POLITECNICO DI TORINO
Repository ISTITUZIONALE

Characterizing Web Pornography Consumption from Passive Measurements

Original
Characterizing Web Pornography Consumption from Passive Measurements / Morichetta, A., Trevisan, M., Vassio, L.. - ELETTRONICO. - 11419:(2019), pp. 304-316. (20th International Conference on Passive and Active Network Measurement Puerto Varas, Chile March 27-29, 2019) [10.1007/978-3-030-15986-3_20].

Availability:
This version is available at: 11583/2728556 since: 2019-03-15T18:38:09Z

Publisher:
Springer International Publishing

Published
DOI:10.1007/978-3-030-15986-3_20

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-030-15986-3_20

(Article begins on next page)

18 November 2026

Characterizing web pornography consumption from passive measurements

Andrea Morichetta, Martino Trevisan, Luca Vassio
Politecnico di Torino
first.last@polito.it *

Abstract. Web pornography represents a large fraction of the Internet traffic, with thousands of websites and millions of users. Studying web pornography consumption allows understanding human behaviors and it is crucial for medical and psychological research. However, given the lack of public data, these works typically build on surveys, limited by different factors, e.g., unreliable answers that volunteers may (involuntarily) provide.
In this work, we collect anonymized accesses to pornography websites using HTTP-level passive traces. Our dataset includes about 15 000 broadband subscribers over a period of 3 years. We use it to provide quantitative information about the interactions of users with pornographic websites, focusing on time and frequency of use, habits, and trends. We distribute our anonymized dataset to the community to ease reproducibility and allow further studies.

Keywords: Passive Measurements; Web Pornography; Adult Content; User Behaviour; Network Monitoring.

1 Introduction

Pornography and technology have enjoyed a close relationship in the last decades, with technology hugely increasing the capabilities of the porn industry. From the limited market reachable through public theatres, the introduction of the videocassette recorder in the 1970s abruptly changed the way of accessing pornography, allowing to access pornography in the privacy and comfort of each individual home. Then, the birth of cable networks and specialty channels in the 1990s, allowed a further step towards accessibility and privacy, giving the possibility to retrieve content directly from home. Finally, the Internet revolutionized again the market, guaranteeing direct desktop delivery to every individual with a connection, interactivity through forums and webcams, free content and, at the same time, anonymity. In 2017, the most used pornographic platform in the world (Pornhub, according to Alexa ranking)[1], claims 80 million daily accesses

* The research leading to these results has been funded by the Vienna Science and Technology Fund (WWTF) through project ICT15-129 (BigDAMA) and the SmartData@PoliTO center for Big Data technologies.

[1] As of October 17^{th}, 2018 www.alexa.com/topsites

to its website.[2] The importance of Internet pornography as a prevalent component of popular culture and the need of its study has been recognized for a long time [6].

In this work we do not attempt to classify Internet pornography. Rather, we refer to the term web pornography (WP) to any online material that, directly or indirectly, *seeks to bring about sexual stimulation* [4]. Therefore, the term *pornographic website* is here used to describe services that provide actual pornographic videos, sell sex related merchandise, help in arranging sexual encounters, etc. We refer only to adult pornography websites and we do not advocate the inclusion of child pornography websites in our research, thus this paper has no application whatsoever to child pornography. The word pornography, in the context of this article, refers exclusively to legal content in the territories of EU and USA.

Through the years, WP has been the subject of many studies, that aimed at describing how people make use of it, or pinpointing eventual pathological situation correlated to an excessive use. However, such works typically come from the medical and psychology communities, and are based on surveys that cover a very small number of volunteers. Moreover, previous studies [12,10] report that people tend to lie, either consciously or unconsciously, when answering to private-life concerning surveys, especially about sexuality; there are people who declare more accesses than real (e.g., to show to be uninhibited) and others who understate the actual consumption, fearing social blame. Both these behaviours, called social desirability biases, and egosyntonic/egodystonic feelings (i.e., being or not in accordance with their self-image) make surveys less reliable with respect to other sources of information.

In contrast to previous works, in this study we investigate WP by means of passive network measurements, collected from about 15 000 broadband subscribers over a period of 3 years. MindGeek, a company operating many popular pornographic websites, switched to encryption only in April 2017, being the first big player of WP industry to adopt HTTPS.[3] As such, the vast majority of WP portals used plain-text HTTP up to March 2017, allowing us to leverage HTTP-level measurements, and obtain detailed results of WP consumption. Using recent advances in data science, we extract only user actions to WP portals from a deluge of HTTP data.

