracting fraud in Video Advertisement: detection of fake views in Video Content Portals
  - YouTube Background
  - Basic Methodology
  - How well does Youtube counts (discounts) real (fake) views?
  - A closer look into YouTube False Positives detection algorithm
  - Appropriateness of thresholds used in YouTube detection algorithm
  - Detection of Fake Views in Monetized videos
  - Conclusion

Information Society Technologies

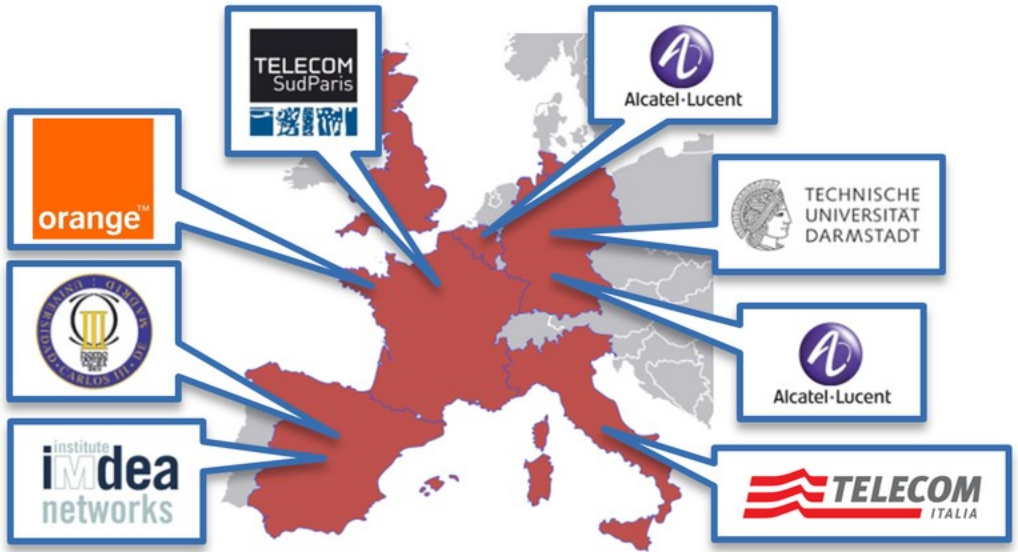
eCOUSIN Overview

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° 318398.

eCOUSIN overview

CONSORTIUM

- 2 Operators (Orange, TI)
- 1 Manufactures (AL-BL)
- 3 Universities (**UC3M**, TSP, TUD)
- 1 research institute (**IMDEA Networks**)



FP7-ICT-318398 (NOV 2013 – APR 2015)

MOTIVATION

- Current solutions for online content distribution (e.g., CDNs) work reasonably well for popular content
- However, with the proliferation of OSNs, Video Platforms, etc. non-popular content is becoming more and more relevant and current solutions are struggling to predict where this content is going to be consumed in order to prefetch/cache it.

eCOUSIN overview

institute imdea networks

GOAL

- Leverage available social information in order to improve online content distribution solutions including
  - Caching algorithms in traditional Content Distribution solutions like Content Delivery Networks (CDNs)
  - Caching algorithms in novel Content Distribution solutions like Content Centric Networking
  - Prefetching algorithms for traffic offloading in cellular networks
  - etc

MAIN CONTRIBUTIONS (SO FAR)

- Set of sophisticated measurement tools to gather data from:
  - Main Online Social Networks (Facebook, Twitter and Google+)
  - Main Content Distribution Platforms (YouTube, BitTorrent)
- Demonstrators for solutions of social enhanced content distribution algorithms in both traditional CDNs and CCN infrastructure.
  - Improvement factor around 20% with respect to existing solutions
- Implementation of social-based mobile applications for content prefetching

Information Society Technologies

Counteracting Fraud In Video Advertising: detection of fake views in video content portals

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° 318398.

Motivation

- Online advertising sustains a large percentage of Internet business
  - $42B revenue in 2013, 17%+ 2012 (Source: IAB)
  - It attracts fraudulent activity, e.g., click fraud
- New forms of advertising (like video ads) are becoming more important
  - $3B revenue in 2013, 7% whole revenue (Source: IAB)
  - New opportunities for fraud are placed (buy fake views for Youtube, Vimeo, Dailymotion)

Information Society Technologies

Motivation

- Impact in Content Distribution and relation to eCOUSIN
  - Any efficient Content Distribution (either pull or push strategies) rely on predicting where a content is going to be consumed
  - Fraudulent Activity (fake views) provides misleading information that may lead to content to be placed in locations where it is actually not going to be consumed
- Content Providers, Network Operators, CDN Operators would benefit from reliable techniques to identify/eliminate fake views

