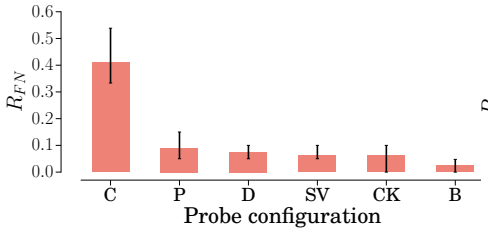

Figure 3: False negative rate obtained for each of the experiment configurations.

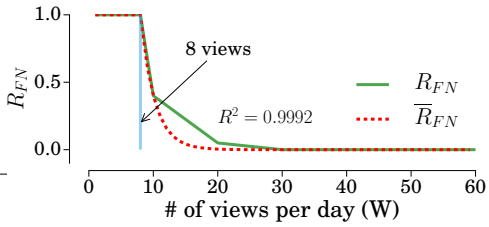

Figure 4: False negative rate to one video depending on the number of views per day.

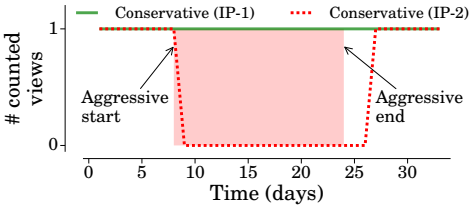

Figure 5: Number of views counted by YouTube for both IP-1 (*conservative* probe) and IP-2 (*conservative* and *aggressive* probe).

- Deterministic (D): The goal of this behavior is to define a simple, and completely deterministic pattern of views. This behavior eliminates any randomness by setting to constant values the view time (40 secs.) and the time between views (72 mins.). All other parameters take their default values from Table 2. We expect this behavior to be easily identified.

- Vary view burst (B): The goal of this behavior is to study the impact of making views in bursts. In particular, the probes run the **D**eterministic behavior, setting the time between consecutive views to 0, and generating a burst of N = 20 consecutive views every day. The time between consecutive bursts can be configured, and is set to 24 hours in the experiments. Since bursts of views from a given IP address, to a single video are atypical for users in *YT-1* and *YT-2*, we expect this behavior to be easily identified, and to have low false negative ratio.

- Vary inter-view wait time (P): The goal of this behavior is to measure the impact of varying the time between views over a day. The probe runs the **D**eterministic behavior, but varies the time between two consecutive views. With this behavior, we aim to determine whether adding some noise to the inter-arrival pattern of views has any impact when compared to a deterministic pattern. In the following, we use a Poisson with $\lambda = 20$.

- Short Views (SV): The goal of this behavior is to measure the impact of making very short views to videos. In the following, the probe runs the **D**eterministic behavior, but sets the duration of video views to 1 sec. Since consecutive short views are atypical for real users, we expect to see this behavior will be heavily penalized.

- Cookies (CK): The goal of this behavior is to measure to what extent audit systems rely on user identifiers when auditing views. We use cookies since they are the most commonly used method to track users [50]. We consider the extreme case in which the probe uses the **D**eterministic behavior, and performs all views using the same cookie.

- Complete (C): The goal of this behavior is provide a baseline by emulating some real-user like features. Therefore we enable all the parameters in Table 2, except the cookies. Specifically, the view duration time and wait time between views are set to Poisson processes with λ = duration of the targeted video, and λ = 72min, respectively. Finally, the Referrer and User-Agent fields are selected randomly. Given the variation in the parameters, we expect this behavior to be the least penalized.

For each behavior, Figure 3 gives the average and max/min R_{FN}. As expected, the **C**omplete behavior yields the highest false positive rate (≈40%), and is on average 4x larger with respect to the other behaviors ($R_{FN} < 10\%$). This indicates that adding some randomness to basic HTTP parameters such as the User-Agent, or the Referrer makes it significantly harder for YouTube to detect fake views.

Looking at the impact of varying the wait time between views (**P**, **D** and **B**), we observe that the view audit system penalizes **B**ursty behavior the most heavily, discounting 98% of the views. Comparing the **D**eterministic and the **S**hort **V**iews behaviors, contrary to our expectation, they are both similarly penalized. We observe that the audit system counts as valid 7% and 6% of views for the **D** and **SV** configurations respectively . Finally, we observe no significant change to enabling/disabling user tracking via the cookies. The differences in false negative ratios with cookies(**CK**) and without (**D**, **SV**, etc.) cookies are negligible.

In summary, *we find that YouTube is able to identify the simplest suspicious behavior patterns, schemes using static HTTP connection parameters are easily identified. Indeed, the view audit system is able to remove more than 90% of fake views generated under these attack configurations. We observe however that adding some variability to HTTP connection parameters may increase the effectiveness of attacks up to ∼30%*. While these results explain the false negative rate difference between the considered configurations and the benchmark, they do not explain the significant number (60%) of discounted fake views common to all the configurations. The only variable common to all the configurations, and which may be responsible for such large a penalization is that they each perform their views from a unique public IP address. This along with the fact that IP addresses are one of the strongest online users identifiers [49], and one of the key parameters many security online services use [13, 51, 52] leads us to believe that the video viewing pattern from an IP address is a key element for the fake view detection mechanism of YouTube. We analyze this hypothesis in the next subsection.

5.2 Influence of Video Viewing Pattern in the detection

In this subsection we analyze the response of YouTube's view audit systems to the fake view patterns of an IP address. We first look at the impact of view patterns to a single video, then explore the cases for a single IP viewing multiple videos, and finally a single video receiving views from multiple IP addresses.

