Tales from the Porn: A Comprehensive Privacy Analysis of the Web Porn Ecosystem

Pelayo Vallina
IMDEA Networks Institute /
Universidad Carlos III de Madrid
pelayo.vallina@imdea.org

Álvaro Feal
IMDEA Networks Institute /
Universidad Carlos III de Madrid
alvaro.feal@imdea.org

Julien Gamba
IMDEA Networks Institute /
Universidad Carlos III de Madrid
julien.gamba@imdea.org

Narseo Vallina-Rodriguez
IMDEA Networks Institute / ICSI
narseo.vallina@imdea.org

Antonio Fernández Anta
IMDEA Networks Institute
antonio.fernandez@imdea.org

ABSTRACT

Modern privacy regulations, including the General Data Protection Regulation (GDPR) in the European Union, aim to control user tracking activities in websites and mobile applications. These privacy rules typically contain specific provisions and strict requirements for websites that provide sensitive material to end users such as sexual, religious, and health services. However, little is known about the privacy risks that users face when visiting such websites, and about their regulatory compliance. In this paper, we present the first comprehensive and large-scale analysis of 6,843 pornographic websites. We provide an exhaustive behavioral analysis of the use of tracking methods by these websites, and their lack of regulatory compliance, including the absence of age-verification mechanisms and methods to obtain informed user consent. The results indicate that, as in the regular web, tracking is prevalent across pornographic sites: 72% of the websites use third-party cookies and 5% leverage advanced user fingerprinting technologies. Yet, our analysis reveals a third-party tracking ecosystem semi-decoupled from the regular web in which various analytics and advertising services track users across, and outside, pornographic websites. We complete the paper with a regulatory compliance analysis in the context of the EU GDPR, and newer legal requirements to implement verifiable access control mechanisms (*e.g.,* UK's Digital Economy Act). We find that only 16% of the analyzed websites have an accessible privacy policy and only 4% provide a cookie consent banner. The use of verifiable access control mechanisms is limited to prominent pornographic websites.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**.

IMC '19, October 21–23, 2019, Amsterdam, Netherlands

ACM ISBN 978-1-4503-6948-0/19/10…$15.00
https://doi.org/10.1145/3355369.3355583

1 INTRODUCTION

Pornographic (porn) websites are among the most visited and lucrative online services since the early days of the World Wide Web [92]. Pornhub, the most visited porn website according to Alexa's domain rank [4], had 33.5 Billion visits and was returned in 30.3 Billion web searches in 2018 [72]. MindGeek, Pornhubs' parent company, has reported over half a billion dollars of revenue in the 2015 fiscal year [56].

Modern privacy regulations like the EU General Data Protection Regulation (GDPR) [32] and California's Consumer Privacy Act (CCPA) [18] consider sexual information of an individual as highly sensitive data. All these privacy regulations require organizations with an online presence to request informed consent from users prior to any data collection [18, 32, 44, 46]. However, as in the case of regular websites, pornographic ones also integrate third-party components – *e.g.,* advertising and analytics libraries – with the capacity to track users' interaction with such services and, therefore, potentially infer a visitor's sexual orientation and preferences. The collection of this information, in addition to the absence of secure network protocols like HTTPS, could put at risk visitors of those websites, specially those connecting from countries where certain sexual orientations are prosecuted [16, 39, 42, 77, 84].

Despite the many research efforts that have taken place in the last decade to identify and quantify the presence and use of tracking technologies in the web, no study has deep dived yet into the privacy risks of sensitive websites, like pornographic ones. It is unclear, as a result, whether pornographic websites can pose a privacy risk to their visitors, and if they comply with the provisions set both by privacy regulations and by newer rules to control minor's access to adult content like the UK's Digital Economy Act [53]. In fact, anecdotal evidence suggests that there are significant differences between the third-party organizations operating in the porn and the regular web tracking industry [11] as large online ad networks such as Google Ads set strict constraints for porn-related publishers, prohibiting the advertising of adult-oriented products and services [39]. These restricting terms of services – possibly driven by fear of damaging their brand reputation – opened new market opportunities for other actors who have specialized in providing advertising and tracking technologies to adult sites. This context has created, as a result, a parallel ecosystem of third-party service providers in the porn ecosystem who has not been scrutinized by regulators, policy makers, and the research community.

In this paper, we develop and use a methodology to perform the first holistic analysis of pornographic websites from a privacy, transparency, and regulatory compliance perspective. Our main contributions are:

- We design a semi-supervised method to compile a representative sample of pornographic websites using publicly available resources (Section 3). After manually inspecting and removing false-positives, we identify 6,843 different pornographic websites.
- We develop and use a methodology to study the presence of third-party services in the porn ecosystem (Section 4). We compare the presence of third-party services present in pornographic websites with those embedded in the most popular web sites according to Alexa's rank. We find 3,673 third-party services embedded in porn websites, including companies specialized in the porn industry (*e.g.,* ExoClick), well-known advertising companies (*e.g.,* DoubleClick), analytics services (*e.g.,* Google Analytics), and domains associated to data brokers (*e.g.,* Acxiom). 84% of the third-party services embedded on pornographic websites do not appear in the most popular non-pornographic websites.
- We study the behavior of pornographic websites and the third-party tracking services embedded in them (Section 5). We find the presence of third-party HTTP Cookies in 72% of the analyzed pornographic websites, while 5% of them also use advance fingerprinting techniques like Canvas Fingerprinting to identify visitors uniquely. Interestingly, 91% of the scripts we found using canvas fingerprinting are not indexed by EasyList and EasyPrivacy [25].
- We quantify behavioral differences on porn websites depending on the user's location and jurisdictional area (Section 6). We conclude that the number of third-party services is quite stable across countries, yet there are regional third-party services that only operate in specific regions: *e.g.,*27 advertising and tracking services (ATS) only appear in Russia.
- We develop and validate a method to automatically analyze the transparency and regulatory compliance of pornographic websites (Section 7). Specifically, we study the presence of cookie consent banners, privacy policies, and age-verification mechanisms. Our analysis reveals a significant absence of privacy policies and consent forms across pornographic websites in spite of their sensitivity. This pattern holds even in regions with strict regulatory frameworks like the European Union: only 16% of the websites have privacy policies when accessed from a machine located at a EU state member. Finally, only 4% of the analyzed porn websites implement cookie consent forms.

Our study reveals a concerning lack of transparency in pornographic websites, despite the large presence of third-party trackers embedded in them and an increasing regulatory pressure. Therefore, we believe that our study will contribute to stress the importance of studying subsets of the world wide web that offer sensitive services and content in depth. This type of effort is not only needed to effectively inform the privacy debate, but also to promote user awareness.

2 BACKGROUND

The many privacy abuses inflicted by the online industry in the latest decades have motivated regulatory and legislative efforts to protect consumers' privacy and digital rights. New comprehensive data protection laws such as the European General Data Protection Regulation (GDPR) – which became effective on May 25th, 2018 [20] – aim to bring transparency to web services and to empower users with control over their personal identity in the web and beyond. In the case of online services, this objective is achieved by forcing companies with a digital presence to obtain explicit consent from any European visitor before collecting, processing, or sharing personal data on their sites. The GDPR, which will be complemented by the ePrivacy directive in 2019 [19, 30], also gives users the right to access, correct, and delete their personal data collected by online services, revoke their collection consent at any time, and object to automatic data processing.

In the context of pornographic websites, the GDPR imposes additional requirements and restrictions on data controllers due to the sensitive nature of their services. Article 9.1 [32] of the GDPR states that *"processing of data concerning a natural person's sex life or sexual orientation shall be prohibited."* Until the ePrivacy regulation becomes effective, the GDPR will require website owners to obtain explicit consent from users to install and use tracking methods, such as HTTP cookies, except when it is strictly required to provide a service requested by the user, to fulfill a legal mandate, or to carry out certain transmissions [31].

Similar regulatory efforts are taking place in other jurisdictions that used to have a traditional *laissez faire* attitude towards privacy. Notable examples are California's Consumer Privacy Act [18] (CCPA) (passed in June 2018), the Japanese Act on Protection of Personal Information [46] (effective since May 2017), and the Indian Personal Data Protection Bill of 2018 (PDP) [44]. All the aforementioned regulations classify and consider information regarding a user's sexual life and orientation as sensitive personal data that require special treatment.

2.1 Access Control in Pornographic Sites

For two decades, many laws failed to effectively prevent children from viewing pornography and other harmful materials on the Internet [38]. The 2017 Digital Economy Act [53] in the United Kingdom – which became effective on July 15th, 2019 [93] – aims to enforce the deployment of age verification mechanisms to block minors from accessing pornographic material. To comply with the new age-verification law in the UK, the industry designed and developed tools such as AgeID, a technology proposed by MindGeek [10] that is expected to become an industry standard [85]. A complementary effort to the aforementioned methods is the proposal made by the Association of Sites Advocating Child Protection (ASACP) [13]. This non-for-profit organization has created a Restricted-for-Adults (RTA) meta tag to assist parents to prevent their children from accessing pornographic material. The fact that there are companies from the online porn industry among the members [12] of this association is considered as a good example of collaboration between the porn industry and external organizations to increase safety and regulatory compliance.

Other regions in the world have followed more drastic and polemic strategies. The Russian government requires Pornhub users to login with a social network profile that is linked to their passport number [16, 94]. This measure has raised several ethical and privacy

concerns. World countries like most Middle East countries, India, Iran, or China actively ban, prosecute, and prohibit access to pornographic content altogether [68, 90]. The 2013 Anti-Pornography Act in Uganda prosecutes the broadcasting and trading of pornography [45], while the Anti-Homosexuality Bill Act in 2014 prosecutes LGBTI communities [77].