The main contributions of this paper are:

- We provide a thorough characterization of WP consumption leveraging measurements from 15 000 broadband subscribers over a period of 3 years.
- We show how users moved to mobile devices through the years, even if the time spent onWP remains constant.
- We show that typical WP sessions last less than 15 minutes, with users rarely accessing more than one website. Less than 10% of users consume

[2] `www.pornhub.com/insights/2017-year-in-review`

[3] `www.washingtonpost.com/news/the-switch/wp/2017/03/30/porn-websites-beef-up-privacy-protections-days-after-congress-voted-to-let-isps-share-your-web-history`

WP more than 15 days in a month, and repeated use within a single day is very sporadic.
– We release our dataset to the community in anonymized form for further investigation [1]. To the best of our knowledge, this is the only public datasets that includes WP accesses from regular Internet users.

The employed metrics are taken from the surveys reported in medical literature and from WP portal reports that we use throughout the paper. We restrict our analysis only to those that we were able to verify given our data. Our results enhance the visibility and understanding of those topics, and give a less mediated overview of users behaviors, mostly confirming what emerges from medical surveys.

The remainder of the paper is organized as follows: Section 2 summarizes related work. Section 3 describes data collection, processing and privacy issues, while Section 4 presents the results. Finally, Section 5 concludes the paper.

2 Related Work

Most previous works that investigate the interaction between users and WP leverage the information contained in surveys proposed to groups of volunteers. Vaillancourt-Morel *et al.* [18] examine the potential presence of different profiles of pornography users and their relation with sexual satisfaction and sexual dysfunction. The investigation is conducted over a poll that involved 830 adults, and they group users' behavior in three clusters. Daspe *et al.* [5] investigate the relationship between frequency of WP consumption and the personal perception of this behavior, pointing out that often there are strong discrepancies. Another analysis of the phenomenon is provided by Grubbs *et al.* [9], where the analysis is conducted over two participants sets, divided in students and adults, showing that moral scruples can infect the self-impression over their consumption. Wetterneck *et al.* [14] propose a critical analysis of WP, showing the various limitations of the state of the art of studies that assessed online pornography usage, concerning its definition, consumption, and the variability of its measurements.

Fewer works used network measurements to study WP. Tyson *et al.* [17] extract trends and characteristics in a major adult video portal (YouPorn) by analyzing almost 200k videos, together with meta-data such as page content, ratings and tags. In a similar direction, Mazieres *et al.* [11] produce and analyze a semantic network of WP categories, extracted from the portal xHamster, in order to find which are the most dominants and if they are actually meaningful. Ortiz *et al.* [13] study a Chilean websites containing human images and classify them in normal, porno and nude, with the objective of automatically discovering WP websites. Finally, Coletto *et al.* [3] study users' activity in social networks related to WP, in order to extract information about the seclusion of those communities with respect to the rest of the population and their characteristics in terms of age and habits and gender. To the best of our knowledge, we are the

first to use passive measurements to study the behavior of users accessing web pornography.

3 Measurements

3.1 Data Collection

In this work, we rely on network measurements coming from passive monitoring of a population of broadband subscribers over a period of 3 years (from March 2014 to March 2017). We have instrumented a Point-of-Presence (PoP) of a European ISP, where $\approx 10\,000$ ADSL and $\approx 5\,000$ FTTH customers are aggregated. ADSL downlink capacity is 4–20 Mbit/s, with uplink limited to 1 Mb/s. FTTH users enjoy 100 Mb/s downlink, and 10 Mbit/s uplink. Each subscription refers to an installation, where users devices (PCs, smartphones, etc.) connect via WiFi or Ethernet through a home gateway. Important to our analysis, the ISP provides each customer a fixed IP address, allowing us to track her over time. Nevertheless, a small fraction of customers abandoned the ISP during the observation period, and few new ones joined. All ADSL customers are residential customers (i.e., households), while a small number of business customers exist among the FTTH customers.