Note that we have checked that reusing IP addresses or using IPs from the same IP prefix in the experiments does not bias the obtained results. Details on the tests conducted to validate these aspects can be found in Sections 5.4 and 5.5 of the Technical Report [53] of this paper. In addition, we did not observe any difference in the response of YouTube's detection system to IPs from PlanetLab proxies and those from our /24 IP prefixes in Germany and Spain.

One video, One IP address

We start by examining how YouTube discounts the views generated by a single IP address to a single video. In particular, we are interested in understanding how the view penalization threshold(s) are triggered, when varying the number of views per day. We conduct a simple experiment, in which the probe generates $W = [1, 4, 7, 8, 9, 10, 20, 30, 40, 50, 60]$ views per day, to a given video, for 8 days. We use the previously defined **D**eterministic behavior for this experiment.

The results of this experiment are presented in Figure 4, which reports the R_{FN} for the different numbers of views (W). We observe that the view audit system counts all the views up to a rate of 8 per day. From 9 views on, the R_{FN} decays exponentially and is 0 for more than 30 views per day. We observe that the R_{FN} with respect to the views per day (W) follows an exponential decay function, and can be modeled with the following parameters, with an R^2 = 0.999:

$$\overline{R_{FN}}(W) = \begin{cases} 1 & \text{if } W \leq 8, \\ e^{-0.455n} & otherwise \end{cases}$$

For the previous experiments, we used newly uploaded videos. To understand whether this has any impact on the results obtained, we look at the response of the audit system when we generate views for videos previously uploaded to YouTube and are moderately popular, and repeat the experiment. With the permission of the uploaders, we use two videos with roughly 12K (100 in the last month) and 300K (5K in the last month) registered views at the start of the experiment. To identify the activity of the probe in the results, we configure it to use very rare User-Agents (Bada, HitTop, MeeGo and Nintendo 3DS). Before starting the experiment, we validate that the targeted videos have not received any views from the selected User-Agents in the previous 6 months using YouTube Analytics.

Setting $W = [8, 9, 10, 20]$ views per day, we find that the view audit system again starts discounting views from 8 views per day, for a given IP address, and R_{FN} follows the same decay pattern. This suggests that view audit system of YouTube are triggered by a fixed threshold regardless of a video's popularity.

Multiple videos, One IP address

Having observed how the views from a single IP address to a single video are penalized, we now look at the response of the view audit system when a single IP address spreads its views over several videos. Given the previous result, we expect that aggressive IP addresses will be heavily penalized, independent of the number of videos they target.

In the following, we first test this hypothesis, and then present the results of a large scale measurements to understand how the rate of false negative varies with respect to the number of videos viewed, and views performed, per IP address.

To begin to understand how the view audit system differentiates between IP addresses, we define two simple probe behaviors; *conservative* and *aggressive*. The *conservative* probe performs 1 view per day, while the *aggressive* probe performs 30 views per day. We set up the following experiment: in two IP addresses, IP-1 and IP-2, we launch an instance of the *conservative* probe to a different video for 34 days. Moreover, in IP-2, we also launch an instance of the *aggressive* probe, starting at day 8 and stopping at day 24, while there is no aggressive probe in IP-1.

Figure 5 gives the number of views the audit system counts over time, for the *conservative* probes in the two IP addresses. Since *conservative* probes perform just 1 view per day, to the video, we expect to see either; 1, if it is counted, or 0, if it is penalized. We find that the audit system counts all the views from the *conservative* probe in IP-1, and penalizes the *conservative* probe in IP-2 for the days that the aggressive probe is also running from IP-2 (days 9-26). We observe that view penalization starts 24 hours after the launch of the *aggressive* probe in IP-2, and ends two days after the probe stops. Repeating the experiment three times in total, we obtain the same results. From this, we conclude that YouTube's fake view audit system labels and tracks the behavior of IP addresses based on their global view behavior across all videos that they visit.

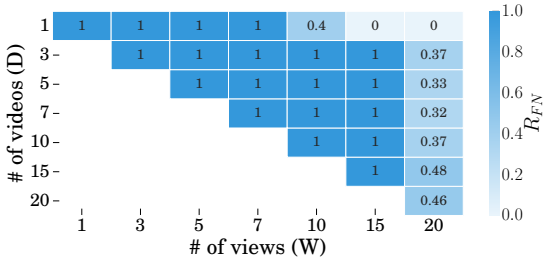

Figure 6: R_{FN} for several combinations of the number of views W and the number of watched videos D.

Having observed that YouTube's view audit system labels IP addresses based on their behavior, we now look at how it penalizes the behavior of an IP address across video views. To do so, we conduct a large scale experiment in which we perform $W = [1, 3, 5, 7, 10, 15, 20]$ views per day, uniformly distributed across $D = [1, 3, 5, 7, 10, 15, 20]$ videos (with $W \geq D$), over a period of 7 days.[6] In total, we ran 28 combinations of views and videos. Finally, we use the **D**eterministic behavior for the probe.

Figure 6 reports the R_{FN} across the 28 combinations considered. Looking at the evolution of R_{FN} for a fixed number of videos, we observe the exponential decay revealed in Figure 4. However, in this case, view penalization is triggered after 15 views per day, when a viewer watches 3 or more distinct videos per day ($D \geq 3$), whereas in Figure 4 it was triggered after 8 views per day (for $D = 1$). With respect to the evolution of R_{FN} for a fixed number of daily views, we observe that when all views are to a single video, the penalization is much more severe, than when they spread across three or more videos.