3 DATA COLLECTION AND METHOD

The first challenge in our study is compiling a representative list of pornographic websites. For that, we implement a semi-supervised approach that combines three different data sources and steps with varying levels of accuracy:

(1) We combine all the pornographic websites indexed by three websites specialized in aggregating, recommending, and classifying pornographic content [2, 3, 64]. This process provides us with 342 porn websites.
(2) We extract 22 websites classified as *Adult* sites by the Alexa's website categorization service [87].
(3) We look for websites indexed by Alexa's rank [4] (throughout 2018) that are potentially offering pornographic content by searching for keywords related to pornographic and adult content in their URLs (*e.g.,* "porn", "tube", "sex", "gay", "lesbian", "mature" and "xxx"). We find 7,735 websites matching these substrings.

The combination of these three methods allows us to identify 8,099 potential pornographic websites. However, the third keyword-based method introduces false positives if not done with care, since the chosen bag of words is not exclusively related to pornographic material (*e.g.,* PornTube offers pornographic content while YouTube does not). To identify and remove false positives, we implement a purpose-built crawler to download their content (DOM and screenshots) which are then manually inspected. In total, we find 1,256 false positives, many of which are because of unresponsive websites at the time of the crawl (we investigate below the stability of these domains). After this sanitation process, we obtain a corpus of 6,843 pornographic websites of various kinds, including websites hosting user-uploaded videos and live streaming content, or websites acting as proxies to pornographic material (*e.g.,* `pornsource.com`), among others. Finally, we use a reference dataset containing 9,688 popular non-pornographic websites[1] to study the commonalities and differences between sensitive pornographic websites and regular ones.

Popularity of Pornographic Websites: We use a longitudinal dataset containing the Alexa top-1M sites throughout 2018 as a proxy to measure the stability, popularity, and representativeness of our corpus of pornographic websites. Figure 1 shows the best and median rank value for each one of the identified porn websites, as well as the percentage of days each website was in the Alexa top-1M over the whole year.[2] We find that 1,103 websites (16%) were always present in the Alexa top-1M, and just 16 of them were always within the top-1K websites during the one-year period (*e.g.,* `pornhub.com`, `xvideos.com` or `livejasmin.com`).

[1]Websites extracted from Alexa's top-10K, in the 10th of January 2019.
[2]We consider their popularity for a whole year in order to account for any eventual bias caused by the Alexa ranking [81].

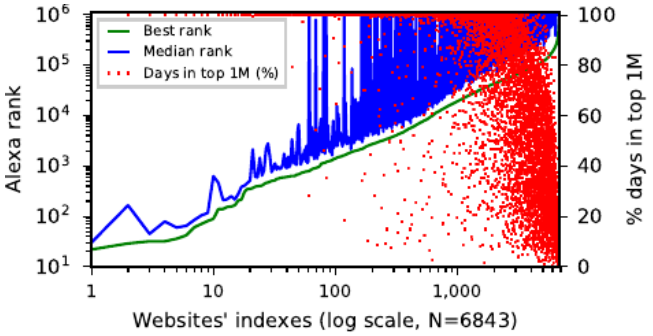

Figure 1: Best (green) and median (blue) Alexa rank for each pornographic website, and the percentage of days that each one of them were indexed in the top-1M throughout 2018. The pornographic websites are ordered in the x-axis by their best Alexa rank.

3.1 Web Crawlers

Our analysis and data collection workflow uses two complementary crawlers to study the behavior of pornographic websites as shown in Figure 2. First, we use a OpenWPM-based crawler to collect evidence of the behavior of each website and used tracking technologies, as well as the presence of third-party libraries and privacy consent forms. In both cases, we only crawl the landing page of websites so our study presents a lower-bound estimation of the privacy risks of pornographic websites as we do not interact with them beyond their landing page. Second, we use a Selenium-based crawler to automatically interact with each pornographic website to pass through the age verification mechanism (when available) and collect the privacy policy. We provide further details about each crawler and their purpose below.

OpenWPM: Rather than implementing yet another crawler, we use OpenWPM [27] because of its simplicity, stability, and the versatility of the features that it offers. OpenWPM is based on Firefox version 52 and allows (1) collecting all the HTTP and HTTPS requests and responses generated while crawling a website; and (2) detecting different tracking technologies, including advanced ones like canvas fingerprinting [27]. Nevertheless, we extend OpenWPM capabilities to analyze other other aspects of pornographic websites. First, we develop methods to extract the chain of requests caused by Real-Time Bidding (RTB) processes (*i.e.,* the inclusion chain [14]) to identify third-party services dynamically embedded in the target websites [14]. Specifically, we analyze the HTTP `Referrer` headers and remove those third parties not directly called by the publisher. Finally, we also enable mechanisms in OpenWPM to automatically record both HTTP cookies and cookie consent forms, so that we can estimate the transparency and regulatory compliance for each pornographic website (Section 7). We use the same browser session – *i.e.,* we do not close and open the browser between visits – for the duration of the crawling process, in order to be able to capture cookie synchronizations (Section 5.1.2).[3] It is also important to note

[3]We established a timeout of 120s for loading a website in order to prevent our crawlers from becoming stagnant.

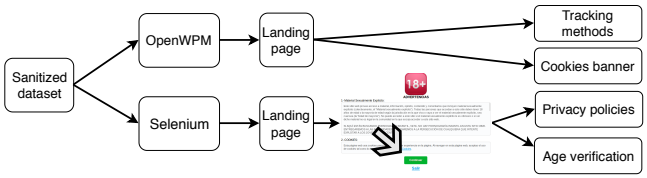

Figure 2: Workflow of our data collection.

that we only crawl each page once, giving us a lower bound on tracking activities [34].

Selenium: We implement a second purpose-built Selenium-based Chrome crawler to (1) detect and bypass age-verification mechanisms in pornographic sites; and (2) fetch their privacy policies when available. We separate this data collection process from OpenWPM crawls to avoid any instrumentation bias introduced by the need of interacting with each website in order to identify their privacy policies. To detect and quantify the support for age verification mechanisms, our crawler parses the landing page of a website and searches for floating elements and the words "Yes", "Enter", "Agree", "Continue" and "Accept" in 8 languages.[4] To eliminate false-positives introduced by using keyword-based matching, our crawler inspects the HTML DOM and the text of the parent and grandparent elements of those containing any of these keywords to identify and verify the presence of age verification mechanisms or warning messages about the content of the webpage. If a relevant message is found, then the crawler clicks on the element to access the landing page. Finally, we fetch privacy policies by searching for URL links containing the keywords "Privacy" and "Policy" in any of the 8 languages. We manually validate the accuracy of our method in Section 7.2.

Geographical diversity: One of the goals of this work is to study whether pornographic websites behave differently depending on the user location and jurisdiction. To answer that, we run our crawls from a vantage point located in Spain, and use two commercial VPN providers – NordVPN [66] and PrivateVPN [73][5] – to gain access to vantage points in other EU state members, Singapore, India, Russia, USA, and the UK [6]. When crawling from Russia and India, we could not access 21 and 168 pornographic websites, respectively. Unfortunately, we can not assert whether this is due to country-level censorship or server-side blocking [59].

4 THE PORN WEB ECOSYSTEM

As of today, the research community lacks of generalizable and robust methods to classify domains by the type of service that they offer, and to identify their parent company [76]. However, gaining this knowledge is critical not only to identify websites offering sensitive content and to be able to identify the organization providing tracking services, but also to assess the accountability of these organizations. In this section we explore (1) the main organizations providing pornographic content and their business models (Section 4.1); and (2) the analysis of third-party tracking technologies embedded in pornographic sites (Section 4.2).

4.1 Discovering Website Owners

Discovering the parent company or organization supporting a given website is a hard problem that requires applying complementary methods. We start this analysis by crawling and measuring differences across pairs of pornographic websites at the landing page and privacy policy (when available) of each pornographic website to search for organization-level information. For the majority of pornographic websites, this information is either vague or incomplete: *e.g.,* some websites only report a postal address rather than a company name accompanied by legal information. Second, we apply the term frequency-inverse document frequency statistical method (TF-IDF) [79] to measure the similarity between privacy policies and the HTML `<head>` element of each pair of pornographic websites to automatically find clusters that might belong to the same organization. We manually analyze each pair and cluster to remove potential false-positives. This method allows us to identify over 80 porn websites that belong to six different companies, including AFS Media LTD., Techpump, Gamma Entertainment, and PaperStreet Media. To increase the coverage and improve the accuracy of our attribution process, we leverage DNS, WHOIS, and X.509 certificate information and insights obtained from white papers, scientific articles, and public reports about the pornographic industry [37, 86].

The combination of these methods only allows us to accurately find 24 companies owning 286 pornographic websites. We could not find reliable organization-level information for 96% of the pornographic websites in our dataset. This lack of corporate or organizational transparency is particularly concerning for websites – data controllers in the context of GDPR – engaging in user tracking or embedding third-party services as their visitors will not be able to effectively exercise their privacy rights to any corporation (*e.g.,* demanding access, corrections, or deletion of their data as indicated in the GDPR). We further discuss in Sections 5 and 7 the presence of trackers in pornographic websites, and their long way towards regulatory compliance, respectively.

Main pornographic website operators: Table 1 shows the 10 largest clusters of organizations ordered by the number of individual pornographic websites that they own. These companies own and operate 3% of the total websites in our corpus. The reasons behind these clusters or pornographic websites are manifold. Typically, these clusters are created through acquisitions and mergers between companies, similar to the industry trends present in the online advertising and tracking industry [65, 76]. Furthermore, pornographic websites are typically federated. This gives them the ability to reach out larger audiences and increase advertising revenues through affiliated services, while also re-publishing and sharing pornographic material across sites.

Monetization Models: The majority of pornographic websites combine different monetization mechanisms, such as online advertising-based models (see Section 4.2), subscription (premium) services, and, in some cases, even through cryptomining services (see Section 5.3).