To gather measurements we use Tstat [15], a passive meter that collects rich per-flow summaries, with hundreds of statistics regarding TCP/UDP connections issued by clients. Beside, Tstat includes a DPI module that creates log files containing details about observed HTTP transactions. For each transaction, it records the URL, a client identifier as well as other HTTP headers of requests and responses. Our measurements are based on the inspection of HTTP headers, and, as such, neglect all encrypted traffic. However, no big WP portal used encryption at the time our dataset was collected. Generated log files are copied to our back-end servers with a daily frequency. Data is stored on a medium-sized Hadoop cluster to allow scalable processing. All processing is done using Apache Spark and Python. The stored data covers 3 years of measurements, totaling 20.5 TB of compressed and anonymized flow logs (around 138 billion records).

3.2 Definition of user and its limitations

Our PoP is located at the Broadband Remote Access Server (BRAS) level. Each subscription is identified by a unique and fixed IP address. However, subscriptions typically refer to households where potentially more than one person surf the Internet sharing the same public IP address. As such, relying on the client IP to identify a user would not be precise enough to study habits and behavior. Thus, in our work we define a *user* as the concatenation of the client IP address and the user-agent as extracted from the corresponding HTTP header. Note that with this definition a single person may appear multiple times with different identifiers if she uses multiple devices or her device incurs software updates that modify the user agent string. Analyses are thus performed on a per-browser

fashion – i.e., each user-agent string observed in a household. Privacy requirements limit any finer granularity.

The evaluated dataset includes only a regional sample of households in a single country. Users in other regions may have diverse browsing habits. Equally, mobile devices have been monitored only while connected to home WiFi networks. As such, our quantification of browsing on mobile terminals is actually a lower-bound, since visits while connected on mobile networks are not captured.

3.3 Data filtering and session definition

Starting from a HTTP-level dataset, we need to filter only entries referring to WP websites. Studying innovative methodologies to automatically isolate traffic towards particular services is out of the scope of this work. We employ a blacklist based approach to perform classification. We build on public available lists, achieving robustness by combining three different sources.[4] These three lists provide a set of domain names that offer different WP content (ranging from video streaming to thematic forums). To avoid false positives, we consider only those domain names contained in at least two over three lists. We come up with 310 252 unique entries, arranged over 460 top-level domains.

After filtering entries referring to WP websites, we perform a further step to identify *sessions* of continuous activity. To this end, we group data by user, and process HTTP transactions by start time. We then identify session as follows: when a user accesses a pornographic website we open a new session and account to this all subsequent entries to WP websites. We terminate a session if we do not observe any entry to WP for a period of 30 minutes. While defining a browsing session is complicated [8], we simply consider a time larger than 30 minutes as an indication of the session end as it is often seen in previous works (e.g., [2]), and in applications like Google Analytics.[5]

3.4 User actions extraction

Subsequently, we further filter the dataset to isolate only those HTTP requests containing an explicit user action by the user. This step aims at isolating users' behavior discarding all HTTP traffic related to inner objects of webpages such as images, style-sheets, and scripts To this end, we implement the methodology described in our previous work [19] that builds a machine-learning model to pinpoint intentionally visited URLs (i.e., webpages) from raw HTTP traces. The followed strategy has as core module a supervised classifier, which is able to correctly recognize user actions in HTTP traces. It results to reach an accuracy of over 98%, and it can be successfully applied to different scenarios, including smartphone apps [20].

[4] `www.shallalist.de/categories.html`, `www.similarweb.com`, and `dsi.ut-capitole.fr/blacklists/index_en.php`

[5] `support.google.com/analytics/answer/2731565?`

In total, after the extraction, we have 58 million user actions/visited webpages towards 59 989 different adult domains. We observe an average of 13 261 different WP users per month. For each user, we determine information about used OS, browser, and if the device was a PC, a smartphone, or a tablet. These information are extracted from the user-agent of the original HTTP request at the time of the capture, using the Universal Device Detection library.[6] We made these data available to the community in anonymized form to guarantee reproducibility of our results and further investigations [1]. In the remainder of the paper, we only take into account user actions, to which we simply refer with the term *visited webpages*.

3.5 Privacy and ethical concerns

Passive measurements potentially expose information which may threaten users privacy. As such, our data collection program has been approved by the partner ISP and by the ethical board of our University. Moreover, this specific data analysis project was also subject to a privacy impact assessment that was done with the data protection officer of our institution.