One video, Multiple IP addresses

Having established, that an IP address is tracked across video views, we now look at the response of the view audit system, when the views to a given video are distributed across several IP addresses. To this end, we use 70 different PlanetLab proxies, and divide them in 3 independent groups of different size $N = [10, 20, 40]$. We assign each group of proxies a different video on YouTube, and configure each proxy to generate views to its corresponding video. We again utilize the **D**eterministic behavior of the probe, and report the results with each PlanetLab proxy group to generate 3 views per day. Overall, the experiment generates 30, 60 and 120 views per day to a video, which should result in $R_{FP} = 0$, if coming from a single IP address.

From this experiment, we observe that the growth in number of views over time is linear for all behaviors, and that overall $R_{FN} > 73\%$ in all three experiments. This indicates that distributing activity across multiple IP addresses results in a substantial increase in the R_{FN} enabling attackers to inflate view counters easily.

This experiment suggests that *YouTube is vulnerable to attacks that employ many IP addresses* (such as those from botnets), and such attacks can apparently achieve an arbitrarily large number of views.

5.3 Impact of NATed IP addresses on the audit system

As NAT devices aggregate traffic, they typically contain the video viewing activity from multiple, usually private, IP addresses. In large NATed networks, such as campus networks, corporate net-

[6]Note that we only run experiments for $W \geq D$. For instance, in the case of $W = 5$ we run experiments for $D = 1, 3, 5$.

Experiment	W (views/day)	# U (users behind the NAT)	U/W	R_{FN}
Loc. 1	20	∼50	∼ 2.5	0.9
Loc. 2	75	∼100	∼ 1.33	0.43
Loc. 3	100	∼50	∼ 0.5	0.36

Table 5: R_{FN} and information about the three scenarios for the experiments we conduct from NATed IP addresses.

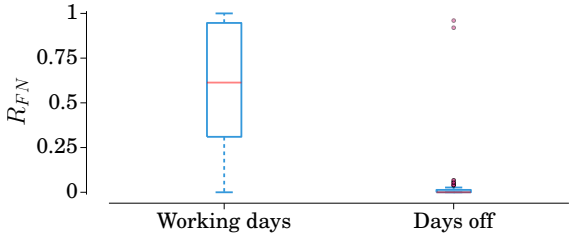

Figure 7: Distribution of daily R_{FN} for working days and days off for Location 2.

works, and in some cases ISP networks, this activity may be significantly large.

Therefore, in the following set of experiments, we investigate how the view audit system of YouTube penalizes the views originating from NATed networks. To do so, we install the probes on three machines accessing the Internet from NATed networks located at three different locations, and we configure them to perform 20 (Location 1), 75 (Location 2), and 100 (Location 3) views per day for a period of 8 days. We again use the **D**eterministic behavior.

Table 5 reports the R_{FN} for each experiment along with information of the different NATed scenarios. Note that, although the probe generates views aggressively, the R_{FN} is surprisingly large in all cases. This suggests that the YouTube's view audit system has problems in properly identifying suspicious activity from NATed networks. To confirm this finding, we separately analyze the R_{FN} on working days and days off (i.e., weekends and holidays) in Location 2, and run the experiment for 194 days. Note, during working days the volume of NATed traffic from the network is high, whereas it is low during the days off. Figure 7 shows the distribution of the daily false negative rate for working days and days off in the boxplots. The results confirm that YouTube discounts almost all views during days off, i.e., when the traffic is more exposed, but has problems in discount views (median R_{FN} = 60%) for workdays, i.e., when the views are hidden by larger volumes of traffic. Hence, this suggests that malicious users can dramatically increase the efficiency of their activity by gaining access to machines located behind large (active) NATed networks, e.g., a public campus network.

In summary, *the view audit system of YouTube implements an exponential discount factor of the number of views performed from a single IP address that increases with the rate of views. However, the results show that some simple modifications in a fraudster's strategy can considerably increase the false negative rate. In practice, i) adding some randomness in the HTTP connection attributes such as the User-Agent or the Referrer, ii) distributing the malicious activity across multiple IP addresses, or iii) performing fake views from NATed networks, are shown to be effective.*

6. YOUTUBE'S AUDIT SYSTEM FOR MONETIZED VIEWS

Surprisingly, the results in Section 4 indicate that YouTube monetizes (almost) all the fake views we generate, while discounting them from the public view counters. In this section we study in more detail the audit mechanism applied to monetized views, to further understand this seemingly anomalous behavior.

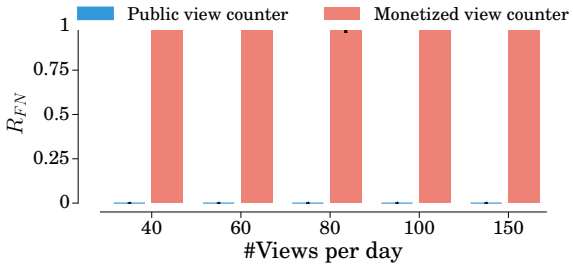

Figure 8: Comparison of false negative ratio for the public and monetized view counters of YouTube for different daily rates of generated views W.

We reuse the configuration described in Section 4.2 for YouTube, and conduct a new set of experiments, whereby we increment the number of views per day the probe generates from a single IP address, to a single video. In particular, we set $W = [40, 60, 80, 100, 150]$ to cover a wide range of aggressive configurations. We conduct each experiment for 10 days. Figure 8 reports the R_{FN} for both the public and the monetized view counters. Again, the main bars and error bars represent the average and the max/min R_{FN}, respectively.

We find that the monetized view counter's audit system penalizes a negligible portion of views in all the considered configurations, while the public view counter's audit system penalizes most of the fake views. These results confirm the preliminary observation in Section 4; YouTube applies different penalization schemes to the fake views in the monetized and public view counter, with the former being much more permissive than the latter.