Our Goal

- Analyzing the efficiency of existing fake view detection algorithms
  - False Positive Rate: Fake views counted as real ones
  - False Negative Rate: Actual views discounted

- We will focus in YouTube
  - It dominates the market of user generated online video
  - It has expressed its interest on fighting this phenomenon, then we assume it will have a sophisticated system in place

- We consider *InStream* ads (i.e., video ads). There are other type of ads (more similar to banner ads)

Background

- YouTube provides two different statistics:
  - Public view counter
  - YouTube Analytics (accessible to the uploader)
- Monetization program
  - Registration required
  - Ads are associated to videos and the user receives a share of the revenue generated by each view
  - Monetization information available in YT Analytics
- This model is subject to fraudulent activity
  - A user registers its videos in the monetization program
  - Afterwards it generates artificial (fake) views

Information Society Technologies

Methodology

False Negatives

- We embedded our YouTube video in a website and ask people to watch it
  - We register all the visits to our website, whether the users watched or not the video and for how long
  - Since it is our video we know the YT view count

False Positives

- We developed a sophisticated modular bot (based on Selenium WebDriver that generate fake views emulating from very close to human behavior to a very simple behavior
  - We use it to generate fake views on our own videos
  - Since it is our video we know the YT view count

Information Society Technologies

HOW WELL DOES YOUTUBE COUNT (DISCOUNT) REAL (FAKE) VIEWS?

False negatives

- Exp 1 → We advertise our video in social media

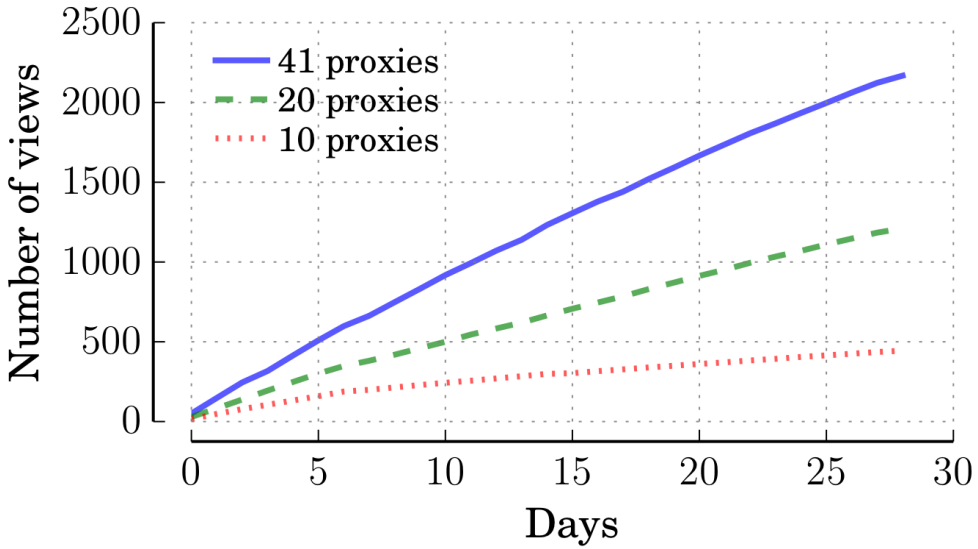
- Exp 2 → Viewers from a crowdsourcing webs

$$R_{FN} = 1 - \frac{\#YouTube\ counted\ views}{\#Views\ registered\ in\ the\ webpage}$$

	# performed real views	#counted views	R_{FN}
Exp. 1	430	340	20,9%
Exp. 2	990	572	42,2%

Table 3: False Negative Ratio for the two conducted experiments.

False positives



$$R_{FP} = \frac{\#YouTube\ counted\ views}{\#Bot\ performed\ views}$$

Using our methodology

- 3 views per proxied bot per day
- False Positive Ratio (100%, 98% and 95%, respectively)

Information Society Technologies

First insights

- It presents a non-negligible rate of false negatives
  - Real views discounted → Harming video publishers

- It presents a high rate of false positives
  - Fake views counted → Harming advertisers
  - Worrisome!!!

