Privacy on Adult Websites

Ibrahim Altaweel
Good Research LLC
UC Santa Cruz
Santa Cruz, United States

Maximilian Hils
Good Research LLC
University of Munster
Munster, Germany

Chris Jay Hoofnagle
School of Information,
School of Law, UC Berkeley
Berkeley, United States

***Abstract*— As it does in polite conversation, pornography goes unmentioned in policy discussions. This paper begins a conversation about this major use of the web, one that is sensitive and could lead to embarrassment or even blackmail of users if publicized. In countries where pornography is illegal, tracking of these behaviors could have profound consequences for users. Viewing such material is legal in the United States, yet authorities may wish to avoid the topic of protecting its consumers.**

We document and discuss the user tracking dynamics on the most popular adult-oriented websites (N=11). Tracking dynamics are different on adult sites than other popular sites. There are relatively fewer third-party tracking companies involved and fewer cookies on adult sites than on comparably popular sites. However, we found that Google trackers (Google Analytics and/or DoubleClick) were present on almost all the sites and that search terms were often leaked in plaintext to third parties and sometimes encoded in cookies. Finally, the dominance of video in pornography could explain the presence of Flash on almost half the sites. We found Flash being used to read HTTP cookie values, but we did not find any evidence of Flash cookies respawning.

Keywords—*adult websites, pornography, privacy, tracking, extortion, HTTPS, search leakage*

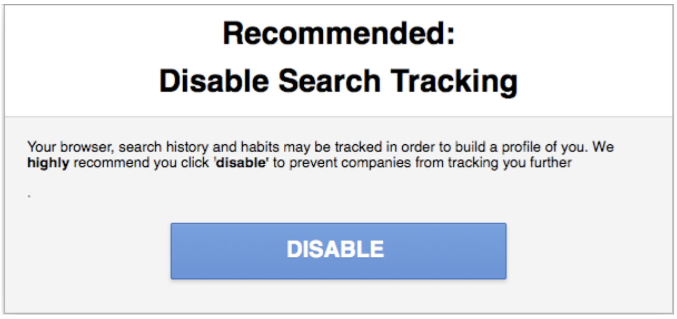
I. Introduction

In this paper, we analyze web tracking on the most popular adult websites in the United States. This paper focuses on just eleven (N=11) websites — every adult-oriented website ranked in the top 500 US sites by Alexa.

It is important to analyze these sites because pornography appears to be a major use of the web. Even though statistics on the amount of the web that is pornographic are not very reliable [1],[2], some adult-oriented websites clearly have a large amount of traffic, based on their relative rankings among the most frequented sites in the US. According to Alexa, the most visited site in our study is on par with Buzzfeed.com in popularity. The least visited site in our study is still more popular than Vox, Disney Go, PBS, and Mit.edu.

These rankings indicate that there is a great deal of pornography consumption online, yet our society has strongly expressed preferences for condemnation of and even prohibition of pornography. In May 2016, the Gallup Poll found that 61% of Americans think that pornography is morally wrong [3]. This figure is remarkably consistent: Gallup has found similar numbers going back to 2011. About a third of Americans favor laws that prohibit distribution of pornography to adults [4], and this sentiment appears to have support even among 18–24-year-old respondents [5]. Of course, such restrictions almost always run afoul of the First Amendment. However, legal protection for consumption is not the norm in all countries, and in some, pornography is filtered or is prohibited by the criminal law. In September 2016, BBC News reported that Russian media regulator Roskomnadzor ordered the nation's ISP's to filter certain popular pornography sites [6].

The moral disapprobation of pornography is so strong that it seems to also cause people to deny that they use the internet to visit adult websites. As late as 2005, the Pew Internet & American Life Project found that 87% of internet users claimed not to view adult websites [7]. Thus, it is obvious that there are many users of adult websites who do not want others to know about their pornography consumption. Consider that in the abstract, we know that married couples have intercourse, and we even arrange public policy to encourage childbearing, yet direct or indirect evidence of intercourse is embarrassing and most people are discreet about it.