[4] We select English, Spanish, French, Portuguese, Russian, Italian, German, and Romanian for being the most common default languages in our list of pornographic websites. We choose these keywords after a manual inspection of part of the websites in our corpus.

[5] We select those VPN providers because 1) they do not appear to manipulate traffic according to our experiments, and 2) they forward traffic through VPN servers rather than through real users in a P2P fashion [50].

[6] We perform these measurements in the UK to study websites' compliance with the Digital Economy Act [53]

Table 1: Largest clusters of pornographic sites, grouped by their parent company. For each company, we report the number of individual websites owned and the one with the highest Alexa rank throughout 2018. A larger cluster size does not necessarily translate into popularity.

Company	# sites	Most popular site (rank)
Gamma Entertainment	65	`evilangel.com` (5,301)
MindGeek	54	`pornhub.com` (22)
PaperStreet Media	38	`teamskeet.com` (10,171)
Techpump	25	`porn300.com` (2,366)
PMG Entertainment	15	`private.com` (7,758)
SexMex	12	`sexmex.xxx` (122,227)
Docler Holding	10	`livejasmin.com` (36)
Mature.nl	9	`mature.nl` (6,577)
Liberty Media	7	`corbinfisher.com` (26,436)
WGCZ	5	`xvideos.com` (32)
AFS Media LTD	5	`theclassicporn.com` (13,939)
AEBN	5	`pornotube.com` (31,148)
Zero Tolerance	5	`ztod.com` (40,676)
Eurocreme	5	`eurocreme.com` (110,012)
JM Productions	5	`jerkoffzone.com` (147,753)

We perform a semi-automatic classification of these websites to infer their business models. First, we parse the landing page of the websites in our dataset and look for keywords that may indicate the option to create an account (*e.g.,* "Log In", "Sign Up") or "Premium" services. We use this signal as a proxy to identify which websites may offer subscription-based services after authentication. Then, we manually label the subscription model as "free" (*i.e.,* the content is freely available after registration), or "paid" (*i.e.,* the content is protected by a payment wall) by inspecting the website. We also verify that the keywords for creating an account and for detecting premium services remain stable independently of the language of the webpage. Thanks to this method, we can conclude that 14% of the porn websites in our corpus offer subscription options; and only 23% of the websites require a payment. While the study of the privacy risks of subscription-based services is outside the scope of this paper, it may be possible that once a user creates an account, all of their actions might be also linked to their profile and banking information.

4.2 Third-Party Services in Porn Websites

A large number of pornographic websites rely on online advertisements to monetize their user base and content and on analytics services for tracking their audiences. However, many ad networks set strict limitations on the usage of their services in pornographic websites, possibly as a measure to protect their brand reputation [39]. This state of affairs has given birth to lesser known ecosystem of advertisement and tracking services (ATS) specialized in adult content which have escaped research and regulatory scrutiny. We conjecture that our current limited understanding of trackers in sensitive websites has been caused by the low penetration of some of these trackers across the whole web landscape, hence falling in the long-tail. In fact, many pornographic websites are rarely indexed in domain ranks so they might not be present in studies that crawl a one-day sample of popular domain ranks [81].

Table 2: Number of first party and third-party domains found on our dataset of pornographic and regular websites. ATS makes reference to third-party Advertisement and Tracking Services.

Domain category	Pornographic websites (P)	Regular websites (R)	$\lvert P \cap R \rvert$
Corpus size	6,346	8,511	—
First-party	727	3,852	—
Third-party	5,457	21,128	889
Third-party ATS	663	196	86

In this subsection we study the third-party services and organizations operating in the online porn industry, and compare them with those present in regular websites. With our OpenWPM-based crawler, we find 5,457 different third-party domains embedded in the set of 6,346 pornographic websites that we could successfully crawl (out of our 6,843 sanitized dataset of pornographic websites). An eyeball analysis of these domains reveals that the majority of them belong to third-party analytics and advertising services, but also to CDN providers and social networks. To obtain a more accurate picture of the third-party tracking ecosystem in pornographic websites, we use the following complementary heuristics to (1) label and classify the domains embedded in pornographic websites as first-, third-party, or third-party advertising and tracking (ATS) services; and (2) attribute hostnames to organizations:

(1) **Third-party service extraction:** We collect all the URLs from all the HTTP(S) requests triggered by our OpenWPM-based crawler to identify the presence of third parties. For comparison, we run our crawl both for our pornographic and regular website datasets. For each URL and HTTP(S) request, we compare its fully qualified domain name (FQDN) and its X.509 certificate information (when available) along with the FQDN and certificate information of the host website, to determine whether a service is a first or third party. If we cannot establish a relationship between a host website and an embedded service based on the previous method, we compute the similarity between the two FQDNs using the Levenshtein distance [55]: if the similarity is higher than 0.7, we then consider the FQDNs to belong to the same entity. We manually verified the results and found this method to be accurate. This method also allows us to group together domains such as `doublepimp.com` and `doublepimpssl.com`, but also to make the distinction between *e.g.,* `doublepimp.com` and `doubleclick.net`. We can successfully label as third party domains 91% of the 6,017 FQDNs contacted when crawling all the porn websites by using this technique.

(2) **ATS classification:** We rely on EasyList and EasyPrivacy blacklists [25] – downloaded on Jan. 29th, 2019 – to identify domains belonging to well-known ATSes . These blacklists are designed and used by the AdBlock [6] and AdBlockPlus [7] browser extensions. Since they are based on rules that consider the whole URL request (*e.g.,* `bbc.co.uk` is not blacklisted, but

bbc.co.uk/analytics is), we match the full URL provided by OpenWPM with these blacklists to identify actual instances of tracking. We relax the matching method to the base FQDN domain to identify the presence of 12% third-party ATS organizations [100].

(3) **Finding the parent company for third-party services:** To better understand the trackers and organizations involved in the ecosystem, it is also critical to associate third-party domains to their parent company. We initially considered using Disconnect's domain-to-company mapping [23] but we soon realized that it is incomplete. We designed a method to complement Disconnect's list with organization-level information found in the X.509 certificate of each third-party domain[7], hence improving significantly its accuracy and coverage. For instance, we could assign to Oracle several third-party trackers like addthis.com (AddThis) [8] and bluekai.com (BlueKai) [17] services [8]. After this process, we found the parent company for 4,477 (74%) FQDNs, accounting for 1,014 companies, while using Disconnect's list yields only 142 of them.

4.2.1 Third parties in regular versus porn websites. Table 2 compares the number of third-party domains present in our set of pornographic websites with those present in our reference set of regular websites. This comparative analysis uncovers significant differences. In aggregated terms, we found 21,128 third-party domains (FQDNs) in our set of regular websites but only 5,457 in the pornographic ones. However, when looking specifically at ATS services, we see that they are more widespread and diverse in pornographic websites as 12% and 1% of all the third-party domains found in pornographic and regular websites are associated with ATSes, respectively. The intersection between the set of ATSes operating in the regular and pornographic websites is also low: only 86 third-party advertising and tracking services are present in both types of websites. This analysis reveals that a majority of advertising and tracking services operating in the online pornography ecosystem are unlikely to be present in regular websites. For instance, exosrv.com and exoclick.com, both belonging to Exoclick, are found in 2,709 pornographic websites (43% of the corpus) but only in 6 regular websites. These figures only represent a lower-bound estimation of the presence of advertising and tracking services in pornographic websites due to the well-known limitations of existing domain classifiers and blacklists [48, 76]. In Section 5, we will inspect the behavior of each third-party service to identify more trackers.

4.2.2 A closer look at the long-tail. The set of third-party services present in pornographic websites varies with the popularity of the hosting site. More concretely, the more unpopular the pornographic website is, the more obfuscated and opaque are the third-party domains it embeds.

Table 3 shows the presence of third-party services in porn websites when grouped in different popularity intervals (according to their highest Alexa rank throughout 2018). Only 3% of third-party domains, regardless of their purpose, are present in the four different tiers of popularity. Amongst those we find cloud providers such as cloudflare.com and large advertising companies (*e.g.,* DoubleClick by Alphabet), but also ATS companies specialized in adult websites such as doublepimp.com or exoclick.com. We would like to stress that Alphabet Inc. has specific policies about the type of content that can be distributed through their ad network as well as on the hosting site [39].[9]

Table 3: Third-party presence by popularity interval (per Alexa's 2018 highest rank). For each interval we show the total number of third-party domains ("Total") and the third-party domains found only in this interval ("Unique")

Popularity Interval	Number of porn websites	Third-party domains (Unique to the interval)
0 — 1k	73	407 (119)
1k — 10k	536	1,327 (531)
10k — 100k	3,668	3,702 (2,115)
100k+	2,069	2,363 (1,007)

In order to get a better understanding of the implications of low popularity – and possibly reputation – in terms of third-party services embedded in porn website, we take a deeper look at 2,069 unpopular pornographic sites that never got indexed by the Alexa Top 100K rank throughout 2018. This detailed analysis confirms that it is more likely to find advertisement and analytic services that are not commonly used by the prominent websites in unpopular pornographic websites. In fact, we find that 18% of the third-party services embedded on all porn websites appear only in the less popular ones according to Alexa. This is the case of analytic services like adultforce.com and zingyads.com [9, 102], for which we could not find a privacy policy on their homepages. We also found four Russian tracking services (betweendigital.ru, datamind.ru,adlabs.ru and adx.com.ru) on pornovhd.info, a Russian porn website. . Finally, we remark the presence of a potentially malicious domains (according to Dr. Web) such as the traffic trade webpage itraffictrade.com [24].

