

# Everybody's got something to hide: Analysis of the Online Adult Ecosystem

Pelayo Vallina <sup>\*†</sup>, Alvaro Feal <sup>\*†</sup>, Julien Gamba <sup>\*†</sup>, Narseo Vallina-Rodriguez <sup>\*‡</sup>, and Antonio Fernandez Anta <sup>\*</sup>

<sup>\*</sup>IMDEA Networks <sup>†</sup>Universidad Carlos III de Madrid <sup>‡</sup>ICSI

## Online Adult Ecosystem

Porn websites are among the most visited websites globally according to Alexa's Rank (10/24/2018):

- 31 porn sites appear amongst the top-1,000 sites.
- pornhub.com is ranked in position 30.

Adult websites pose various privacy risks which remain largely unknown: **online services in the porn ecosystem can infer sensitive information such as the sexual behavior, orientation and preferences.**

Many studies have analyzed privacy risks in the web, but there is a dearth of knowledge about the privacy implications of adult websites and the third-party tracking ecosystem supporting it.

- Large advertising and analytic services operating in the web don't provide support to adult websites to prevent brand reputation loss.
- Modern personal data protection laws – e.g., GDPR – prohibit the collection and processing of sexual information to build user profiles [1].
- New digital legislation are being proposed to prevent access to adult content to minors in UK [2].

## Goals

Due to the lack of knowledge of the Online Porn Ecosystem, we aim to:

- Identify its actors, and their role, their tracking and profiling activities, and measure the differences with the web ecosystem.
- Study organization's regulatory compliance with the different digital regulations.
- Study organization's capacity to adapt to new regulatory frameworks.

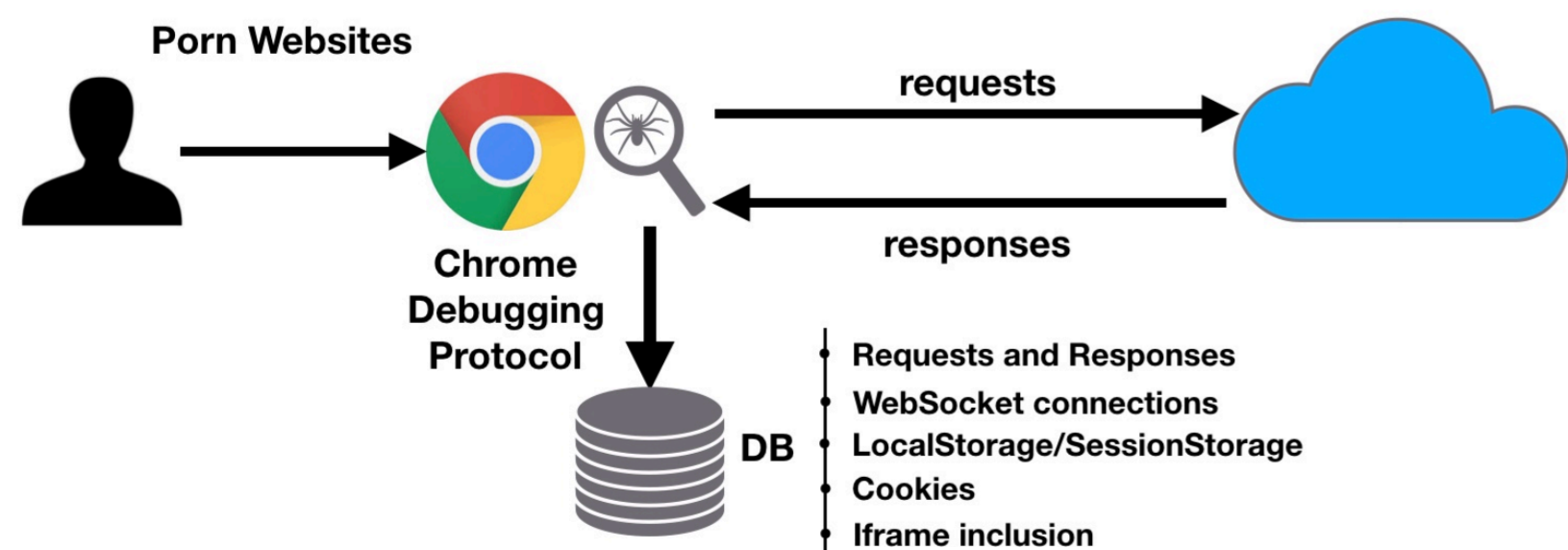
## Methodology

We have built a corpus of 6,744 different porn sites collected from different sources:

1. Websites specialized in recommending and classifying pornographic content like "toppornsites.com" or "only4adults.net".
2. Analyzing the top 100,000 Alexa rank and select those websites with keywords related to pornographic and adult content – e.g., "porn" or "sex".
3. **Removing false-positives:** we use a web crawler to took a screenshot of the landing page and manually check whether it contains pornographic content.

We implemented a crawler using "Chrome-Debugging-Protocol" that visits the main page of each website. This allows us to collect information both at the browser- and network-level such as:

1. All the network requests and responses generated from the browser.
2. All the cookies and permanent browser storage information like "LocalStorage".
3. All incoming communications to detect and analyze "cookie sync-ing".
4. All the websocket connections and the data transferred through it.
5. Analyze the iframe inclusion.
6. Take screenshots of the main page to detect and study consent forms.



We collect all the privacy policies of each website using a crawler based on Selenium WebDriver to:

1. Identify if they have privacy policies.
2. Compare empirical observations with privacy claims and assess if they comply with their own policies and privacy laws.

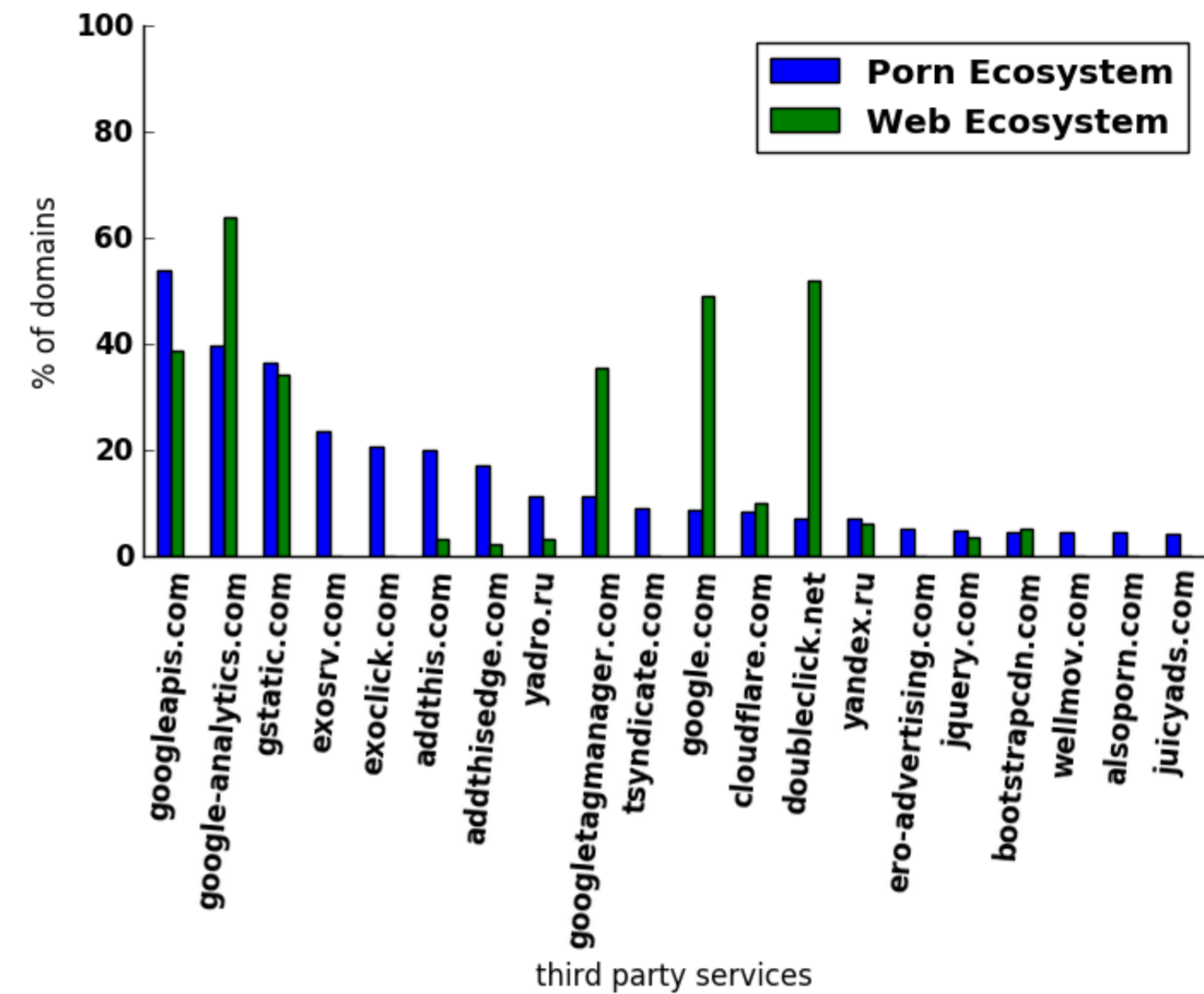
We use third-party services to analyze the nature of each third-party service presented in websites.