Information Society Technologies

A CLOSER LOOK INTO YOUTUBE FAKE VIEWS DETECTION ALGORITHM

Information Society Technologies

Methodology

- We use different version of our bot in order to infer which parameters are relevant in YouTube's fake views detection algorithm
- Modules: User-agent, Referrer, Signed-in users, Proxies, View duration, Inter-arrival time between views
- Robots:
  - Complete (C)
  - Single Sources (SS)
  - Deterministic (D)
  - Without Likes (W/O L)

Parameters used to detect fake view

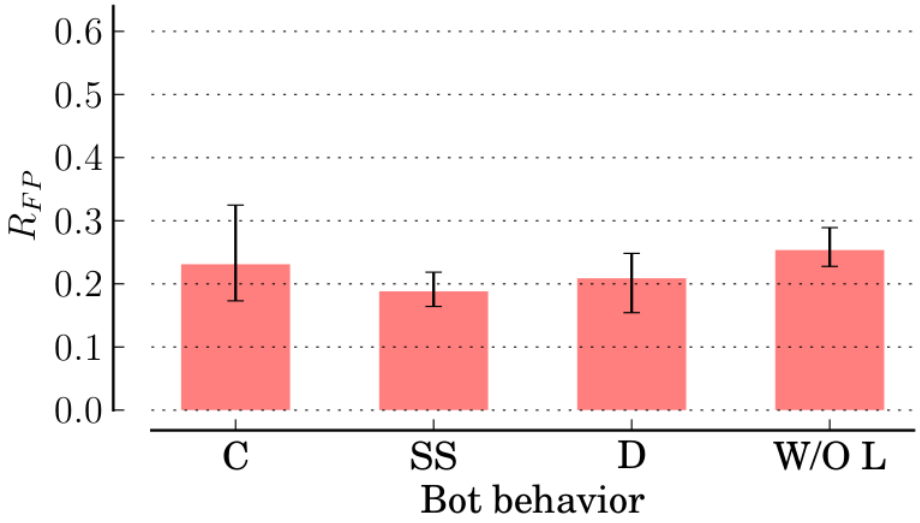


Figure 2: False Positive Ratio associated to the different types of bot.

- All robots get a similar rate of False Positives
- Thus, the parameters that made each robot "different" is not used by YT in the detection mechanism

- But… all robots have a commonalty
  - All activity is conducted from a single IP address