Recommended:
Disable Search Tracking

Your browser, search history and habits may be tracked in order to build a profile of you. We **highly** recommend you click '**disable**' to prevent companies from tracking you further

.

Fig. 1. *In our testing we found that some sites gave users privacy advice. This popup led to a search company (https://www.searchincognito.com/) that promised more privacy in queries.*

Similarly, consumption of pornography can be embarrassing in itself as it can reveal sexual preferences or fantasies. See Figure 1, for instance, as an example of a pornographic website giving privacy advice to consumers. However, the academic literature is thin on consumers' privacy expectations when consuming adult material online as well as the technical tracking that adult sites employ [8].

There is reason for concern. Revelations of pornography consumption can have effects on people in profound ways. Professor Andrew Gilden recounts several examples of how pursuing online sexual fantasy has influenced real-world legal relationships, such as in custody battles, divorce proceedings, and in criminal trials, where fantasy is used as propensity evidence [9]. The data can even be used for fishing expeditions. Recall that ten years ago, the Department of Justice issued subpoenas to several search engines in order to obtain user data about the prevalence of pornography online [10]. Google laudably fought the subpoena, but at the same time, the company would not characterize the search terms involved as "private" information.

Pornography consumption can also be used to infer other facts about an individual that could be used to extort or embarrass a person. One need only look to the fallout from the recent Ashley Madison breach, where the information of tens of millions of users of the website for extramarital affairs was leaked online [11]. The breach involved a public leak, one where anyone could have downloaded the user corpus. But despite the availability of the data, criminals could still approach individual victims and extort them with threats of giving broader publicity to the data, such as by telling coworkers or family members about membership on the site. Turning to private-sector tracking of pornography consumption, private leakage and tracking also increases the risk of extortion. One site leaked data to Russia-based Yandex. We found that nine of the websites we visited had a Google tracking script (DoubleClick or Google Analytics) and that seven leaked search terms to third parties and/or coded pornographic search terms into cookies. Some of these parties, particularly Google, could trivially and secretly re-identify these users by relying on data collection from other sites. One could imagine the uses — for instance, would you sue a company for privacy violations if in discovery, the company suggested that it could use its extensive activity logs to elucidate one's web use at a public trial?

For these reasons, we focus here on the kinds of tracking and other privacy-relevant technical activity on websites featuring pornography.

II. Methods

In previous studies, we employed OpenWPM to perform large-scale crawls of the web to count HTTP, Flash, and HTML5 cookies [12]. Because this study involved a very small sample (N=11), we performed a manual crawl using Firefox, and documented data flows with mitmproxy [13]. The sample was comprised of all eleven adult websites that appeared in the Alexa Top 500 US-ranked websites as of September 2016. We focused on this small number of popular sites because although the Web has a "long-tail," user attention is strongly focused on the most popular websites. We directly typed in the target URL and selected three links that pointed to resources on the same domain. We did not play videos.

Table 1: List of Sites Tested

Pornhub.com
Xvideos.com
Xhamster.com
Bongacams.com
Txxx.com
Chaturbate.com
Xnxx.com
Upornia.com
Redtube.com
Livejasmin.com
Youporn.com

We also employed Mezzobit to perform automated analyses of the eleven websites. Mezzobit is a cloud-based crawling platform that assesses privacy, usability, and website performance. The two methods reveal different data because mitmproxy tracks all connections during the browsing session, including popups and other resources that are loaded, while Mezzobit focuses more on the target URL. In so doing, Mezzobit provides enhanced analysis that assists in spotting whether a subdomain is operated by a third party, or whether it is simply cloud infrastructure operated by the first party.

