Privacy on Adult Websites

Ibrahim Altaweel, UC Berkeley School of Information & UC Santa Cruz School of Engineering
Maximilian Hils, UC Berkeley School of Information & Good Research
Chris Jay Hoofnagle, UC Berkeley School of Information & School of Law

INTRODUCTION

Many people use the web to consume pornography, yet little is known about how users are tracked on pornographic websites.

Consumption of pornography is legal in the US, yet pornography goes unmentioned in policy discussions. Our research begins a conversation about this major use of the web, one that is sensitive and could lead to embarrassment and harm to users if publicized. 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 where fantasy is used as propensity evidence in criminal trials. Pornography consumption can also be used to infer other facts about an individual that could be used to extort or embarrass a person.

METHODS

We analyzed every adult-oriented website ranked in the top 500 US sites by Alexa **(N=11)**. 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.

Manual Crawl

We visited the sites with Firefox in a clean state and documented with mitmproxy.

Automated Crawl

We used Mezzobit, Netograph, and supplemented these tools with Palantir Contour for link and statistical analysis.

ACKNOWLEDGMENTS

We thank Nathaniel Good for feedback and advice on this project. We thank Palantir and Mezzobit for their significant technology donations.

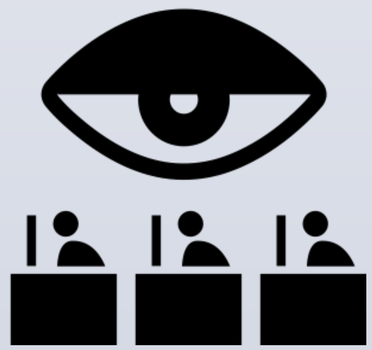
RESULTS AND DISCUSSION

Search Term Leakage

Seven sites "leaked" search terms "in the clear."

Fig 1: The search term "lynchrim" is leaked in URLs (red text) to Google and Russia-based Yandex, and sometimes encoded in plain text in cookies.

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").



Third-party Tracking

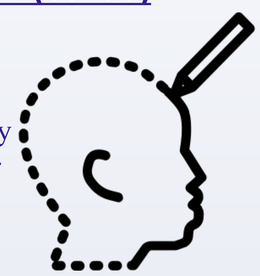
- Google trackers were present on almost all the sites.
- A small number of third-party trackers present that appear to specialize in pornographic ads.
- Adult sites generally lack "social buttons," and just one site had a Facebook tracking script.
- AddThis and Twitter buttons were present on a small number of sites.

Local Storage

No HTML5 local or session storage.

RESULTS AND DISCUSSION (Cont'd)

HTTPS



Only two of the eleven sites used HTTPS by default. This means that many intermediaries, be it the WLAN operator or intelligence agencies, can view preference and even second-by-second decisions about consumption.

HTTP Cookies

Fewer cookies on adult websites than comparably popular non-adult websites.

Just 44 pages on eleven adult websites generated almost 1,100 cookies.

Websites had an average of ten first-party cookies, 84 third-party cookies, and these third-party cookies were served on average by 25 hosts.

Flash Cookies

Flash **on five of the eleven websites**. Flash was being used to read HTTP cookie values.

No evidence that Flash was being used to reinstate, or "respawn," deleted HTTP cookies.

CONCLUSION

- Search terms and category tags, which may reveal sexual fantasy, are leaked in the clear to third-parties.
- Just a handful of sites use HTTPS, leaving full URL strings visible for monitoring by others.
- Our results point to the need for careful consideration of whether consumption of pornography, deserves attention from consumer protection authorities.