Information Society Technologies

Limit in the # of daily views per IP

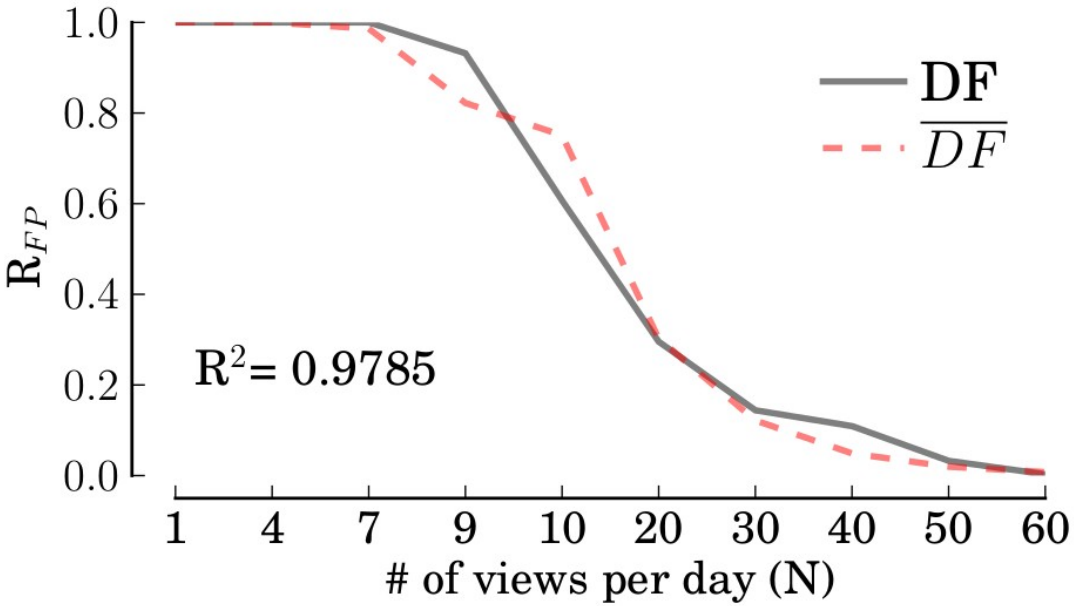


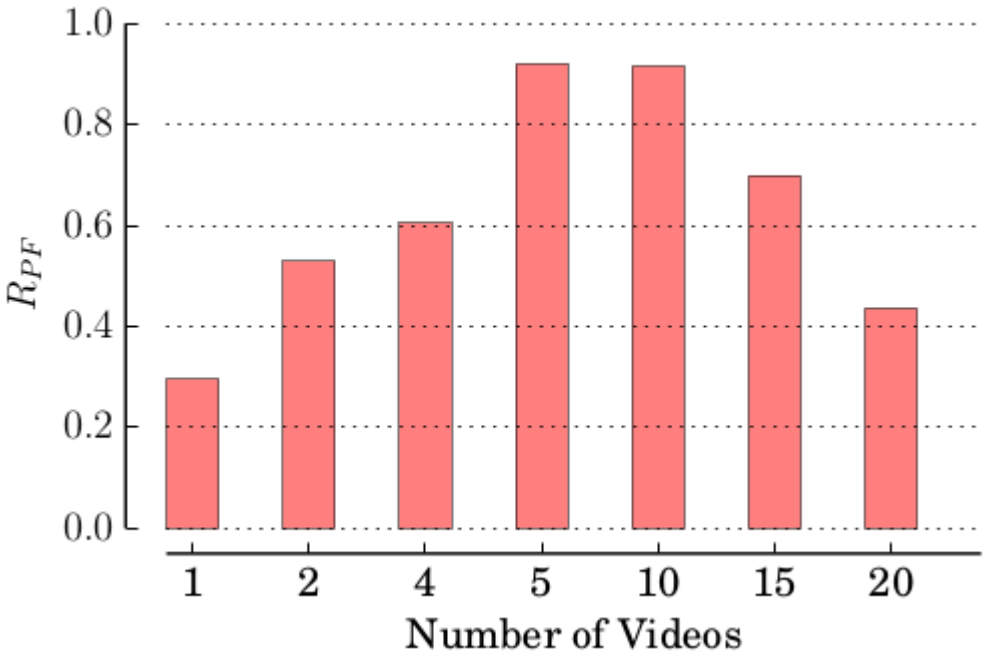
Figure 3: False Positive ratio for different number of fake views performed from an IP address to a video

$$\overline{DF}(N) = \begin{cases} 1 & \text{if } N \leq 7, \\ 1.8645e^{-0.091n} & otherwise \end{cases}$$

Information Society Technologies

A hint on the punishment strate

- Experiment: a robot performs 20 views/day homogenously distributed across x videos
- Extreme cases present the highest discount rate

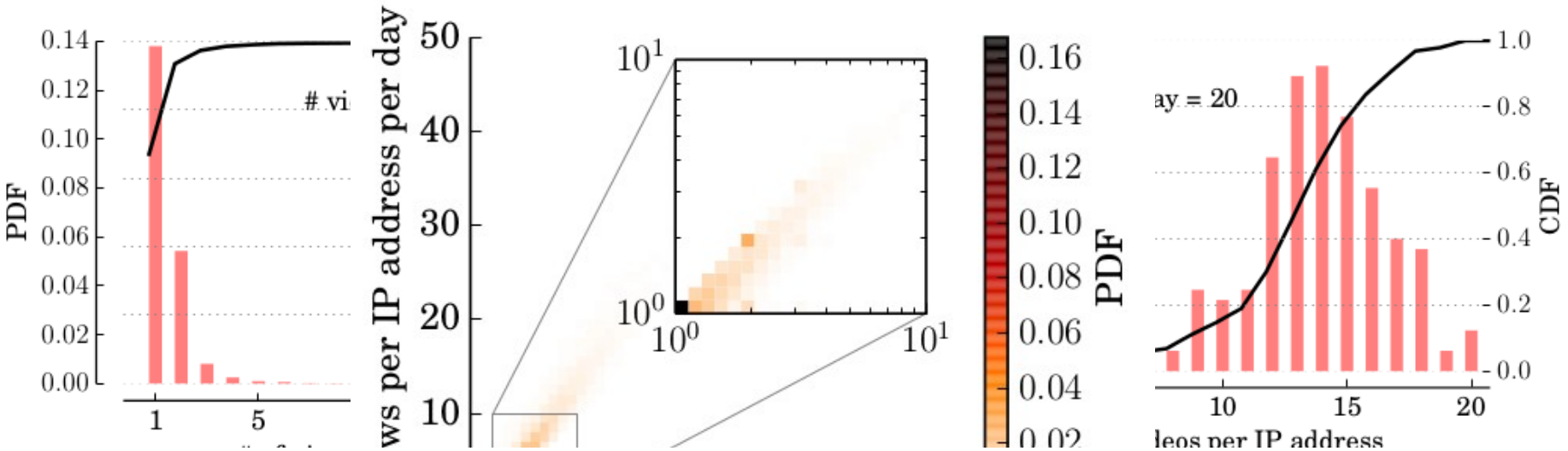


APPROPRIATNESS OF THRESHOLDS USED IN YOUTUBE DETECTION ALGORITHM

Dataset & Methodology

- Dataset: 3.9 million YouTube sessions, 1.3 million videos, requested by 28.000 IP addresses over 49 days.
- We validate the appropriateness of the previously unveiled thresholds based on the statistical information of our trace