4.2.3 An organization-level analysis. We now present an organizational level analysis of the third-party domains operating in pornographic websites, regardless of their role. Figure 3 shows the 19 companies offering third-party services to most of the studied pornographic websites. As we can see, Alphabet is – as in the regular web – the most prevalent organization (74% of the total pornographic websites). Exoclick and Cloudflare services[10] are second and third with 40% and 35% of prevalence, respectively. When comparing with the third-party companies present in the regular web, we find that several ones solely operate in the adult industry. While some of them are well-known actors like Exoclick [33], others are lesser known companies like JuicyAds (4%) [49] and EroAdvertising (4%) [29].

[7] In some cases, the Subject field only contains the domain name of the website instead of the company name. We choose not to take the certificate information of these websites into account.

[8] Oracle operates a data marketplace, the largest third-party data marketplace for "open and transparent audience data trading" according to their own sources [69].

[9] Performing an analysis on whether the host sites are in compliance with Google Ads Policies is outside the scope of this paper.

[10] In this specific case, we cannot confidently confirm that Cloudflare is operating these domains. It might be possible that other companies, advertising services or tracking services might be using Cloudflare's infrastructure.

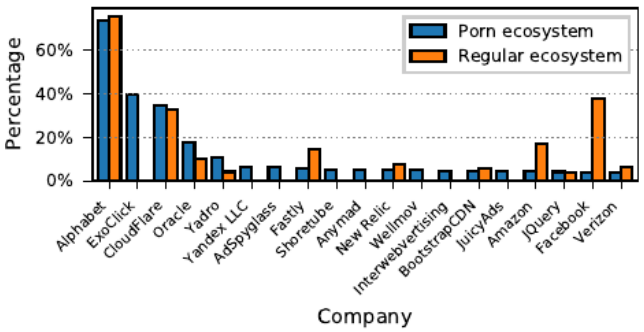

Figure 3: Most relevant third-party organizations in the porn ecosystem. We show their prevalence in the regular ecosystem for comparison.

In general terms, the presence of Alphabet services (*e.g.,* Doubleclick, Google Analytics) is very similar in both regular and pornographic websites. Yet, the prevalence of each individual service varies greatly: `google-analytics.com` is present in 39% of porn websites, while `doubleclick.net` – an ad-network – appears in 12% of them (for reference, 60% of the analyzed non-porn websites connect to Doubleclick domains). The higher presence of Oracle in porn websites is caused by its `addthis.com` service, which provides web developers features like social network integration and content sharing (*e.g.,* pictures or videos). Another interesting case is the domain `alexa.com` which is related to the Amazon-owned browser extension that populates such list. Another interesting case is the presence of the domain `rlcdn.com` in four pornographic websites, one of them offering *bestiality porn*, a practice considered illegal in many countries of the world. This domain belongs to RalpLeaf which is a subsidiary of TowerData/Acxiom [1], one of the largest data brokers in the world [15, 97]. Finally, while Facebook is highly popular in the web ecosystem, its presence in pornographic websites is really low.

5 PRIVACY RISKS

The sensitive nature of pornographic websites, and the quite unique ecosystem of third-party ATSes operating in them highlight the importance of studying in depth the behavior of these websites and their use of tracking technologies. In this section, we perform a multi-dimensional analysis of the various privacy risks to which visitors of pornographic websites might be exposed to (Section 5.1). We also provide an analysis of the use of insecure protocols (*e.g.,* HTTP) who may allow in-path observers like censors to monitor users' browsing habits (Section 5.2), and report on the presence of known malware in these sites (Section 5.3).

5.1 User Tracking Techniques

We leverage our customized version of OpenWPM to measure the use of various tracking techniques in pornographic websites, specifically HTTP cookies, cookie syncing, and advanced fingerprinting techniques.

5.1.1 HTTP Cookies. Online companies often use HTTP cookies as a means for tracking users across the web. They do so by generating and storing unique identifiers in end-users' browsers. Using OpenWPM, we can successfully identify 89,009 HTTP cookies installed by 92% of our dataset of porn websites. This includes both first- and third-party cookies. However, not all cookies might be used for the purpose of tracking users (*e.g.,* session cookies). Therefore, we focus our analysis in those HTTP cookies that may potentially contain user identifiers. For that, we discard session cookies and those with a length below 6 characters which are unlikely to contain unique identifiers [36]. After applying this filter, 51,648 HTTP cookies that can potentially be used for tracking users remain. 3% of them are larger than 1,000 characters, even reaching 3,600 characters in the case of cookies installed by third-party ATS services like `juicyads.com`, `tsyndicate.com`, `exoclick.com`, `exosrv.com`, and other porn websites.

Table 4: The 5 most common third-party domains delivering cookies that potentially contain unique IDs.

Third-party domain	% porn websites	# Cookies	ATS	In web ecosystem	% Cookies with user IP
`exosrv.com`	21%	2095	✓	✓	85%
`addthis.com`	17%	1289	✓	✓	0%
`exoclick.com`	14%	434	✓	✓	29%
`yandex.ru`	4%	312	✓	✓	0%
`juicyads.com`	4%	475	✓	✓	0%

We now focus our study on the 30,247 HTTP cookies installed by 3,343 third-party domains in 72% of our corpus of pornographic sites. The 100 most popular cookies (by their unique *name* = *value* combination) appear in over 30% of the total porn websites. Moreover, as shown in Table 4, the main third-party services responsible for installing HTTP cookies in users' browsers are ExoClick, Oracle (*via* AddThis), Yandex, and JuicyAds. While ExoClick and JuicyAds are specific to the online porn ecosystem, AddThis and Yandex are commonly found in regular web services, allowing these firms to potentially track users across the whole web.

Encoded Information in HTTP Cookies. We decode the cookie values using two types of encoding: base64, and URL. We detect 2,183 cookies that store the IP address of our physical machine along with potential IDs. 97% of these cookies belong to different Exoclick domains, which are present in 440 different porn websites as shown in Table 4. In particular, 85% of `exorsrv.com` cookies and 29% of `exoclick.com` cookies follow this pattern. Furthermore, we identify 28 cookies in 15 websites that store approximate geolocation data, potentially obtained through geo-IP databases [58]. 27 of these 28 cookies are delivered by two third-party domains, `fling.com` and `playwithme.com`. While the former only stores the coordinates, the latter also includes detailed information about the network provider. While the accuracy of geo-IP databases is not very precise in general, it could reveal the precise location of a user in certain scenarios [71].

5.1.2 Cookie Synchronization. For security purposes, modern web browsers limit the access to cookies to the service that has installed them [62]. To circumvent this security mechanism and ease cross-site tracking, third-party services use a technique called cookie synchronization (cookie syncing, in short) that allows them to share their cookie data with other services by embedding the cookie in the

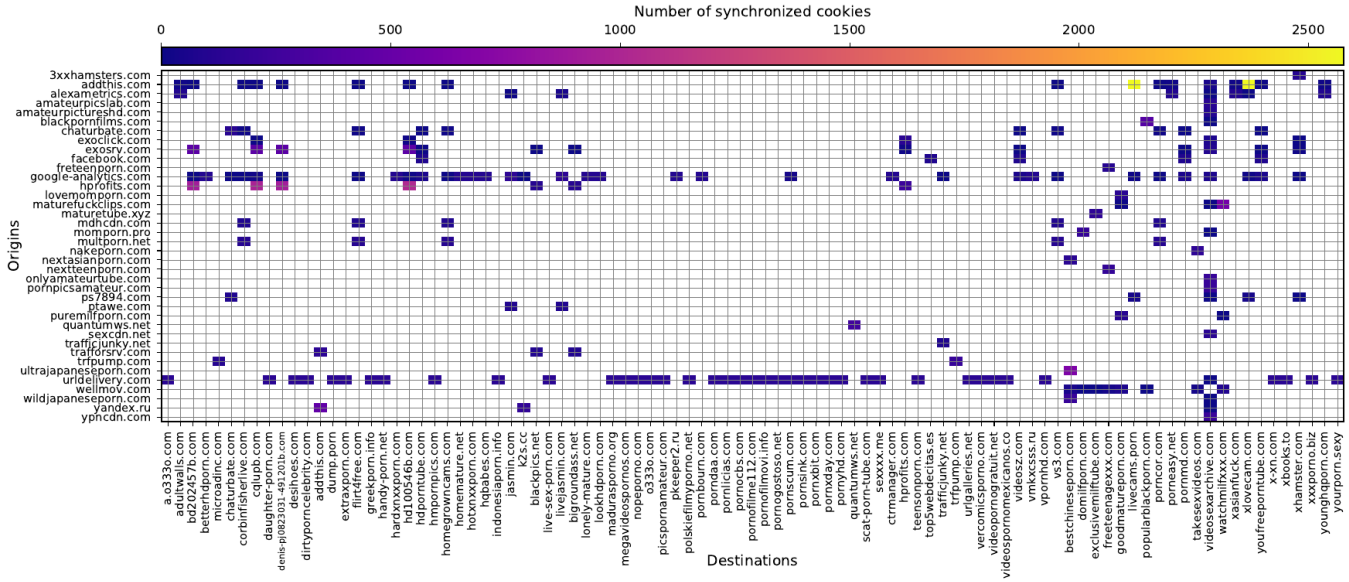

Figure 4: Cookie syncing between organizations. Pairs of domains that exchanged at least 75 cookies are shown.

URL [40, 70]. We study the use of cookie syncing in pornographic websites by checking if any of the observed HTTP cookies are later embedded in subsequent HTTP requests. To avoid introducing false positives, we do not split the cookie value by delimiters like "-" or "=". Hence, our findings offer a lower bound estimation of the prevalence of this technique.