We undertake several countermeasures to avoid recording any personally identifiable information. Before any storage, all client identifiers are anonymized using Crypto-PAn algorithm [7], and URLs are truncated to avoid recording URL-encoded parameters. Encryption keys are varied on a monthly basis, to avoid persistent users tracking. Sensitive information such as cookies and `Post` data are not monitored at all. Logs are stored in a secured data center in an encrypted format. We emphasize again that in our research we only refer to adult pornography websites, obtained through open datasets, referring exclusively to legal content in the territories of EU and USA.

4 Results

In this section, we report the most significant results emerging from our dataset. We first focus on the time dimension, showing the evolution of WP consumption from 2014 to 2017 in terms of quantity and device type. We then focus on users, characterizing duration and frequency of their WP use. Finally, we provide some figures about the popularity of services.

4.1 Usage trends

Our first analysis aims at describing WP consumption trends from 2014 to 2017. In Fig. 1a, we focus on the time spent on WP by monitored users. The blue (solid), red (dash-dot) and green (dashed) curves report, respectively, the 25^{th}, 50^{th} and 75^{th} percentiles of the total per-user daily time spent on WP, i.e., the sum of the duration of all the WP sessions. Curves are calculated only for active

[6] `github.com/piwik/device-detector`

users, i.e., users visiting at least one WP website during one day. Curves are not continuous, for the lack of data due to outages in out PoP. The outcome shows a rather stable trend over the observation period, with half of the users spending less than 18 minutes per day on WP; however almost 25% of users reaches 40 minutes of daily activity. These are day-wise statistics, and do not provide figures about the repeated use of WP across multiple days by the same user, as we will see later. Measuring the overall share of users accessing WP portals is not easy using our data, as a single identifier – the client IP address – identifies a broadband subscription, potentially shared by multiple users. However, we notice that every day 12% of subscribers access WP websites, and this value is constant across years. A further analysis on WP pervasiveness is given in Sec. 4.2.

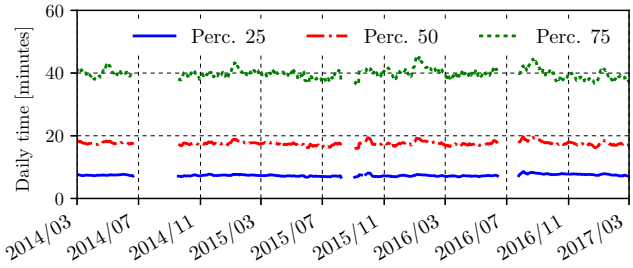
Those results can be used as a comparison with surveys statistics, fortifying or confuting what the participants declare. Vaillancourt-Morel *et al.* [18] study the characteristics of WP consumers. The majority of the chosen sample uses WP for recreation only, on average 24 minutes per week, a value consistent but slightly higher compared to our data.

Then, we investigate the evolution in device categories use (PCs, tablets and smartphones). We compute, for each device category, its share in terms of number of sessions. Fig. 1b shows the results. We notice that smartphones (blue surface), have largely increased their share from 27% to 42% at the expense of PCs. Tablets pervasiveness, reported in green, is instead rather constant. Not shown, the evolution of daily time spent with different devices did not changed too much throughout the years (see Sec. 4.2 for more details).

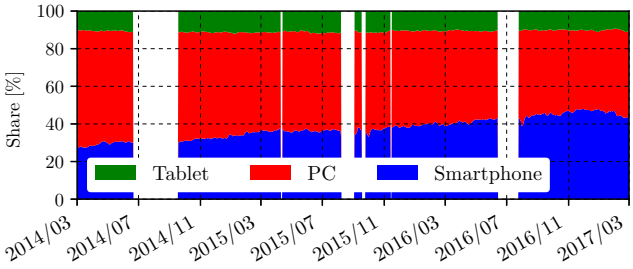
4.2 Usage frequency, duration and habits