6.1 Counting monetized views from the advertiser's perspective

To gain insight into the monetary implications of the above finding, we designed a tailored experiment in which we assume the role of an advertiser exposed to fraudulent views. To do so, we first create an advertiser account using the Google AdWords service. AdWords enables us to configure advertising campaigns in YouTube, so that our video ads can target YouTube videos whose uploaders participate in the monetization programme. We then create a video ad and build an advertising campaign to target experiment videos that we have previously uploaded to YouTube. In this way, we play both the role of the advertiser and the publisher in the campaigns, and can build a complete picture of the trade.

AdWords offers a wide range of tools to aid in the design of video advertising campaigns. Advertisers can tailor campaigns to reach specific YouTube viewer demographics (per interests, country, language, gender, age), or target specific YouTube videos. With the aim of checking if YouTube actually charges advertisers in presence of fake views, we configure a campaign to target the views from the countries where the proxies are located (accepting all languages, genders and ages) and headed to the experiment videos. Then, we use the probe to generate views to these videos.

YouTube deploys a sophisticated bidding algorithm that selects in real time the ad to target to a specific video. Briefly, this algorithms implements a variant of a Vickrey auction, named Generalized Second-Price auction [54] for which the winner (advertiser) pays the price of the second highest bid. Note that winning bids vary over time and targeted videos. In addition to the bid price, the algorithm also considers other parameters including the profile of the viewer watching the video, the advertiser's daily budget, etc.

In setting up these experiments, we faced several challenges to configure a successful campaign able to target a large number of ad views in the videos. Our initial trails were unsuccessful; we used a small daily budget of 50€ and the campaign had an unusual configuration, since it targeted very specific and relatively unpopular videos. To overcome this, we took advice from an online advertising expert to; i) increase the bidding prices per ad view up to 10-15€ (the recommended bid price for YouTube was 0.04-0.05€), ii) configure the video uploader's AdSense account to accept only the specific type of ads defined in the campaign, iii) configure different campaigns with different accounts that compete for the same videos (viewers), iv) to vary the pattern of views to the videos.

Having done so, we launched new experiments, whereby we targeted a set of videos from different IP addresses and different rates of views per IP address (between 10 and 70 views per IP address). In particular, the campaigns targeted 14 videos, using the **D**eterministic configuration of the probe. Of the 14 trails, 5 videos were able to attract ad views from the campaigns, meaning that we bid and won - in effect the ads were targeted to the uploaded videos, and watched by the probes, which are configured to view in full any ad target, as well as the video.

Table 6 summarizes the main characteristics of the view pattern configuration of these videos. Moreover, it shows the number of monetized ad views from the campaigns, as well as the number of counted views in the public view counter for the days of the experiment in which our ad was delivered. We observe that in all the cases the number of monetized views are larger than the number of counted views, i.e., views considered suspicious are removed from the public view counter, but monetized.

Our videos received a total of 301 ad views in 5 days.[7] In the case of Video 3 and 4, views were initially added to the bill of our advertiser account. However, 5 days after the first ad view was delivered, YouTube rightfully labeled the probe's activity as suspicious and suspended the video uploader account in AdSense. In addition, YouTube notified us via email of the suspension of the uploader's account due to suspicious activity . Finally, the ad views associated to fake views were removed from the advertiser account and 4.85€ refunded. We believe that the peculiar setup of the campaign, coupled with the aggressiveness of the experiment triggered alarms in YouTube's view audit system. In case of Video 4, we repeated this experiment twice obtaining the same result (AdSense account closed).

In the case of Video 1 and 2, 91 ad views were shown, for which we were charged 5.65€, whereas just 25 views were counted in the public view counter. Google indicates through its AdWords support website that *"If we find invalid clicks that have somehow escaped the automated detection in the past two months, we'll give you credit for these clicks"* [55]. In the case of Video 1 and 2, all the ad views were made more than 8 months before the conclusion of this work. Therefore we can consider that the probe's actions have gone unnoticed by Google's fraud detection algorithm.

In summary, we conclude that *YouTube uses a seemingly permissive view audit system to discount fake monetized views. This exposes advertisers to the risk of building their advertisement campaigns on unreliable statistics, and may make them initially burden the risk of fraud. Conversely, the public view counter is much more discriminative, demonstrating that YouTube has effective means to identify fake views. Our results also reveal that whenever the permissive threshold for the detection of fake monetized views is crossed, YouTube severely penalizes the uploader of the video by suspending her AdSense account, preventing the uploader from monetizing any of the videos associated to the suspended account.*

[7] Note that after finishing the experiments, these videos have received only 16 views in 8 months. Based on this, we have high certainty that our video ad was only viewed by the probes, and not by legitimate users.

	# IPs	Daily Views per IP	Monetized view counter	Public view counter
Video 1	1	10	31	18
Video 2	1	20	60	7
Video 3	8	10	178	147
Video 4	2	70	15 (17)	0

Table 6: Experiments configuration of videos attracting ads from our advertising campaigns. The reported numbers of monetized and public counted views correspond to the sum of views of the days in which ads were shown. The number 17 for Video 4, reflects the second trail of the experiment.

7. RELATED WORK

The research community has devoted an important amount of effort to the identification of malicious behaviors in online services and to the design of countermeasures to such behaviors [56–58]. Similarly to YouTube's fake view detection mechanism, most of the detection system designs rely on the IP address as the main id to track and identify malicious behaviors. Some examples of such mechanisms are the classical monitoring tools looking for sources of attacks, such as port scanning [59] and DDoS attacks [60], or the detection systems which counteract malicious users in P2P applications [61]. Only those systems requiring the user registration to gain access to the service, e.g., Online Social Networks, implement detection mechanisms that use both the IP address and the user id as basic units to detect inappropriate behaviors. For instance, Facebook traces the requests pattern from a given account, if it is unusual, the user is warned and if the behavior persists the account is closed [62].