The number of pornographic websites for which we have observed this practice is 2,867. This covers 58% of the top-100 most popular porn websites according to Alexa. However, the matching of the pairs of organizations (at the domain level) involved in this practice yields 4,675 different pairs as shown in Figure 4 (for clarity reasons, we only show the pairs of domains that exchange at least 75 cookies). Specifically, we can find 1,120 origin and 727 destination services. Cookie syncing can also occur between domains belonging to the same organization. For instance, the third-party domains `hd100546b.com` and `bd202457b.com` synchronize HTTP cookies with `hprofits.com`. The X.509 certificates for these three domains suggest that all of them belong to `hprofits.com`, an ad exchange platform according to their website.

5.1.3 User Fingerprinting. Fingerprinting techniques allow trackers and services to create a unique user identifier by accessing and processing several characteristics of the user's device using JavaScript APIs. As opposed to cookie-based tracking, this sophisticated method can be used to persistently track users and their activities across websites without having to rely on cookies.

First, we analyze pornographic websites and third-party services using either canvas or canvas font fingerprinting techniques [27]. HTML Canvas Fingerprinting is a tracking technique that exploits system differences between devices in how they render images. These scripts use the `CanvasRenderingContext2D` and the `HTMLCanvasElement` JavaScript APIs to generate images using specific height, width, fonts, and background colors, among other characteristics. Font fingerprinting, instead, is a variation of canvas fingerprinting in which a tracker can leverage the fonts that each browser has installed to generate a unique ID of the device. This is achieved with the `measureText` method of the HTML Canvas API which allows to draw text using different fonts. Depending on the size of the written text, the tracking service can infer if a particular font is installed.

Table 5: Third-party domains using different tracking-techniques. The `ATS` and `Regular web` columns indicate whether these services are indexed in EasyList/EasyPrivacy or if they are present in the regular web, respectively.

Domain	Presence in porn sites	ATS	Regular web	Canvas fingerprinting	WebRTC
`adsco.re`	152	-	✓	0	1
`ero-advertising.com`	33	✓	✓	32	0
`cloudfront.net`	31	✓	✓	8	0
`cloudflare.com`	28	✓	✓	2	0
`adnium.com`	26	✓	-	41	0
`highwebmedia.com`	22	✓	✓	1	0
`xcvgdf.party`	18	-	-	18	0
`provers.pro`	15	✓	-	1	0
`montwam.top`	13	✓	-	25	0
`dditscdn.com`	10	✓	✓	1	0

Yet, not all the services that invoke these JavaScript APIs do so for the purpose of tracking users. To eliminate false positives, we follow the methodology proposed by Englehardt *et al.* [27]. In the case of canvas fingerprinting, we exclude: (1) all the canvas with width and height below 16px; (2) scripts that do not use at least two colors or text with more than 10 different characters; (3) scripts that do not call either the `toDataURL` or the `getImageData` methods with an area below 320px; and (4) scripts that use the save, restore, or `addEventListener` methods of the rendering context. Despite these precautions, none of the scripts reported by OpenWPM meet these criteria. As a result, we set stricter conditions

to identify scripts performing font fingerprinting: we only count those that set the font property and call the `measureText` method on the same text at least 50 times. This allows us to find 245 different JavaScripts performing canvas fingerprinting in 315 porn websites. 74% of the JavaScripts are fetched from 49 third-party services including `ero-advertising.com` and `highwebmedia.com`, a service that belongs to `chaturbate.com` (one of the biggest live sex services). These third-party services are present in 4% of all the porn websites in our dataset. We only find one script, delivered by `online-metrix.net`, using font fingerprinting.

We find that the script performing font fingerprinting and 91% of the scripts using canvas fingerprinting have not been previously indexed by tracking blacklists like EasyList and EasyPrivacy. As a result, these services could track users even when they use plugins such as ABP [7]. One example is the script delivered by `xcvgdf.party` (see Table 5) which performs canvas fingerprinting on 18 different porn websites, including a website offering transsexual/transgender pornography, `ladyboy-porno.com`.

5.1.4 Other Potential Tracking Methods. Our methodology allows us to find instances of other methods that could be potentially used for tracking purposes. However, we did not gather sufficient evidence to demonstrate that these JavaScript APIs are actually used for such purposes. One case is WebRTC [43], a technology to establish real-time peer-to-peer communications between browsers. WebRTC APIs allow collecting the IP address of the users, as well as the local network address. Through the combination of WebRTC with other tracking techniques [27], online services can discover networking information such as devices hosted behind the same NAT for cross-device tracking [78], or identify whether the user connects through a VPN [50]. In our dataset, there are 27 different JavaScripts using WebRTC present in 177 different pornographic sites, 21 of which use other tracking mechanisms in conjunction. Two of the 13 different third-party services using WebRTC, appear in the regular web and are classified as ATSes by EasyList. These services are `traffichunt.com` and `online-metrix.net`, an advertisement platform and a web analytic service, respectively.

5.2 (Lack of) Network Security Standards

Safeguarding users' privacy and security should be a priority for providers of pornographic content, particularly if users can be subject to censorship and surveillance at the network level [75, 99]. The use of encryption for transmitting data over the network is also a provision in privacy laws such as the GDPR (Article 32 [20]) and CCPA [18]. To identify the lack of security protocols in pornographic websites, we measure HTTPS support in porn websites by inspecting the requests triggered by our OpenWPM crawler. By default, we crawl each website using HTTPS, only downgrading to HTTP when HTTPS is not supported by the server.

Table 6 shows the use of HTTPS in pornographic websites depending of their highest Alexa rank in 2018. We find that over 92% of the most popular websites (in the top-1K of the Alexa ranking) do support HTTPS. However, the ratio of porn websites supporting HTTPS drops as their popularity does: HTTPS support decays to less than 25% for websites whose highest Alexa rank in 2018 was 10,000 or lower. This trend is similar for third-party services: those included in popular porn websites are more likely to support HTTPS. Nevertheless, we can find that 4,663 pornographic websites (68% of the total) are not fully HTTPS: either the website or one of its embedded third-party do not support HTTPS. By inspecting the content of these flows, we can identify that 8% of these websites upload cookies containing sensitive data in the clear as shown in Section 5.1.1.

Table 6: HTTPS usage in pornographic websites

Interval	Feature	HTTPS
0 — 1k	Porn websites (75)	92%
	3rd-party services (407)	90%
1k — 10k	Porn websites (552)	63%
	3rd-party services (1,327)	48%
10k — 100k	Porn websites (3,886)	32%
	3rd-party services (3,702)	25%
100k+	Porn websites (2,330)	22%
	3rd-party services (2,363)	16%

5.3 Potential Malicious Behaviors

We conclude this section with a short study of the presence of potentially malicious behaviors in pornographic websites according to VirusTotal [95]. To minimize false positives, we only report domains flagged as malicious by at least 4 of the 70 different malware scanners aggregated by VirusTotal. There are 7 porn websites classified as a potentially malicious by VirusTotal. Further, malicious and deceptive behaviors also extend to 16 third-party services embedded in 41 porn websites.

We highlight the presence of three cryptocurrency mining services: `coinhive.com`, `jsecoin.com` and `bitcoin-pay.eu` in 8 porn websites. The latter domain, `bitcoin-pay.eu`, is not active anymore, but is related to `crypto-webminer.com` [21]. This suggests that owners of pornographic websites had explored alternative monetization schemes beyond online advertisement and subscription-based models. Whether these practices are performed with user consent is beyond the scope of this study.

6 MEASURING GEOGRAPHICAL DIFFERENCES

This section measures whether pornographic websites adapt their behavior – including the presence of trackers – to the geographical location of the user, possibly to meet the requirements of different regulatory frameworks. For that, we launch our crawls from different vantage points using commercial VPNs and our physical vantage point located in Spain.

6.1 Third Party Services

Table 7 shows the number of third-party services embedded in porn websites per country. We can see that the total number of third-parties in each location remains rather stable but for Russia, which has over 700 third-party services less. When looking at individual instances of third-party services, we find that there are hundreds of domains that are unique in each country but around 10% of them are related to CDNs or porn websites that generate

Table 7: Comparison of the domains found on porn ecosystem from different geographical points. The values do not include domains loaded dynamically on the websites.

	FQDN	Web Ecosystem	Unique Country	ATS	Unique Country
USA	5,483	16%	357	635	25
UK	5,364	15%	231	620	20
Spain	5,494	16%	561	592	59
Russia	4,750	16%	373	542	27
India	5,340	15%	275	607	21
Singapore	5,310	15%	233	608	16
Total	7,813	14%	2,030	816	168

arbitrary domains such as `img100-589.xvideos.com`. If we look at ATS domains specifically, we can see that Google services dominate at a global scale, regardless of users' geolocation.

6.2 Malware Presence

The number of third-party domains considered as malicious by VirusTotal varies per country: from 15 third-party domains when accessed from Russia to 19 when accessed from India. Yet, 13 of these malicious domains are present regardless of users' geolocation (*e.g.,* the cryptomining domain `coinhive.com`). When counting the number of pornographic websites that contain such malicious content, the figure varies from 29 websites in Russia to 42 in Spain. Nevertheless, 26 pornographic websites always contain malware regardless of the country of access. This indicates, that some of the organizations serving malicious content might target users located at specific world regions.

7 REGULATORY COMPLIANCE

We now evaluate pornographic websites' efforts to comply with regulation. Specifically, we verify: (1) the presence and use of cookie consent forms as required by the EU GDPR and ePrivacy regulation; and (2) the use of verifiable age verification mechanisms in the context of the UK's Digital Economy Act. We also investigate the lack of privacy policies and potential inconsistencies that exist between these legal documents and the behavior observed in each pornographic websites in terms of user tracking and the presence of third-party ATSes.

7.1 Cookie Consent Notice

The ePrivacy directive will require websites to obtain consent from European users before installing and using cookies, unless the cookie is strictly necessary for the webpage functionality. As this legislation is not yet into effect (it will take effect in 2019), the use of cookies is currently regulated by the GDPR, which indicates that users must consent to the use of any technique that may uniquely identify them [20]. This is typically done through cookie consent forms.