We comptrasted our manual and automated crawl data, performed spot checks using different browsers and ran a separate analysis using Netograph [14]. We used Palantir Contour, a relatively new service offered by Palantir Technologies, to do link analysis on the corpus of data.

III. Results and Discussion

Here, we detail the high-level tracking dynamics on our sample of adult websites.

A. *Third-party Tracking Overview*

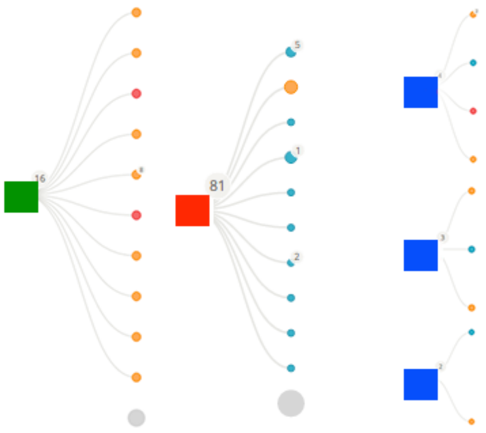


Fig. 2. *Mezzobit nicely illustrates website communication with third parties. Compare a popular medical information website (green square) that has 33 third-party vendors and a website comparably popular as the adult sites we analyzed (red square), which has relationships with 20. The blue squares represent typical popular adult sites – they have far fewer relationships with third parties, having on average four third-party relationships.*

Tracking dynamics are different on adult websites than popular non-pornographic sites [15]. There is a relatively small number of third-party trackers present that appear to specialize in pornographic ads. Adult sites generally lack "social buttons," and just one site had a Facebook tracking script. This is remarkable because in 2015, we found Facebook on over half of top 1,000 most popular websites. AddThis and Twitter buttons were present on a small number of sites. [12]

Aside from Google Analytics and DoubleClick, mainstream behavioral advertisers are also not present on adult sites. When using Mezzobit to analyze tracking on adult sites, the median site in our small sample of adult websites sent data to four third parties, with the highest sending data to ten other third parties and the lowest to just one third party. On average, sites made 25 separate communications with these third parties. However, it is important to note that Mezzobit does not measure all popup sites spawned by the adult sites.

Turning to our analysis using mitmproxy, we find dramatically more third-party tracking than average, but upon inspection, this increase is due to two websites that opened popup windows for the same third-party news website (popularscience.tv), where there was much tracking (in one round of testing, the news website instantiated 448 third party cookies from 135 different third-party hosts).

Table 2: Summary Statistics on Adult Websites

Total number of cookies generated by 11 sites	1092
Average number of cookies	95
Median number of cookies	19
Median number of first party cookies	11
Median number of third party cookies	4
Sites with Flash content	5
Sites that leaked search content in plaintext	7
Sites with HTTPS by default	2

1) *Explaining Third Party Tracking on Adult Websites*

What explains this absence of third party trackers on the adult sites themselves? One hypothesis is that pornography, as a subject matter, is simply too personal and creepy to track. In her survey of privacy policies in seven markets, Professor Florencia Marotta-Wurgler found that adult website privacy policies were more likely to comply with Federal Trade Commission recommendations in several respects than other non-adult sites. Specifically, she found highly-ranked adult sites (N=17) to be more restrained in data collection and sharing and to have shorter, yet more rights-protective privacy notices [16]. The market for pornography could simply demand that adult sites are more private.

At the same time, if adult sites are too creepy to track, it stands to reason that the market would also pressure medical sites to limit tracking. However, as Figure 2 shows, a top 500 medical website sponsored by respected mainstream medical institutions has far more third-party tracking—Mezzobit reported 32 third-party vendors and the site placed over 30 cookies—than any adult website we analyzed.