For detailed analyses, we restrict to the last month of our dataset that does not include nor public holidays nor measurement outages, i.e., October 2016. We first characterize WP sessions in terms of duration. Fig. 2a shows the empirical cumulative distribution function (CDF) of session duration, expressed in minutes. The duration is larger for PCs than for tablet and smartphones. While most of the sessions are rather short, i.e., less than 15 minutes for PCs and 10 for smartphones, we observe sporadic longer sessions up to more than one hour. We now draw the attention on the number of webpages accessed during WP sessions, whose CDF is reported in Fig. 2b. Here the difference among devices is limited, with users accessing in median 5 or 6 webpages in a session, with 28% of them limited to one or two. However, in some cases tens of webpages are accessed. Similarly, in Fig. 2c we report the distribution of the number of *unique* websites accessed during a WP session. Results show that smartphone users tend to focus on a single WP website at a time (78% of sessions), while PC users are more prone to visit more websites. For all the devices, very few sessions include visits to 4 or more different websites. Finally, Fig. 2d reports the number of daily sessions for an active user. The figure shows that users hardly make repeated use of WP within a day, without differences among devices.

We next focus on the frequency of WP consumption by users over the month. In Fig. 3 we report the CDF of number of days of activity for WP users in the



(a) Per-user daily time spent on WP. 25^{th}, 50^{th}, and 75^{th} percentiles are shown.

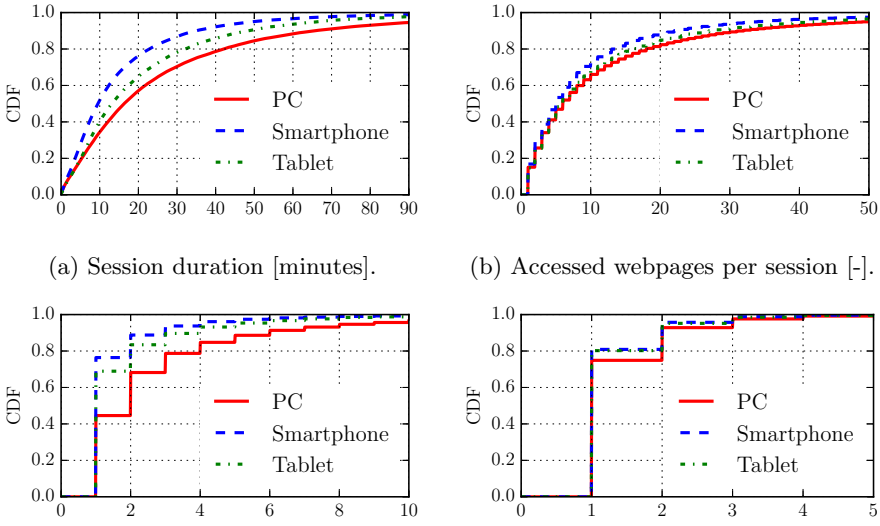


(b) Device category share for accessing WP.

Fig. 1: Usage trends from March 2014 to March 2017.

dataset. The figure indicates that the monthly frequency is generally low, with 76% of the users visiting WP 5 or less days in a month. Still, there are some users with a reiterate usage, with 8% of them consuming WP more than 15 days. These results confirm what is found by Daspe *et al.* [5], who show that the 73% of the participants to a survey access pornography no more than once or twice per week, and only 11% more than 5 times per week. Given the nature of our dataset, we cannot estimate the number of users *not* consuming WP. Still, an analysis of per-*subscription* traffic to WP is provided later in this section.

WP consumption also changes during different time of day. Fig. 4 provides the average percentage of sessions across the 24 hours of the day (red solid line). For ease of visualization, we start the x-axis from 4am, correspondent to the lowest value of the day. The two higher peaks are immediately after lunch time (2pm - 4pm) and after dinner (9pm - midnight). In addition to WP traffic, the figure also reports the overall trend considering all HTTP transactions, regardless their nature (dashed blue line). Comparing WP to total traffic, some differences are noticeable; the peaks do not overlap, and the latter is definitely more balanced over daylight hours. An hypothesis for those divergences may be related to the fact that accessing pornographic websites is likely to be a private and leisure



(a) Session duration [minutes].

(b) Accessed webpages per session [-].

(c) Accessed websites per session [-].

(d) Sessions per day for an active user [-].

Fig. 2: CDF of WP session characteristics, divided by device type.