1. VirusTotal: Aggregates information from multiple antivirus and online scan engines.
2. OpenDNS: Categorizes websites.
3. EasyList: Identify trackers and advertisement services.

## Web Ecosystem VS Online Porn Ecosystem

We compare our corpus of 6,744 adult websites against the top-7K ones according to Alexa's rank. We exclude pornographic websites from the top 7K.

	#Third-party Services	% Unique	#Trackers- Advertisement	%Unique
Porn Ecosystem	3,191	86%	112	48%
Web Ecosystem	7,362	94%	439	86%



According to VirusTotal, 18 domains, present in 36 websites, are potentially malicious.

- Notably, "coinhive.com" (crypto-mining service) is presented in 11 porn website wherein the web ecosystem only appears twice.

## Profiling in Porn Ecosystem

We found third-party cookies in more than 57% porn websites delivered by 700 different third-parties.

- There are 287 different third-party cookies that appear in more than 10 porn websites, where 6 of them appear in more than 1K different porn sites.
- Two of them belong to advertisement services ("exosrv.com" and "exoclick.com") and the other 4 to "addthis.com", a social bookmarking service.

We found 7 porn websites collecting geo-location and IP-based location of the user in the localStorage without explicit consent.

All these information together can pinpoint a single user, and according to GDPR should be collected only when the user grants explicit consent. We didn't found any consent form.

## Conclusions and Future Work

- Users of pornographic websites can be exposed to privacy risks and profiling by not well understood online services and companies.
- Use of black-box testing techniques to infer user profiling on the server-side and personal data dissemination between companies.
- Study geographical bias and capacity to adapt to different regulatory jurisdictions. We will run more experiments from different geographical locations using commercial VPN services to gain vantage points at a global scale.
- Study the privacy risks of pornographic websites across platforms and browsers. We will extend our analysis by using different type of devices like mobiles and different browsers.
- Assess the effectiveness of the commercial antitracking techniques such as plugins or browser incognito mode to protect user's privacy.
- The online porn ecosystem remains opaque due to the lack of systematic and comprehensive studies focused on this ecosystem.

## References

[1] E. Union, "Processing of special categories of personal data," 2018. [Online]. Available: <http://doi.acm.org/10.1145/2934664>

[2] T. Guardian, "Millions of porn videos will not be blocked by uk online age checks," 2018. [Online]. Available: <https://www.theguardian.com/technology/2018/oct/18/millions-of-porn-videos-will-not-blocked-uk-online-age-checks>