Another more powerful explanation is that adult websites have low utility for non-pornography advertising. Reputation is a key issue, as mainstream advertisers probably do not want their products displayed next to pornographic content. Also, it could be that preferences for pornography simply do not have relevance for targeting non-pornographic ads. Furthermore, as Dr. Kate Darling explains, there is a lack of trust in pornographic websites because of historical problems of unauthorized charges and malware [1]. Medical websites on the other hand are more trusted and offer more opportunities to mention specific products.

Finally, adult websites compete with free resources, as observed by Professor Benjamin Edelman, and such free resources may be better from the consumer perspective because of concerns about fraud and monitoring [17]. Obtrusive and obvious tracking might cause users to turn to peer-to-peer compilations of pornography. Yet these "free" resources may have other hidden costs, such as malware.

We did not examine browser fingerprinting, but a testing platform we used, Mezzobit, contains an automated fingerprint risk estimate score. According to its analysis, three sites were more likely than not to fingerprint browsers.

B. *Tracking Mechanisms*

1) HTTP Cookies

Just 44 pages on eleven adult websites generated almost 1,100 cookies, yet we found that there are fewer cookies on adult websites than comparably popular non-adult websites. Our 2015 crawl found an average of 135 cookies (with a median of 91 cookies) on the top 1,000 most popular websites. [12] Adult sites have fewer – the median is 19, while the average is 95. The high average is a reflection of the two sites that popped up a third-party website. When those two sites are excluded, the average lowers to 33, because one of the two tracking-intensive sites had 465 cookies while and the other had 288.

Websites had an average of ten (with a median of 11) first-party cookies, 84 (median 8) third-party cookies, and these third-party cookies were served, on average, by 25 (median 4) hosts. All of the third-party summary statistics are influenced by the two sites with many cookies, and as a result, the median values are more useful here.

2) Flash Cookies

We detected Flash on five of the websites. In most cases, Flash was being used to read HTTP cookie values, usually from the same domain. We found no evidence that Flash was being used to reinstate, or "respawn," deleted HTTP cookies.

3) Local Storage

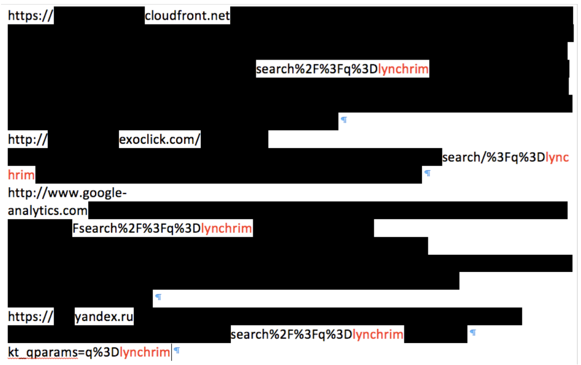
We did not encounter any use of HTML5 local or session storage.

C. *Security Issues*

1) Plain Text Search Term Leakage

We found that seven sites "leaked" search terms "in the clear. [18] That is, if a user visited a site and performed a search, the search query was transmitted to third parties in plain text.

Additionally, the search term was often encoded into a cookie in plain text. Recipients of search terms included Google (both Analytics and DoubleClick), Russia-based Yandex, and other marketing and ad tech services. In addition to search terms, "category" tags were often encoded in plain text, meaning that a click on a specific interest ("blonde," "trans," and so on) were also transmitted in plain text rather than as a code (e.g. category "38273").



https://[redacted]cloudfront.net[redacted]search%2F%3Fq%3Dlynchrim[redacted]¶
http://[redacted]exoclick.com/[redacted]search/%3Fq%3Dlynchrim[redacted]¶
http://www.google-analytics.com[redacted]Fsearch%2F%3Fq%3Dlynchrim[redacted]¶
https://[redacted]yandex.ru[redacted]search%2F%3Fq%3Dlynchrim[redacted]¶
kt_qparams=q%3Dlynchrim¶

Fig. 3. T*he search term "lynchrim" is leaked in URLs (red text) to Google and Russia-based Yandex, and sometimes encoded in plain text in cookies.*

2) Lack of HTTPS

In retail stores, pornography and sex toys are typically sheathed in a brown paper bag before leaving the store. Purchasers can pay cash and use the product in the privacy of their home. Protecting one's privacy was as straightforward as being discreet in a context that ordinary people understand.