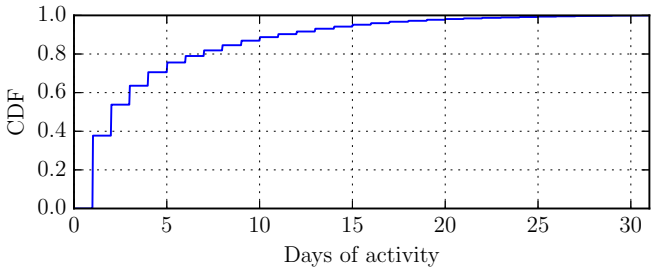


Fig. 3: Number of distinct days in which users consumed WP in a month.

activity confined to intimate moments. At a global scale, Pornhub service has found similar results.[7] The average session time reported by Pornhub for Italy is 9 minutes and 30 seconds, similar to what observed from our analysis. We also provide a breakdown across both hours and days of the week, with Fig. 5 showing the heat-map of the percentage variations from the gross weekly average (white color). Warmer tones register values below average, while colder ones show values above. Notice some clear diminishing traffic on Saturday evening

[7] See footnote 2

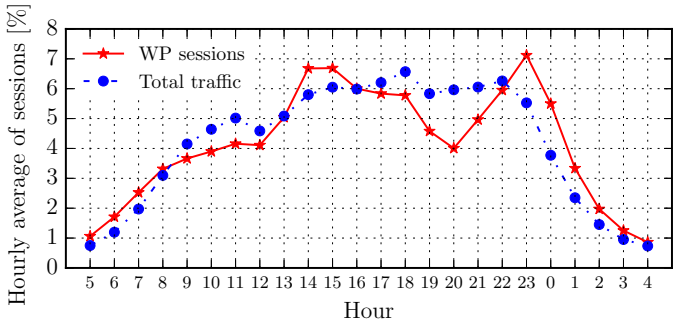


Fig. 4: Average hourly percentage of number of WP sessions and total traffic.

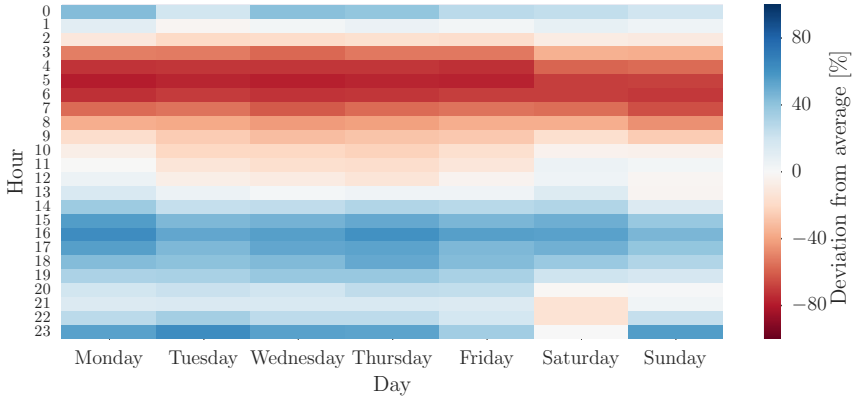


Fig. 5: Weekly breakdown of hourly WP usage. Heat-map of deviation from hourly average.

(7pm - midnight) and some increasing traffic on Saturday, Sunday and Monday morning (9am - 1pm). Indeed, many commercial activities are closed on Monday morning in the monitored country, perhaps influencing this behavior. Again, Pornhub data shows comparable results, with their heatmap having peaks of traffic in more or less the same time frames (2pm - 5pm) and (10pm - midnight). Considering the cumulative daily accesses, Mondays register the highest values and Saturdays the lowest.

Finally, we provide an overall picture about the fraction of all monitored subscribers accessing WP website. Although our dataset does not contain fine-grained details about WP pervasiveness, we can still show the fraction of subscriptions where at least one user accessed WP during our period of observation. In Fig. 6, the x-axis represents the 31 days of our reference month (being

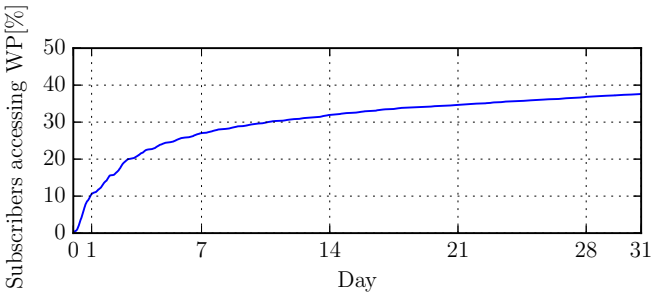


Fig. 6: Cumulative percentage of subscriptions accessing WP at different time in the trace.