More recently, the rapid proliferation of botnets and specialized bots to attack specific services has led the research community to work on the identification, characterization and elimination of botnets and bots [63–71]. Additionally, following a similar methodology to the one we use in this paper, Boshmaf et al. [72] and Bilge et al. [73] have developed their own automatic software to evaluate the effectiveness of the defenses of different social networks from different types of attacks such as user impersonation.

In the field of fraud detection and mitigation in online advertising, most of the literature focuses on traditional type of ads such as search or display ads. In this case, the fraud problem is referred to as "click fraud" since the fraudulent activity is associated to fake clicks on ads, typically performed from bots. Metwally et al. [13] present an early study in which they use the IP address as the parameter to detect coalition of fraudulent users or *fraudsters*. In a more recent work, Li et al. [74] propose to analyze the paths of ad's redirects and the nodes found in the content delivery path to identify malicious advertising activities. Furthermore, Stone-Gross et al. [14] managed to get access to a command-and-control botnet used for ad fraud in which the bot master sends commands with fake referrers. On a complementary work, Miller et al. [75] study the behavior of two clicking robots: Fiesta and 7cy. Moreover, Dave et al. [76] design an algorithm to identify click fraud from the advertiser perspective; to design this algorithm, the authors propose to measure different aspects of the user behavior in the advertiser webpage such as the mouse movements or the time spent in the website. Based on their initial work, the same authors propose, implement and test ViceROI [12], a solution to discount fake clicks from ad networks. The basis of ViceROI detection algorithm is the fact that click-spammers will lead to a higher ROI (Return of Investment) than a legitimate

publisher, as the authors claim that a realistic ROI is difficult to obtain with robots. Fraudsters can perform other types of attacks in the online advertisment ecosystem. For example, Snyder et al. [77] present a study of the prevalence of fraud in affiliate marketing networks. These networks encourage publishers to promote online shops on their webpages, receiving later some amount of money if the user, that has clicked in the promoted link, makes a purchase in that online shop. Fraudsters setup a webpage forcing user's browser to click the promoted link. Later if that user buys an item in the promoted online shop, the fraudster will receive credit for it. Another example is presented by Thomas et al. [78]. They study the impact of *ad injection* in the advertisement ecosystem. They identify mainly Chrome extensions and Windows binaries responsible of this source of fraud. Finally, Meng et al. [79] present a new type of attack taking advantage of the different prices paid depending on the user's profile. They claimed that fraudsters could increase their revenue as much as 33% by "polluting" user's profiles with high paying preferences.

All the above works establish a very solid basis for the design of tools to mitigate fraud associated to traditional ads. However, they are (in general) not applicable to fraud associated to video ads due to the different nature of video ads and click-based ads. To the best of the authors knowledge, there is only a very recent study that analyzes fraud in video ads [16]. The authors of this study use traces from a video platform in China to identify statistically outlying video viewing patterns and, based on these observations, suggest a fake view detection algorithm built on parameters such as the number of views made from an IP address to a video or the number of different IP addresses watching a given video. Unfortunately, as the authors acknowledge, they do not count with a ground truth dataset to validate their designed solution as legitimate views cannot be distinguished from fake ones in their dataset. In contrast to this work, our study focuses on five major video portals, including YouTube, the most important video platform worldwide, and pursues a different goal. We propose a methodology to generate ground truth scenarios so that we can evaluate the performance (and unveil basic functionality principles) of different video portals' audit systems for both the number of counted and monetized views. As our methodology is extensible to other video platforms, the authors from [16] could use it to validate their proposed solution in their considered video platform.

8. ETHICAL ASPECTS AND FEEDBACK FROM THE INDUSTRY

While, to the best of our knowledge, there is not a methodology that could obtain the results presented in the paper without any effect on advertisers and/or video portals, we would like to highlight that the experiments performed in this paper have an extremely low impact on both video portals and advertisers.

Video portals have to dedicate storage resources to host our videos and bandwidth to serve views to the probes. However, the number of videos uploaded and views generated in the experiments is very small (negligible in comparison with the volume managed by these portals) and therefore has practically no impact on the operation of the services.

Some advertisers have lost money during the experiments by having their ads shown in the videos viewed by the probes. However, based on the reported revenue by Google AdSense accounts associated to the videos, we can confirm that the total monetary losses produced by our experiments for advertisers are estimated to be lower than 6€. These losses are distributed across all those advertisers having their ads exposed in the videos, and thus the individual economical impact on each of them is negligible.

In addition, we would like to highlight that we have not received any payments while running these experiments, and all the statistics we report, were retrieved from the YouTube Analytics channel page, Google AdWords page and the Dailymotion Publisher page.

Finally, we have reported our findings to YouTube and Dailymotion. YouTube has contacted us via email, stating that they recognize the validity of our results, and have not indicated any ethical concerns with our methodology. We plan to present the Dailymotion feedback and explanations, once we receive them. Advertisers have also reacted positively to our research after the technical report of this work attracted media attention, and was published by several organizations, including the Financial Times [81], The Guardian [82], Business Insider [83] or the BBC [84]. Major advertising companies and associations have welcomed the work, without raising any ethical concerns. Based on our results, they have urged Google and other major players to increase their transparency, when accounting for advertising expenditure, as well as to more effectively address the problem of fraud in online advertising [30, 31].