Turning to the internet, the online equivalent of the brown paper bag is HTTP over SSL (HTTPS). Network providers can tell that the user visited a website, but HTTPS would protect the specific pornographic content consumed while it was sent from its source to the user.

We found surprising dynamics concerning HTTPS. One might assume that adult websites would use HTTPS in order to limit ISP and other monitoring. However, only two of the eleven sites we tested used HTTPS by default for content delivery, and one provided HTTPS at the user's election. The remaining eight either would not load or would forward the user to HTTP if the website URL was entered with HTTPS manually.

Additionally, we found that various individual communications used HTTPS on almost every site, but more likely than not, the communications secured by HTTPS were those of third party trackers and advertising delivery companies. That is, on these sites, we found that tracking efforts and ad delivery were sometimes transmitted over HTTPS, while adult content was delivered over HTTP.

We think this lack of HTTPS is an important privacy problem that users may not understand. Many intermediaries, be it the WLAN operator or intelligence agencies, can view preference and even second-by-second decisions about consumption. A reputable advocacy group, the Center for Democracy & Technology, has announced an initiative to increase adoption of HTTPS on adult websites [19]. Without these protections, the user may feel private, because one can view these sites in seclusion, protected by the walls of the home. In reality, the user is consuming the product before a one-way mirror.

IV. CONCLUSION

As this paper is being submitted, a pair of lawyers pled guilty to conspiracy charges in a scheme where they uploaded pornography to peer-to-peer services, waited for users to download it, and then sued the users for copyright infringement. The lawyers assumed users would be either too poor to afford a good defense, or too embarrassed to admit in public court filing that they had downloaded pornography. The lawyers collected $6 million in the scheme, often in $3,000 "settlements" with users.[20]

In a world with criminal prosecution for pornography possession, of large-scale leaks of private information, and of growing extortion attempts based on personal information collected online, we need to consider whether consumption of pornography, a popular, yet unsympathetic web activity, deserves attention from consumer protection authorities.

In this brief paper, we explored the kinds and of amount of tracking on popular adult websites. Adult websites have a smaller universe of trackers than popular non-adult websites. This lower level of tracking is best explained by the marketability of adult websites. Mainstream advertisers do not want their content placed next to pornography, nor are they likely to be able derive actionable marketing intelligence from users' specific preferences for adult material.

However, several privacy risks are present: search terms and category tags, which may reveal sexual fantasy, are leaked in most cases in the clear and to third parties. Furthermore, just a handful of sites use HTTPS, leaving full URL strings visible for monitoring by others.

The academic literature is thin on privacy expectations in porn consumption. Our work complements Professor Marotta-Wurgler's on privacy policies, in that it adds technical analysis of adult websites' functioning that are consonant with her findings of adult sites' stated policies.

A. Authors and Affiliations, Conflicts of Interest

Ibrahim Altaweel is a student at University of California, Santa Cruz; Web Security Engineer at Good Research LLC. Altaweel's work has been supported by NSF.

Maximilian Hils is a graduate Student at University of Münster; Visiting Scholar, UC Berkeley School of Information; mitmproxy Core Developer. His open-source developments have been supported by Google and other technology-industry companies.

Chris Jay Hoofnagle is an adjunct full professor at UC Berkeley School of Information and School of Law. Investigator Hoofnagle's work has been supported by NSF, and by technology- industry companies, including Microsoft, Google, and Nokia. He is of counsel to Gunderson Dettmer, LLP, a firm that concentrates on representing emerging technology companies.

We conceived of and executed this study on our own initiative, without external financial support. However, we have significant technology donations from Palantir Technologies and Mezzobit.