Degeling *et al.* performed a preliminary analysis of cookie consent forms (cookie banners) in 6,579 websites after GDPR came into effect [22]. They found that around 62% of the websites display a cookie consent-banner and developed a categorization of HTTP cookie banners that considers 6 different groups: (1) *No Option*: This type of cookie banner only informs users about the use of HTTP cookies without giving the possibility of accepting or rejecting them; (2) *Confirmation*: This type informs users about the use of cookies, but users can only show their accordance with the use of cookies, they can not reject them; (3) *Binary*: In this case, users can accept or reject the use of cookies; (4) *Slider*: This type of cookie banner gives users more fine-grained control over the level and type of cookies, that they allow by adjusting a slider; (5) *Checkbox*: This type of banner gives users the capacity to allow/reject cookies for a specific purpose or from a particular third-party service; and (6) *Other*: Any other type of banner that does not match any of the above. These banners tend to have a higher degree of complexity.

Table 8: Usage of HTTP cookie banners in porn websites.

Type	EU	USA
No Option	1.36%	1.39%
Confirmation	2.82%	2.3%
Binary	0.2%	0.06%
Others	0.03%	0.01%
Total (N = 6,843)	4.41%	3.76%

Identifying cookie banners automatically in websites following Degeling's method and taxonomy is not trivial. In fact, we could only instrument our customized OpenWPM to identify the following types: *No option*, *Confirmation* and *Binary*. We merge the *Slider* and *Checkbox* types together in the *Others* category, as we would need to interact with the banner to be able to further categorize them. Our method works as follows: first, we inspect the HTML DOM to find elements that resemble a banner (inspecting the text of the banner). If such an element is found, we extract the text rendered to the user, and take a screenshot that we manually analyze to manually verify that the HTML element is indeed a banner. We repeat this procedure from two countries, Spain and the USA to find potential differences in cookie banner presence.

Table 8 shows the percentage of pornographic websites in which we find HTTP cookie banners. As can be observed, the proportion of pornographic webpages with cookie banners is very small, being only 4% of the total. A second observation from Table 8 is that the difference between accessing webpages from one country or the other is also very small, as only 0.65% more pages show a cookie banner when fetching them from Europe.

The low presence of cookie banners is remarkable when compared with the fact that 72% of the pornographic websites studied contain third-party cookies (Section 5.1.1).[11] Moreover, out of the websites that show a cookie banner, 32% do not give users any control over the use of cookies as the banner only discloses their use (No Option type). While it is possible that not all third-party cookies are actively used for tracking purposes, these figures suggest that many websites offering sensitive content may potentially be in violation of the GDPR.

It is important to note that our methodology uses OpenWPM to crawl the websites and that we do not interact with the webpage

[11] For comparison purposes, Degeling *et al.* showed that 69.9% out of a corpus of 6,357 websites had a cookie consent banner in January 2018 [22].

once we have visited it. Therefore, even in the websites where a cookie banner is present, we never gave actual consent to the use of cookies. As a final note, one might expect that large corporations providing pornographic content would have strong incentives to be in compliance with regulatory requirements. While this is the case for a small fraction of popular pornographic websites, there is not a clear correlation between the use of cookie consent forms and the popularity of porn webpages.

7.2 Age Verification

Some pornographic websites have taken positive steps to implement age verification mechanisms in an effort to comply with increasing regulatory pressures (see Section 2.1). In this section, we study how prevalent and how effective verifiable age-verification mechanisms are in the wild. For that, we use our Selenium-based crawler to parse the landing page of each porn website, and look for warnings and consent forms displayed to the user as detailed in Section 3.1. As our approach relies on string matching to identify such warnings, it is prone to introduce false positives, specially so in age-related keywords that appear often in the content of the websites. Therefore, due to the difficulty to perform this study automatically and at scale, we only investigate a subset of the top-50 most popular pornographic websites manually.

We perform this manual analysis in 4 countries (the US, the UK, Spain, and Russia) to identify regional differences. The results from the USA, UK and Spain are consistent: *i.e.,* the same set of 20% of the pornographic websites implement and show to the end user the same age verification mechanism, consisting of a simple warning text and a button to be clicked on. However, there are significant differences when accessing the same websites from Russia: only 14% of the analyzed websites have an age verification mechanism. Additionally, 8% of the websites that do not verify users' age for the rest of countries do so in Russia, whereas 12% of the websites do not verify user age in Russia but do so in the rest of countries studied. We note that we did not find any instance of AgeID being deployed during our study.

Despite regulatory pressures, the current age verification mechanisms implemented by all these sites are easy to bypass and could not be considered as "verifiable age verification mechanisms". In other words, if our automatic crawler manages to bypass the mechanism, a child could do it as well. We only found one webpage in Russia, `pornhub.com`, implementing a complex age verification mechanisms through social media accounts as requested by the Russian federal government in 2017 [94].

7.3 Privacy Policies vs. Reality

The GDPR [20] requires all websites collecting or processing personal identifiable data from European citizens to portray a privacy policy describing their personal data collection and processing practices, including data collected by embedded third parties. We perform a best-effort crawl to collect the privacy policies, if available, of each pornographic website to crosscheck with our empirical results, and highlight potential privacy violations. We perform this analysis using the method introduced in Section 3.1, only from our physical machine located in Spain.

Our crawler inspects the DOM of the landing page looking for a link to the privacy policy. We are able to find a privacy policy in 16% of the pornographic websites in our dataset. We get these figures after a manual sanitization of our results in which we manually check the privacy policies which are abnormally short in the number of words and found 44 false positives caused by HTTP errors (response codes).

The GDPR forced changes in the way privacy policies are presented to users, forcing publishers to be clear about their data collection, processing and sharing practices, as well as user rights. We use string matching to find that 218 (20%) of the privacy policies make an explicit mention to the GDPR. We dive deeper into the analysis of the privacy policies by first looking at length patterns, in an attempt to understand how similar (or different) policies are. We find that, on average, privacy policies contain 17,159 letters and that the shortest policy we found has 1,088 letters, and the largest 243,649.

While this might hint that there are big differences across policies, we further investigate the similarity of the text in privacy policies. We use the term frequency-inverse document frequency (TF-IDF) [79] to measure the similarity between two texts.[12] We run this measure for the 1,202,312 pairs of different privacy policies in our dataset and found that 76% have a similarity above 0.5 (meaning they are co-related). This can be a direct result of websites belonging to the same company having a very similar privacy policy as well as the prevalence of templates that are highly popular across websites. In fact, finding pairs of websites with a coefficient of 1 helped us discover companies holding a larger number of pornographic websites (Section 4.1).

The opacity of the privacy policies makes it difficult to perform an automatic analysis of their content at scale. To tackle this issue, we use the publicly available tool Polisis[41], which presents a human-readable summarized version of the privacy policy, to extract third-party entities and data collection methods. As Polisis does not provide APIs to access the results in a machine-readable format, we rely on the web version of the tool to further investigate the top 25 websites tracking users (*i.e.,* canvas fingerprinting and cookies) according to our results from Section 5. We manually asses that 72% of this subset of porn websites have a privacy policy in which they clearly state the use of cookies, the type of data collected, and the presence of third parties in their websites. Only one of the websites discloses in its privacy policy the complete list of third-party advertising and tracking services.

These results show that – while privacy policies are becoming more common, complete, and clear to users – there are still many websites engaging in user tracking without privacy policies and other transparency mechanisms. When they do, with only one exception, they do not disclose the whole list of embedded third-parties.

8 ETHICAL CONSIDERATIONS

The data collection process does not involve human subjects. All the experiments were ran automatically on a controlled environment using crawlers. The processes involving manual inspection were

[12] The value goes from -1 (exactly opposite) to 1 (exactly equal) going through 0 (no co-relation).

conducted by the authors of the paper. The members of the research team gave their approval to conduct such work, being aware of the possibility of having to see potentially uncomfortable images in some cases. Also, before running any experiments using the VPNs, we contacted NordVPN and PrivateVPN to inform them about our research work, in order to do not break their terms of uses and make sure that no harm to users were going to be caused.

Furthermore, we do not interact with the consent notices displayed by the websites and we do not surf beyond the landing page to avoid generating advertising revenues and accessing specific content. We do not discard the presence of additional tracking mechanisms and services beyond the landing page.

Before performing our data collection, we also defined a protocol to report any service distributing illegal pornographic content to the authorities in case that this uncomfortable situation arose. Unfortunately, we found one service distributing such content while performing our sanitization process. We immediately reported the case to the national authorities.

9 RELATED WORK

The research community has studied web tracking extensively. We expand the state of the art by looking for the first time at a specific but highly sensitive ecosystem that, to this day, has remained unexplored. We discuss below studies that are more relevant to our work.

Web tracking: Several research studies have made groundbreaking contributions to illuminate and uncover the privacy risks of the web. This includes studying the use of HTTP cookies, cookie syncing [5, 27, 28, 51, 70, 80], persistent tracking mechanisms [5, 54, 91], and advanced fingerprinting techniques [27, 35, 61]. In [14], Bashir *et al.* introduced the notion of inclusion chain to model the diffusion of user's data within third parties. These studies were possible thanks to web crawlers: either customized versions of Chrome or Firefox [54, 63], or headless browsers such as Selenium [74, 82, 96] or purpose-specific ones like OpenWPM [26, 27, 36, 60]. In this work, we leverage many of the techniques and definitions introduced in previous studies.