day 1 October 1^{st}, 2016 and day 31 October 31^{st}, 2016), while y-axis reports the cumulative fraction of subscriptions that accessed at least one WP website. Considering a single day, less than 12% of subscriptions accessed WP, but this fraction raises to 27% after a week. At the end of the month it reaches 38%, meaning that more than one subscription over three generated traffic toward WP websites at least once in a month. For comparison, YouTube and Netflix are daily accessed by 45% and 3% of subscribers respectively. Considering social networks, 60% and 25% of subscribers contact Facebook and Instagram on a daily basis, respectively.[8]

4.3 WP websites at a glance

In this section, we briefly describe WP website popularity and pervasiveness. Similarly to the Internet global trend, the market is dominated by few big players. Looking at the Alexa rank,[9] three WP websites appear among the top-50, namely `pornhub.com`, `xvideos.com` and `livejasmin.com`, with the first one ranked 29th, just behind `linkedin.com`. Considering our dataset, we observe a similar situation, with over-the-top companies leading the rank. In Fig. 7 we show the percentage of users reached by the top-15 WP websites using bars (left-most y-axis), and the cumulative percentage of visits to these services (red line, right-most y-axis). In total users accessed 7 048 different websites during the entire month. The top-3 websites in our dataset match exactly Alexa rank, with `pornhub.com` being accessed by 34% of the users. Global tendencies are reflected in our top-15, with only 2 *omitted* websites as local representative of the monitored country. Considering the percentage of visited webpages, `pornhub.com` alone accounts for 14% of them, and the top-15 together approximately 63% of all WP visits. The percentage reaches 90% considering the top-204 websites, confirming the concentration of users around top services. Interestingly, very

[8] A deeper analysis can be found in our previous work [16].
[9] See footnote 1

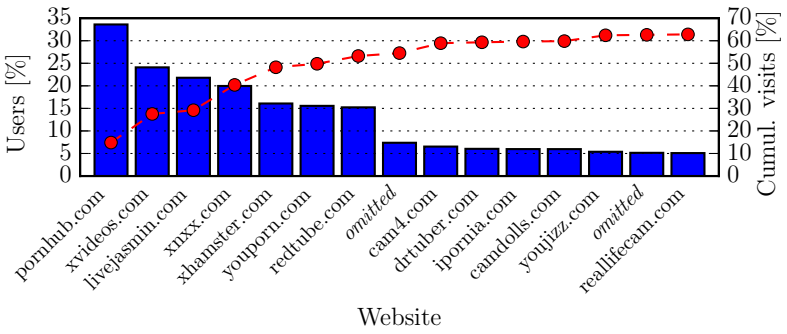


Fig. 7: Top-15 WP websites ranked according to percentage of users accessing them. Cumulative percentages of their visits with respect to all WP visits are also shown.

similar numbers hold for the overall traffic (including also non-WP websites), with top-15 accounting for 61% of traffic and 90% due to 195 websites.

Finally, we notice that 3 out of 15 WP websites of Fig. 7 belong to MindGeek, a company owning `pornhub.com`, `redtube.com`, `youporn.com`, and dozens of other websites.[10] MindGeek websites account for more than 20% of accesses in our dataset, making it a market leader. For comparison, the second website in terms of users and visits is `xvideos.com` (owned by WGCZ Holding), with less than half the users of MindGeek services, according to our data, suggesting a scenario where the ecosystem is lead by few big players in a dominant position.

5 Conclusion

In this paper we offered a quantitative analysis concerning web pornography consumption. To the best of our knowledge, we are the first to use network passive measurements to study the interactions of users with these services. We followed an exploratory approach on data, focusing on questions, topics and metrics typically analyzed in previous surveys and research works, e.g., frequency of fruition and the time spent on WP. We found interesting results, some typical of the observed population and others capable of confirming global trends.

Our results draw the attention to a large and active group of users, and may be helpful for researchers that study web services consumption and human behavior at large. The obtained outcomes can be checked and verified, thanks to the fact that we release our anonymized dataset. Furthermore, the chosen metrics allowed a comparison with outcomes of previously conducted surveys, and mostly confirmed their results.