9. CONCLUSIONS AND FUTURE WORK

To the best of our knowledge, this work is the first one to propose a set of tools to monitor and audit the view audit systems of online video portals, and enable independent and external parties to measure their performance. The application of the tools and methodology to the view counting behavior of five different video portals has highlighted some interesting observations. We find that only YouTube deploys a sufficiently discriminative view audit systems for the public view counter. All the other portals studied are susceptible to very naïve view inflation attacks. Clearly, this raises a problem for users with regard to the accuracy of the numbers that are reported by these portals.

A more careful analysis of YouTube's view audit systems has revealed that it is susceptible to attacks that introduce some randomness to the viewer behavior, including the use of multiple User-Agents, Referrers, multiple IP addresses, or machines within a large NATed network. These are traits that a knowledgeable attacker would be able to configure easily, and we have been reported to be common in large scale attacks using botnets. YouTube is consistently more permissive in the counts for monetized views, when compared to the public view counters. Specifically, fake views are penalized and not counted by the public view counter, but can still be monetized, i.e., have paid for ads delivered in them, and counted in the video owner's monetized views. While YouTube is shown to strive to protect its users and clients, for example by reacting quickly when suspicious behavior is identified, we speculate that its setup seems to place an unnecessary burden of risk on advertisers. For example, fake views can be discounted equally for public and monetized counters, but they are not.

Finally, our analysis in this paper reinforce the call by industry for (i) consistent and independently measurable principles on how [Supply sources (SSPs/exchanges, ad networks, and publishers)] should identify and expunge fraudulent traffic and (ii) more efficient antifraud mechanisms. In future work, we intend to refine and better scale the tools, and methods developed here, and explore how to make them available to the wider community.

Acknowledgments

This work has been partially supported by the European Union through the FP7 METRICS (607728), H2020 TYPES (653449) and ReCRED(653417) Projects, the Spanish Ministry of Economy and Competitiveness through the DRONEXT project (TEC2014-54335-C4-2-R) and the Regional Government of Madrid though the BRADE Project (P2013/ICE-2958).

References

[1] The Interactive Advertising Bureau (IAB), "IAB internet advertising revenue report, 2014 full year results." http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2014.pdf. Last accessed January 2016.

[2] eMarketer, "As Barriers Tumble, Video Marketing Adoption Grows." http://www.emarketer.com/Article/Barriers-Tumble-Video-Marketing-Adoption-Grows/1010374, 2014. Last accessed January 2016.

[3] eMarketer, "Online Video Advertising Moves Front and Center." http://www.emarketer.com/Article/Online-Video-Advertising-Moves-Front-Center/1009886, May 2013. Last accessed January 2016.

[4] W. Luttrell, "Only The Buy-Side Can Solve Our Fraud Problem." http://www.adexchanger.com/data-driven-thinking/only-the-buy-side-can-solve-our-fraud-problem/, 2013. Last accessed January 2016.

[5] S. Vranica, "A 'Crisis' in Online Ads: One-Third of Traffic Is Bogus." http://online.wsj.com/news/articles/SB10001424052702304026304579453253860786362, 2014. Last accessed January 2016.

[6] A. Neal and S. Kouwenhoven, "Quantifying online advertising fraud: Ad-click bots vs humans," tech. rep., Oxford Bio Chronometrics, January 2015.

[7] Solve Media, "Solve Media Survey." http://news.solvemedia.com/post/74832974631/solve-media-bot-survey-2014, January 2014. Last accessed January 2016.

[8] G. Sloane, "Fraud Alert: Millions of Video Views Faked in Sophisticated New Bot Scam." http://www.adweek.com/news/technology/fraud-alert-millions-video-views-faked-sophisticated-new-bot-scam-156883, 2014. Last accessed January 2016.

[9] A. Kantrowitz, "Ad-Fraud Operation Fools Detection Companies, Nets Millions." http://adage.com/article/digital/ad-fraud-operation-fools-detection-companies-nets-millions/293929/, 2014. Last accessed January 2016.

[10] J. Kirk, "Malware campaign inflated views of pro-Russia videos." http://www.techworld.com.au/article/574002/malware-campaign-inflated-views-pro-russia-videos/, May 2015. Last accessed January 2016.

[11] ANA and White Ops, "The Bot Baseline: Fraud in Digital Advertising." https://www.ana.net/getfile/21853, December 2014. Last accessed January 2016.

[12] V. Dave, S. Guha, and Y. Zhang, "Viceroi: Catching click-spam in search ad networks," ACM CCS, 2013.

[13] A. Metwally, D. Agrawal, and A. El Abbadi, "Detectives: Detecting coalition hit inflation attacks in advertising networks streams," ACM WWW, 2007.

[14] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna, "Understanding fraudulent activities in online ad exchanges," ACM IMC, 2011.

[15] The Interactive Advertising Bureau (IAB), "Trustworthy Supply Chain: Anti-Fraud Working Group." http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/pr-091614, 2014. Last accessed January 2016.

[16] L. Chen, Y. Zhou, and D. M. Chiu, "Analysis and detection of fake views in online video services," *ACM TOMM*, vol. 11, 2015.

[17] C. Kang, " The real reasons why YouTube's 5 biggest stars became millionaires." https://www.washingtonpost.com/news/the-switch/wp/2015/07/23/how-these-5-youtube-stars-became-millionaires-and-why-you-wont-be-joining-them-anytime-soon/, 2015. Last accessed 8/10/2015.

[18] C. Tripputi, "Tubrosa threat drives millions of views to scammers' YouTube gaming videos." http://www.symantec.com/connect/blogs/tubrosa-threat-drives-millions-views-scammers-youtube-gaming-videos, 2015. Last accessed January 2016.