Pornographic websites: The online porn industry has remained largely underground. There has been isolated steps towards studying this ecosystem, mainly from a content availability standpoint in a major porn website [89]. Vasey and Abild tackle the topic of pornography in the Internet in [92], comparing what people say about their sexuality with the results of billions of Internet searches. Wondracek *et al.* studied the economic structure of the online porn industry [98], showing that these websites usually present shady schemes to generate revenue (such as traffic trading). Altaweel *et al.* used OpenWPM to study user tracking in 11 of the most popular pornographic websites [11]. They showed that there is lesser presence of tracking in porn websites in comparison with popular non-pornographic sites. Marotta-Wurgler studied the presence of privacy policies in websites [57], including 17 popular adult websites showing that this type of website is more likely to include privacy policies. The differences with our findings can be explained by the fact that our dataset of pornographic websites is significantly larger, and that it includes less popular, but still interesting, websites.

Regulatory compliance: We built on previous work to study GDPR compliance and measure websites' readiness for the ePrivacy directive and UK's Digital Economy Act age-verification. Degeling *et al.* studied the impact of GDPR on web privacy and the prevalence of cookie consent mechanisms in a corpus of 6,579 websites in each of the 28 EU member states [22]. Trevisan *et al.* studied the current implementation status of the EU cookie directive in more than 35,000 websites [88]. Other studies have looked at the presence of privacy policies across websites [101] and mobile apps [67, 83] and ways to automatically study such policies [41]. Kulyk *et al.* performed a user study on user's reaction to cookie consent notices in 50 German websites [52]. Finally, Marotta-Wurgler [57] found that most of the porn websites have policies. However, they only looked at the 17 most relevant porn websites.

10 FUTURE WORK

Our study has opened a number of research questions that we plan to address in the near future.

Our study focuses in basic aspects of GDPR compliance and UK's efforts to control minors' access to pornographic content. This analysis could be extended to look deeper into GDPR compliance, for instance by further analyzing the values of cookies to investigate the prevalence of other tracking IDs and by developing methods to reduce human intervention and supervision. An interesting aspect to study in the future could be characterizing cross border data exchanges as reported by Iordanou *et al.* [47] and Razaghpanah *et al.* [76], or performing a deeper investigation of the connections between online trackers, advertising services, and data brokers.

Another aspect that could be analyzed in future work are the privacy implications on those websites offering subscriptions, analyzing which type of data they require to create the account as well as compare the presence and amount of tracking services between the subscription and free modes. Finally, in this study, we have intentionally not studied aspects such as censorship of pornographic websites and the performance of anti-tracking technologies to protect users' privacy, including safe-browsing modes and popular ad-blockers. We believe that analyzing the effectiveness of such tools in specific ecosystems and longitudinally deserves a dedicated study on its own.

11 CONCLUSIONS

Online porn has been traditionally considered as an obscure subsystem of the Internet. Yet, the porn industry is not different in many aspects from regular web services: it has rapidly integrated advanced tracking technologies to monitor (and in some cases to monetize) users.

In this paper we performed the first comprehensive and large scale analysis of the privacy risks and (lack of) regulatory compliance of porn sites. We identified noticeable differences with regards to regular websites, specially regarding the parallel ecosystem of third parties offering advertising and tracking services to online porn websites. The presence of porn-specific trackers might render many anti-tracking technologies based on blacklists insufficient. We found that 91% of the scripts implementing canvas fingerprinting were not indexed by EasyList and EasyPrivacy lists. Furthermore,

we demonstrated that a large number of porn websites fail to implement the most common security mechanisms such as the use of HTTPS, and basic transparency requirements such as privacy policies and cookie consent forms, even in those websites actively tracking users. Only the companies behind some of the most popular pornographic websites seem to make efforts to comply with current legislation, possibly fearing the high fines of new regulations like the GDPR. Besides data protection and users' privacy aspects, we demonstrated that the efforts made by the online porn industry to prevent children access to inappropriate content are not being widely deployed. While most countries do not have laws to prevent children from accessing pornographic material, we have observed that the deployment of these mechanisms is rare even in jurisdictions where such laws will be applicable soon.

Our work opens new doors for other studies focused in measuring and characterizing the privacy risks of semi-decoupled and highly sensitive web subsystems (*e.g.,* gambling and online health services), while also informing the public debate. Many of these services might fall between the cracks of public scrutiny and research efforts that aim at identifying web privacy problems from a macroscopic perspective.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers and our shepherd Prof. Hamed Haddadi (Imperial College London) for their constructive feedback on the preparation of the final version of this paper.

This work is partially supported by the Spanish grant TIN2017-88749-R (DiscoEdge), the Region of Madrid EdgeData-CM program (P2018/TCS-4499), the European Union's Horizon 2020 Innovation Action program (grant Agreement No. 786741, SMOOTH Project) and "la Caixa" Foundation agreement LCF/PR/MIT17/11820009.

REFERENCES

[1] Acxiom. https://www.acxiom.com.
[2] Only4 adults. http://only4adults.com.
[3] Top porn sites. http://toppornsites.com/.
[4] Alexa Websites Ranking, 2019. https://www.alexa.com/topsites/.
[5] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., and Diaz, C. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (2014), ACM, pp. 674–689.
[6] AdBlock. Block Ads. Browse safe. https://getadblock.com/.
[7] AdblockPlus. Surf the web with no annoying ads. https://adblockplus.org/.
[8] Addthis. About AddThis. https://www.addthis.com/about/oracle/.
[9] Adult Force. Homepage. https://www.adultforce.com/#/.
[10] AgeID. AgeID announced to the industry at European Summit. https://www.ageid.com/press/article/11.
[11] Altaweel, I., Hils, M., and Hoofnagle, C. J. Privacy on adult websites.
[12] ASACP. ASACP Members. https://www.asacp.org/index.php?content=members#top.
[13] ASACP. Association of Sites Advocating Child Protection. https://www.asacp.org/.
[14] Bashir, M. A., and Wilson, C. Diffusion of user tracking data in the online advertising ecosystem. *Proceedings on Privacy Enhancing Technologies 2018*, 4 (2018), 85–103.
[15] BBC. Facebook scandal: Who is selling your personal data? https://www.bbc.com/news/technology-44793247.
[16] BBC News. Russia extends porn site ban. https://www.bbc.com/news/technology-37373244.
[17] Bluekai. Oracle Buys Bluekai. https://www.oracle.com/es/corporate/acquisitions/bluekai/.
[18] California State Legislature. California Consumer Privacy Act. https://www.caprivacy.org/.
[19] CookieBot. The EU ePrivacy Regulation and Cookies - What do I need to do? https://www.cookiebot.com/en/eprivacy-regulation-and-cookies/.
[20] Council of European Union. General Data Protection Regulation 679/2016, 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.
[21] Crypto Webminer . Crypto Webminer - Web mining - Mining in your Browser. https://www.crypto-webminer.com.
[22] Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., and Holz, T. We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy. *arXiv preprint arXiv:1808.05096* (2018).
[23] Disconnect Tracking Protection List. https://github.com/disconnectme/disconnect-tracking-protection.
[24] Dr. Web. Homepage. https://www.drweb.com/.
[25] EasyList. EasyList. https://easylist.to.
[26] Englehardt, S., Eubank, C., Zimmerman, P., Reisman, D., and Narayanan, A. Openwpm: An automated platform for web privacy measurement. *Manuscript* (2015).
[27] Englehardt, S., and Narayanan, A. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (2016), ACM, pp. 1388–1401.
[28] Englehardt, S., Reisman, D., Eubank, C., Zimmerman, P., Mayer, J., Narayanan, A., and Felten, E. W. Cookies that give you away: The surveillance implications of web tracking. In *Proceedings of the 24th International Conference on World Wide Web* (2015), International World Wide Web Conferences Steering Committee, pp. 289–299.
[29] EroAdvertising. Homepage. https://www.eroadvertising.com/#!/.
[30] EUR-Lex. ePrivacy proposal, 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017PC0010.
[31] European Comission. The EU Internet handbook. http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm.
[32] European Comission (Algolia). GDPR Article 9. https://gdpr.algolia.com/gdpr-article-9.
[33] ExoClick. Homepage. https://www.exoclick.com/.
[34] Falahrastegar, M., Haddadi, H., Uhlig, S., and Mortier, R. Tracking personal identifiers across the web. In *International Conference on Passive and Active Network Measurement* (2016), Springer, pp. 30–41.
[35] Fifield, D., and Egelman, S. Fingerprinting web users through font metrics. In *International Conference on Financial Cryptography and Data Security* (2015), Springer, pp. 107–124.
[36] Fouad, I., Bielova, N., Legout, A., and Sarafijanovic-Djukic, N. Tracking the pixels: Detecting web trackers via analyzing invisible pixels. *arXiv preprint arXiv:1812.01514* (2018).
[37] Futurism. Data From British Porn Viewers Might Be In The Hands of One Company, 2018. https://futurism.com/mindgeek-monopoly-uk-porn-viewers-data.
[38] Gomez, R. A. Protecting minors from online pornography without violating the first amendment: Mandating an affirmative choice. *SMU Sci. & Tech. L. Rev. 11* (2007), 1.
[39] Google. Google Policies Help - Adult Content. https://support.google.com/adspolicy/answer/6023699?hl=en, 2019. Accessed: February 12, 2019.
[40] Google Developers. Cookie matching. https://developers.google.com/authorized-buyers/rtb/cookie-guide.
[41] Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K. G., and Aberer, K. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th {USENIX} Security Symposium ({USENIX} Security 18)* (2018), pp. 531–548.
[42] IAB and PWC. The Official xHamster 2019 Trend Report, 2019. https://xhamster.com/blog/posts/911001.
[43] IAB and PWC. WebRTC, 2019. https://webrtc.org/.
[44] Indian Government. The Personal Data Protection Bill. https://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf.
[45] International Amnesty. Uganda's new anti-human rights laws aren't just punishing LGBTI people. https://www.amnesty.org.uk/uganda-anti-homosexual-act-gay-law-free-speech.
[46] International Association of Privacy Profesionals. GDPR matchup: Japan's Act on the Protection of Personal Information. https://iapp.org/news/a/gdpr-matchup-japans-act-on-the-protection-of-personal-information/.
[47] Iordanou, C., Smaragdakis, G., Poese, I., and Laoutaris, N. Tracing cross border web tracking. In *Proceedings of the Internet Measurement Conference 2018* (2018), ACM, pp. 329–342.
[48] Iqbal, U., Shafiq, Z., Snyder, P., Zhu, S., Qian, Z., and Livshits, B. Adgraph: A machine learning approach to automatic and effective adblocking. *arXiv preprint arXiv:1805.09155* (2018).
[49] Juicy Ads. Homepage. https://www.juicyads.com/.