[10] `https://goo.gl/UgLqAj`

References

1. Anonymized datasaset of visits to webpages belonging to web pornographic domains, obtained from network passive measurements. https://smartdata.polito.it/adult-clickstreams/ (2018)
2. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the world-wide web. Elsevier Computer Networks and ISDN systems **27**(6), 1065–1073 (1995)
3. Coletto, M., Aiello, L.M., Lucchese, C., Silvestri, F.: Adult content consumption in online social networks. Social Network Analysis and Mining **7**(1), 28:1–28:21 (2017)
4. Cornog, M.: Libraries, erotica, pornography. The Library Quarterly: Information, Community, Policy **61**(4), 457–459 (1991)
5. Daspe, M.E., Vaillancourt-Morel, M.P., Lussier, Y., Sabourin, S., Ferron, A.: When pornography use feels out of control: The moderation effect of relationship and sexual satisfaction. Journal of Sex & Marital Therapy **44**(4), 343–353 (2018)
6. Dilevko, J., Gottlieb, L.: Selection and cataloging of adult pornography web sites for academic libraries. The Journal of Academic Librarianship **30**(1), 36 – 50 (2004)
7. Fan, J., Xu, J., Ammar, M.H.: Crypto-pan: Cryptography-based prefix-preserving anonymization. Computer Networks **46**(2) (2004)
8. Fomitchev, M.I.: How google analytics and conventional cookie tracking techniques overestimate unique visitors. In: Proceedings of the 19th International Conference on World Wide Web. pp. 1093–1094 (2010)
9. Joshua, G., Joshua, W., Julie, E., Kenneth, P., Shane, K.: Moral disapproval and perceived addiction to internet pornography: a longitudinal examination. Addiction **113**(3), 496–506 (2014)
10. Lewontin, R.C.: Sex, lies, and social science. The New York Review of Books **42**(7), 24–29 (1995)
11. Mazires, A., Trachman, M., Cointet, J.P., Coulmont, B., Prieur, C.: Deep tags: toward a quantitative analysis of online pornography. Porn Studies **1**(2), 80–95 (2014)
12. Ochs, E.P., Binik, Y.M.: The use of couple data to determine the reliability of selfreported sexual behavior. The Journal of Sex Research **36**(4), 374–384 (1999)
13. Ortiz, F., Castaeda, V., Baeza-Yates, R., Verschae, R., del Solar, J.R.: Characterizing objectionable image content (pornography and nude images) of specific web segments: Chile as a case study. In: Web Congress, Latin American(LA-WEB). pp. 269–278 (2005)
14. Short, M.B., Black, L., Smith, A.H., Wetterneck, C.T., Wells, D.E.: A review of internet pornography use research: Methodology and content from the past 10 years. Cyberpsychology, Behavior, and Social Networking **15**(1), 13–23 (2012)
15. Trevisan, M., Finamore, A., Mellia, M., Munafo, M., Rossi, D.: Traffic analysis with off-the-shelf hardware: Challenges and lessons learned. IEEE Communications Magazine **55**(3), 163–169 (March 2017)
16. Trevisan, M., Giordano, D., Drago, I., Mellia, M., Munafo, M.: Five years at the edge: Watching internet from the isp network. In: Proceedings of the 14th International Conference on Emerging Networking EXperiments and Technologies. pp. 1–12. CoNEXT '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3281411.3281433, http://doi.acm.org/10.1145/3281411.3281433
17. Tyson, G., Elkhatib, Y., Sastry, N., Uhlig, S.: Measurements and analysis of a major adult video portal. ACM Trans. Multimedia Comput. Commun. Appl. **12**(2), 35:1–35:25 (2016)

18. Vaillancourt-Morel, M.P., Blais-Lecours, S., Labadie, C., Bergeron, S., Sabourin, S., Godbout, N.: Profiles of cyberpornography use and sexual well-being in adults. The journal of sexual medicine **14**(1), 78–85 (2017)
19. Vassio, L., Drago, I., Mellia, M.: Detecting user actions from http traces: Toward an automatic approach. In: 2016 International Wireless Communications and Mobile Computing Conference (IWCMC). pp. 50–55 (2016)
20. Vassio, L., Drago, I., Mellia, M., Houidi, Z.B., Lamali, M.L.: You, the web, and your device: Longitudinal characterization of browsing habits. ACM Transactions on the Web **12**(4), 24:1–24:30 (2018)