We thank Nathaniel Good for feedback and advice on this project.

REFERENCES

[1] K. Darling, "IP Without IP? A Study of the Online Adult Entertainment Industry," Stan. Tech. L. Rev., vol. 17, pp. 709-771, Nov. 2014.

[2] S. Tarrant, *The Pornography Industry: What Everyone Needs to Know.* Oxford, United Kingdom: Oxford Univ. Press, 2016.

[3] Gallup Organization. Gallup Poll, May, 2016 [survey question]. USGALLUP.052616.R21Q. Gallup Organization [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, iPOLL [distributor], accessed Sep-17-2016.

[4] National Opinion Research Center, University of Chicago. General Social Survey 2014, Mar, 2014 [survey question]. USNORC.GSS14B.Q102. National Opinion Research Center, University of Chicago. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, iPOLL [distributor], accessed Sep-17- 2016.

[5] Berkley Center for Religion, Peace, and World Affairs at Georgetown University. Millennial Values Survey, Mar, 2012 [survey question]. USPRRI.12MILVAL.R22D. Public Religion Research Institute [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, iPOLL [distributor], accessed Sep-17- 2016

[6] BBC NEWS, *Russia extends porn site ban*, Sept. 16, 2016.

[7] Pew Internet & American Life Project. Pew Internet & American Life Project Poll, May, 2005 [survey question]. USPSRA.070605.WEBA12. Princeton Survey Research Associates International [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, iPOLL [distributor], accessed Sep-17- 2016.

[8] J. Kinsley, "Sexual Privacy in the Internet Age: How Substantive Due Process Protects Online Obscenity," *Vand. J.l Ent. & Tech.y L.,* vol. 16, no. 103, pp. 103-131, Feb. 2013.

[9] A. Gilden, *Punishing Sexual Fantasy*, William & MaryL. Rev., vol. 58, pp. 419-491, Sept. 2016.

[10] N. Richards, *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age.* Oxford, United Kingdom: Oxford Univ. Press, 2015.

[11] K. V. Brown. (2015, Dec. 9). Scared, dead, relieved: How the Ashley Madison hack changed its victims' lives. *Fusion* [Online]. Available: http://fusion.net/story/242502/ashley-madison-hack-aftermath/

[12] I. Altaweel, N. Good, and C. J. Hoofnagle. (2015, Dec. 15). Web Privacy Census. *Technology Science* [Online]. Available: http://techscience.org/a/2015121502

[13] A. Cortesi, M. Hils, T. Kriechbaumer, et al. mitmproxy 0.17. https://mitmproxy.org/

[14] A. Cortesi, netograph, http://netograph.io/.

[15] S. Englehardt and A. Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis," presented at the 23rd ACM Conference on Computer and Communications Security, Vienna, Austria, 2016.

[16] F. Marotta-Wurgler, "Understanding Privacy Policies: Content, Self-Regulation, and Markets," NYU L. & Econ. Research Paper No. 16-18, Jan. 2016.

[17] B., "Red Light States: Who Buys Online Adult Entertainment," *J. Econ. Perspectives, vol. 23, no. 1,* pp. 209-220, 2009. *See also* Gilbert Wondracek et al., *Is the internet for porn? An insight into the online adult industry*, WEIS 2010, 9th Workshop on the Economics of Information Security, Jun. 7–8 2010, Boston, USA.

[18] B. Krishnamurthy, K. Naryshkin, and C. Wills, "Privacy leakage vs. Protection measures: The growing disconnect," IEEE Secur. Priv., vol. 11, no. 3, 2011

[19] J. L.o Hall and G. Norcie. (2016, Oct. 7). It's Time to Move to HTTPS. *CDT.* [Online]. Avaliable: https://cdt.org/blog/its-time-to-move-to-https/

[20] U.S. v. John L. Steele, 16-cr-00334 (D.C. Minn. 2017)