[50] Khan, M. T., DeBlasio, J., Voelker, G. M., Snoeren, A. C., Kanich, C., and Vallina-Rodriguez, N. An empirical analysis of the commercial vpn ecosystem. In *Proceedings of the Internet Measurement Conference 2018* (2018), ACM, pp. 443–456.

[51] Krishnamurthy, B., and Wills, C. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th international conference on World wide web* (2009), ACM, pp. 541–550.

[52] Kulyk, O., Hilt, A., Gerber, N., and Volkamer, M. "this website uses cookies": Users' perceptions and reactions to the cookie disclaimer,". In *European Workshop on Usable Security (EuroUSEC)* (2018).

[53] Legislation.gov.uk. Digital Economy Act 2017. http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted.

[54] Lerner, A., Simpson, A. K., Kohno, T., and Roesner, F. Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th {USENIX} Security Symposium ({USENIX} Security 16)* (2016).

[55] Levenshtein, V. I. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady* (1966).

[56] Luxembourg times. Porn empire reports half billion dollars in revenue – but ends year with loss, 2018. https://luxtimes.lu/luxembourg/33248-porn-empire-reports-half-billion-dollars-in-revenue-but-ends-year-with-loss.

[57] Marotta-Wurgler, F. Self-regulation and competition in privacy policies. *The Journal of Legal Studies 45*, S2 (2016), S13–S39.

[58] MaxMind. Detect online fraud and locate online visitors. https://www.maxmind.com/en/home.

[59] McDonald, A., Bernhard, M., Valenta, L., VanderSloot, B., Scott, W., Sullivan, N., Halderman, J. A., and Ensafi, R. 403 forbidden: A global view of cdn geoblocking. In *Proceedings of the Internet Measurement Conference 2018* (2018), ACM, pp. 218–230.

[60] Miramirkhani, N., Starov, O., and Nikiforakis, N. Dial one for scam: A large-scale analysis of technical support scams. *arXiv preprint arXiv:1607.06891* (2016).

[61] Mowery, K., and Shacham, H. Pixel perfect: Fingerprinting canvas in html5. *Proceedings of W2SP* (2012), 1–12.

[62] Mozilla. Same-origin policy. https://developer.mozilla.org/en-US/docs/Web/Security/Same-origin_policy.

[63] Mughees, M. H., Qian, Z., and Shafiq, Z. Detecting anti ad-blockers in the wild. *Proceedings on Privacy Enhancing Technologies 2017*, 3 (2017), 130–146.

[64] mypornbible. My porn bible. mypornbible.com.

[65] New York Magazine. The Geek-Kings of Smut. http://nymag.com/news/features/70985/index4.html#.

[66] NordVPN, 2019. https://nordvpn.com.

[67] Okoyomon, E., Samarin, N., Wijesekera, P., Elazari Bar On, A., Vallina-Rodriguez, N., Reyes, I., Feal, Á., and Egelman, S. On the ridiculousness of notice and consent: Contradictions in app privacy policies. In *Workshop on Technology and Consumer Protection* (2019), ConPro '19.

[68] OpenNet Initiative. Iraq. https://opennet.net/research/profiles/iraq#footnote24_is5a386.

[69] Oracle. Oracle Data Marketplace. https://docs.oracle.com/en/cloud/saas/data-cloud/data-cloud-help-center/Help/AudienceDataMarketplace/AudienceDataMarketplace.html.

[70] Papadopoulos, P., Kourtellis, N., and Markatos, E. P. Cookie synchronization: everything you always wanted to know but were afraid to ask. *arXiv preprint arXiv:1805.10505* (2018).

[71] Poese, I., Uhlig, S., Kaafar, M. A., Donnet, B., and Gueye, B. Ip geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review 41*, 2 (2011), 53–56.

[72] Pornhub. 2018 Year in Review, 2018. https://www.pornhub.com/insights/2018-year-in-review#2018.

[73] PrivateVPN, 2019. https://privatevpn.com/.

[74] Pujol, E., Hohlfeld, O., and Feldmann, A. Annoyed users: Ads and ad-block usage in the wild. In *Proceedings of the 2015 Internet Measurement Conference* (2015), ACM, pp. 93–106.

[75] Razaghpanah, A., Li, A., Filasto, A., Nithyanand, R., Ververis, V., Scott, W., and Gill, P. Exploring the design space of longitudinal censorship measurement platforms. *arXiv preprint arXiv:1606.01979* (2016).

[76] Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., Kreibich, C., and Gill, P. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem.

[77] Reuters. Uganda's "kill the gays" bill shelved again. https://af.reuters.com/article/topNews/idAFJOE74C0HP20110513.

[78] Richter, P., Wohlfart, F., Vallina-Rodriguez, N., Allman, M., Bush, R., Feldmann, A., Kreibich, C., Weaver, N., and Paxson, V. A multi-perspective analysis of carrier-grade nat deployment. In *Proceedings of the 2016 Internet Measurement Conference* (2016), ACM, pp. 215–229.

[79] Roelleke, T., and Wang, J. Tf-idf uncovered: A study of theories and probabilities. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2008), SIGIR '08, ACM, pp. 435–442.

[80] Roesner, F., Kohno, T., and Wetherall, D. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation* (2012), USENIX Association, pp. 12–12.

[81] Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S. D., and Vallina-Rodriguez, N. A long way to the top: Significance, structure, and stability of internet top lists. In *Proceedings of the Internet Measurement Conference 2018* (New York, NY, USA, 2018), IMC '18, ACM, pp. 478–493.

[82] Selenium. What is Selenium? https://www.seleniumhq.org/.

[83] Story, P., Zimmeck, S., and Sadeh, N. Which apps have privacy policies? In *Annual Privacy Forum* (2018), Springer, pp. 3–23.

[84] The Guardian. Gay relationships are still criminalised in 72 countries, report finds, 2017. https://www.theguardian.com/world/2017/jul/27/gay-relationships-still-criminalised-countries-report.

[85] The Independent. Porn website age verification tool officially announced within UK, 2018. https://www.independent.co.uk/life-style/porn-age-verification-tool-uk-announcement-pornhub-ageid-adult-content-websites-mindgeek-a8242476.html.

[86] The Next Web. The (almost) invisible men and women behind the world's largest porn sites, 2016. https://thenextweb.com/insider/2016/03/03/the-almost-invisible-men-and-women-behind-the-worlds-largest-porn-sites/.

[87] Top websites. Adult catogory. Alexa. http://alexa.com/topsites/category/Top/Adult.

[88] Trevisan, M., Traverso, S., Bassi, E., and Mellia, M. 4 years of eu cookie law: Results and lessons learned. *Proceedings on Privacy Enhancing Technologies 2019*, 2 (2019), 126–145.

[89] Tyson, G., Elkhatib, Y., Sastry, N., and Uhlig, S. Demystifying porn 2.0: A look into a major adult video streaming website. In *Proceedings of the 2013 conference on Internet measurement conference* (2013), ACM, pp. 417–426.

[90] USA Today. China creates stern Internet, e-mail rules. https://usatoday30.usatoday.com/tech/news/2002/01/18/china-internet.htm.

[91] Vallina-Rodriguez, N., Sundaresan, S., Kreibich, C., and Paxson, V. Header enrichment or isp enrichment?: Emerging privacy threats in mobile networks. In *Proceedings of the 2015 ACM SIGCOMM Workshop on Hot Topics in Middleboxes and Network Function Virtualization* (2015), ACM, pp. 25–30.

[92] Vasey, P. L., and Abild, M. A billion wicked thoughts: What the internet tells us about sexual relationships, 2013.

[93] verification under the Digital Economy Act 2017, A. BBFC. https://www.ageverificationregulator.com/.

[94] Vice. Russians now need a passport to watch Pornhub. https://news.vice.com/en_us/article/kzgkv3/russians-now-need-a-passport-to-watch-pornhub.

[95] Virus Total. Virus Total. https://www.virustotal.com.

[96] Wang, L., Dyer, K. P., Akella, A., Ristenpart, T., and Shrimpton, T. Seeing through network-protocol obfuscation. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (2015), ACM, pp. 57–69.

[97] Wolfie Christl. Corporate Surveillance in Everyday Life. https://crackedlabs.org/en/corporate-surveillance.

[98] Wondracek, G., Holz, T., Platzer, C., Kirda, E., and Kruegel, C. Is the internet for porn? an insight into the online adult industry. In *WEIS* (2010).

[99] Yadav, T. K., Sinha, A., Gosain, D., Sharma, P. K., and Chakravarty, S. Where the light gets in: Analyzing web censorship mechanisms in india. In *Proceedings of the Internet Measurement Conference 2018* (2018), ACM, pp. 252–264.

[100] Yu, Z., Macbeth, S., Modi, K., and Pujol, J. M. Tracking the trackers. In *Proceedings of the 25th International Conference on World Wide Web* (2016), International World Wide Web Conferences Steering Committee, pp. 121–132.

[101] Zaeem, R. N., and Barber, K. S. A study of web privacy policies across industries. *J Info. Priv. Sec* (2017), 1–17.

[102] ZINGY ADS. Homepage. http://zingyads.com/.