Results



Good threshold for the number of daily views (99% of the users watch < 8 videos a day)

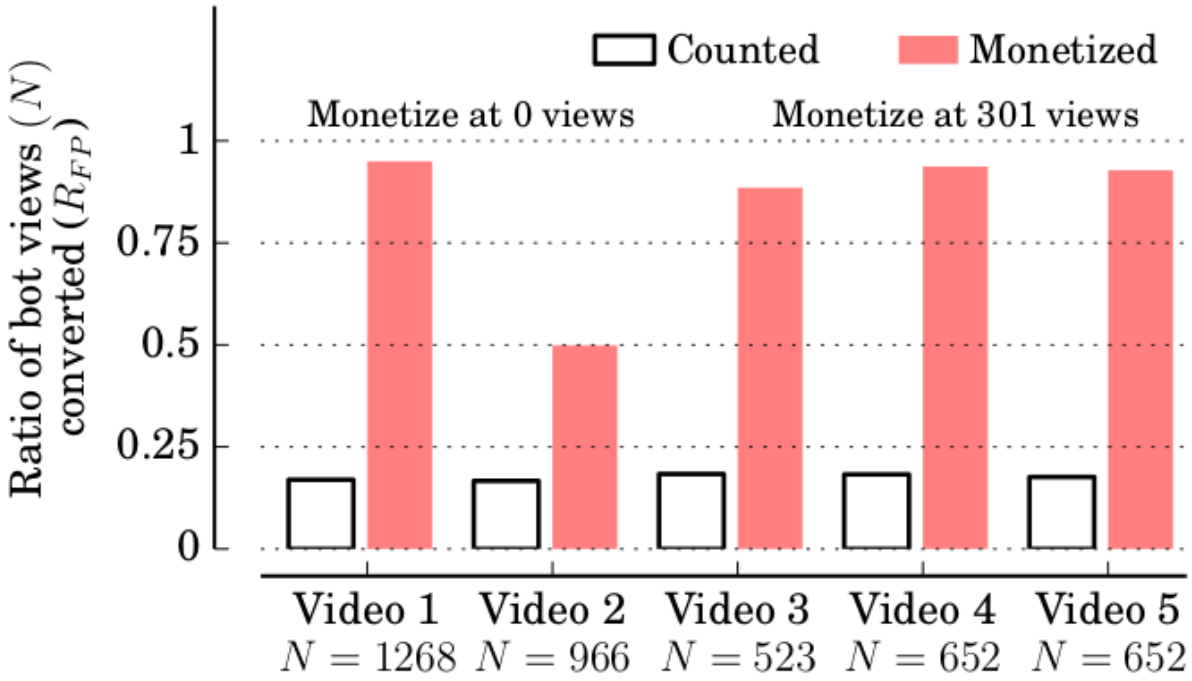
Bad threshold for the case of 20 views per day (20 views on 15 different videos is one of the most common cases, but it is hardly punished) → Potential cause of false negatives

views performed from an IP address to a video

DETECTION OF FAKE VIEWS IN MONETIZED VIDEOS

Monetized videos

- We enrolled 5 videos in the YouTube monetization program
- We performed 20 views/day for each of them



Information Society Technologies

Conclusion

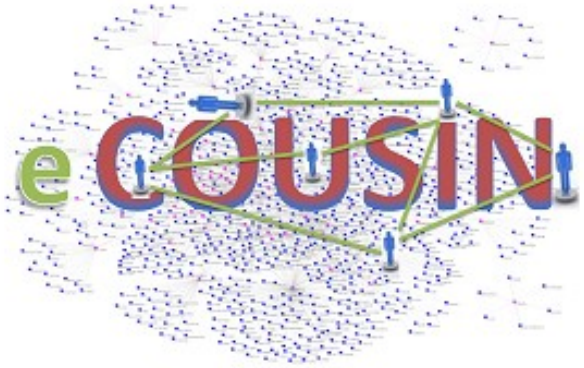
1. Novel Methodology to evaluate the efficiency of fake views' detection mechanism in video portals

1. Datailed analysis of YT fake views detection algorithms:
   - Based on global IP address behaviour (i.e., number of views performed and number of videos watched)

2. From real YouTube sessions, we see that the view-discount factors used by YouTube are not properly defined what may lead to false positives

3. YouTube uses different algorithms to detect fake views when counting the views that are monetized

Information Society Technologies



institute
imdea
networks

Question & Answers

Thank you

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° 318398.