[19] YouTube Help, "Missing YouTube Views." https://support.google.com/youtube/answer/4646474?hl=en, 2015. Last accessed January 2016.

[20] C. Hoffberger, "YouTube strips Universal and Sony of 2 billion fake views." http://www.dailydot.com/news/youtube-universal-sony-fake-views-black-hat/, 2012. Last accessed January 2016.

[21] AdWords Help, "Cost-per-view (CPV)." https://support.google.com/adwords/answer/2472735?hl=en. Last accessed January 2016.

[22] TubeMogul, "Video Advertising Playbook." https://www.tubemogul.com/marketing/TubeMogul_Video_Ad_Playbook.pdf, 2014. Last accessed January 2016.

[23] Supreme Traffic Bot, "Traffic Generation & Automation, Made Easy...." http://www.supremetrafficbot.com/. Last accessed January 2016.

[24] The Interactive Advertising Bureau (IAB), "Anti-Fraud Principles and Proposed Taxonomy." http://www.iab.net/media/file/IAB_Anti_Fraud_Principles_and_Taxonomy.pdf, September 2014. Last accessed January 2016.

[25] "Viewbros." http://www.viewbros.com/. Last accessed January 2016.

[26] "QQTUBE." https://www.qqtube.com/. Last accessed January 2016.

[27] "Buildmyviews." http://www.buildmyviews.org. Last accessed January 2016.

[28] B. Elgin, M. Riley, D. Kocieniewski, and J. Brustein, "The Fake Traffic Schemes That Are Rotting the Internet." http://www.bloomberg.com/features/2015-click-fraud/, 2015. Last accessed January 2016.

[29] R. Fenton, "Has YouTube come of age for modern advertisers?." http://www.theguardian.com/media-network/2015/oct/05/youtube-brands-demand-views-transparency, 2015. Last accessed January 2016.

[30] R. Cookson, "WPP urges Google to tackle problem of fake ad views." http://www.ft.com/cms/s/0/f9da727c-6207-11e5-9846-de406ccb37f2.html, 2015. Last accessed January 2016.

[31] The Incorporated Society of British Advertisers (ISBA), "'Bots', YouTube and advertisers." http://www.isba.org.uk/news/2015/09/24/'bots'-and-youtube, 2015. Last accessed January 2016.

[32] P. Pfeiffenberger, "Keeping YouTube Views Authentic." http://googleonlinesecurity.blogspot.co.uk/2014/02/keeping-youtube-views-authentic.html, February 2014. Last accessed January 2016.

[33] S. Dredge, "Google goes to war on 'fraudulent' YouTube video views." http://www.theguardian.com/technology/2014/feb/05/youtube-fake-views-counts-google, 2014. Last accessed January 2016.

[34] "Youtube Bot Views." http://traffic-bots.com/youtube-bots/youtube-bot-views/. Last accessed January 2016.

[35] "YouTube Bot Views Proxies." https://listingdock.com/Computer-Software/2600/YouTube-Bot-Views-Proxies-Random-Referrer. Last accessed January 2016.

[36] Sysomos, "A Look Inside Online Video Engagement - Part I." https://www.sysomos.com/reports/video, 2009. Last accessed January 2016.

[37] "Online Video market share in the Alexa top 1M." http://www.datanyze.com/market-share/online-video/, 2015. Last accessed January 2016.

[38] Nielsen, "May 2012 - Top U.S. Online Video Sites." http://www.nielsen.com/us/en/insights/news/2012/may-2012-top-u-s-online-video-sites.html. Last accessed January 2016.

[39] Statista, "Leading internet multimedia portals in the United States in August 2014, based on market share of visits." http://www.statista.com/statistics/266201/us-market-share-of-leading-internet-video-portals/, 2014. Last accessed January 2016.

[40] "Comparison of video hosting services." https://en.wikipedia.org/wiki/Comparison_of_video_hosting_services, 2015. Last accessed January 2016.

[41] D. Gayle, "YouTube cancels billions of music industry video views after finding they were fake or 'dead'." http://www.dailymail.co.uk/sciencetech/article-2254181/YouTube-wipes-billions-video-views-finding-faked-music-industry.html, 2012. Last accessed January 2016.

[42] YouTube Help, "Views report." https://support.google.com/youtube/answer/1714329. Last accessed January 2016.

[43] YouTube, "Ad Performance report for partners." https://support.google.com/youtube/answer/2423005?hl=en. Last

accessed January 2016.
[44] Vimeo, "Vimeo - Get Advanced Statistics." http://vimeo.com/stats. Last accessed January 2016.
[45] "Selenium webdriver." http://docs.seleniumhq.org/projects/webdriver/.
[46] "Squid proxy server." http://www.squid-cache.org/.
[47] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman, "Planetlab: An overlay testbed for broad-coverage services," *ACM SIGCOMM CCR*, vol. 33, July 2003.
[48] A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao, "Youtube everywhere: Impact of device and infrastructure synergies on user experience," ACM IMC, 2011.
[49] L. Chen, Y. Zhou, and D. M. Chiu, "Fake view analytics in online video services," ACM NOSSDAV, 2013.
[50] S. Englehardt, D. Reisman, C. Eubank, P. Zimmerman, J. Mayer, A. Narayanan, and E. W. Felten, "Cookies that give you away: The surveillance implications of web tracking," ACM WWW, 2015.
[51] M. P. Collins, T. J. Shimeall, S. Faber, J. Janies, R. Weaver, M. De Shon, and J. Kadane, "Using uncleanliness to predict future botnet addresses," ACM IMC, 2007.
[52] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," *ACM SIGCOMM CCR*, vol. 36, Aug. 2006.
[53] M. Marciel, R. Cuevas, A. Banchs, R. Gonzalez, S. Traverso, M. Ahmed, and A. Azcorra, "Understanding the detection of fake view fraud in video content portals," *CoRR*, vol. abs/1507.08874, 2015.
[54] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords," *American Economic Review*, vol. 97, no. 1, 2007.
[55] AdWords Help, "About invalid traffic." https://support.google.com/adwords/answer/2549113?ctx=tltp&hl=en. Last accessed January 2016.
[56] F. Soldo, K. Argyraki, and A. Markopoulou, "Optimal source-based filtering of malicious traffic," *IEEE/ACM Trans. Netw.*, vol. 20, no. 2, 2012.
[57] Z. Chen, C. Ji, and P. Barford, "Spatial-temporal characteristics of internet malicious sources," IEEE INFOCOM, 2008.
[58] S. Venkataraman, A. Blum, D. Song, S. Sen, and O. Spatscheck, "Tracking dynamic sources of malicious activity at internet scale," in *NIPS*, Curran Associates, Inc., 2009.
[59] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," *J. Comput. Secur.*, vol. 10, July 2002.
[60] T. Peng, C. Leckie, and K. Ramamohanarao, "Proactively detecting distributed denial of service attacks using source ip address monitoring.," in *NETWORKING*, vol. 3042 of *Lecture Notes in Computer Science*, 2004.
[61] R. Cuevas, M. Kryczka, R. González, A. Cuevas, and A. Azcorra, "Torrentguard: Stopping scam and malware distribution in the bittorrent ecosystem," *Comput. Netw.*, vol. 59, Feb. 2014.
[62] M. Gjoka, M. Kurant, C. Butts, and A. Markopoulou, "Practical recommendations on crawling online social networks," *IEEE JSAC*, vol. 29, October 2011.
[63] A. Karasaridis, B. Rexroad, and D. Hoeflin, "Wide-scale botnet detection and characterization," USENIX HotBots, 2007.
[64] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming botnets: Signatures and characteristics," *ACM SIGCOMM CCR*, vol. 38, Aug. 2008.
[65] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots + machine learning," ACM SIGIR, 2010.
[66] J. Zhang, R. Zhang, Y. Zhang, and G. Yan, "On the impact of social botnets for spam distribution and digital-influence manipulation," CNS, 2013.
[67] O. Thonnard and M. Dacier, "A strategic analysis of spam botnets operations," ACM CEAS, 2011.
[68] K. Thomas and D. Nicol, "The koobface botnet and the rise of social malware," in *MALWARE, 2010*, 2010.
[69] G. Stringhini, T. Holz, B. Stone-Gross, C. Kruegel, and G. Vigna, "Botmagnifier: Locating spambots on the internet.," in *USENIX Security Symposium*, 2011.
[70] G. Stringhini, O. Hohlfeld, C. Kruegel, and G. Vigna, "The harvester, the botmaster, and the spammer: On the relations between the different actors in the spam landscape," ACM ASIA CCS, 2014.
[71] L. Bilge, D. Balzarotti, W. Robertson, E. Kirda, and C. Kruegel, "Disclosure: Detecting botnet command and control servers through large-scale netflow analysis," ACM ACSAC, 2012.
[72] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The socialbot network: When bots socialize for fame and money," ACM ACSAC, 2011.
[73] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," ACM WWW, 2009.
[74] Z. Li, K. Zhang, Y. Xie, F. Yu, and X. Wang, "Knowing your enemy: Understanding and detecting malicious web advertising," ACM CCS, 2012.
[75] B. Miller, P. Pearce, C. Grier, C. Kreibich, and V. Paxson, "What's clicking what? techniques and innovations of today's clickbots," Springer-Verlag DIMVA, 2011.
[76] V. Dave, S. Guha, and Y. Zhang, "Measuring and fingerprinting click-spam in ad networks," *ACM SIGCOMM CCR*, vol. 42, October 2012.
[77] P. Snyder and C. Kanich, "No please, after you: Detecting fraud in affiliate marketing networks," WEIS, 2015.
[78] K. Thomas, E. Bursztein, C. Grier, G. Ho, N. Jagpal, A. Kapravelos, D. McCoy, A. Nappa, V. Paxson, P. Pearce, N. Provos, and M. A. Rajab, "Ad injection at scale: Assessing deceptive advertisement modifications," in *IEEE S&P*, 2015.
[79] W. Meng, X. Xing, A. Sheth, U. Weinsberg, and W. Lee, "Your online interests: Pwned! a pollution attack against targeted advertising," ACM CCS, 2014.
[80] S. S. Krishnan and R. K. Sitaraman, "Understanding the effectiveness of video ads: A measurement study," ACM IMC, 2013.
[81] R. Cookson, "Google charges for YouTube ads even when viewed by robots." http://www.ft.com/cms/s/0/53ac3fd0-604e-11e5-a28b-50226830d644.html, 2015. Last accessed January 2016.
[82] B. Quinn, " Google charges advertisers for fake YouTube video views, say researchers ." http://www.theguardian.com/technology/2015/sep/23/google-advertisers-fake-youtube-video-views-adwords-bot, 2015. Last accessed January 2016.
[83] J. D'Onfro, "Google charges for YouTube ads even when it thinks a robot viewed them, says study." http://uk.businessinsider.com/google-charges-advertisers-for-robot-views-2015-9, 2015. Last accessed January 2016.
[84] K. Rawlinson, "Google 'charges for YouTube adverts viewed by bots'." http://www.bbc.com/news/technology-34335971, 2015. Last accessed January 2016.