ETH zürich

Department of Computer Science

CookieBlock & CookieAudit: Fixing Cookie Consent with ML

Karel Kubicek, Dino Bollinger, Adrian Zanga, Carlos Cotrini, David Basin (Department of Computer Science, ETH Zürich)

1 Motivation

- Web tracking technologies are ubiquitous: over 90% of websites use cookies for stateful tracking.
- EU ePrivacy Directive (ePD) and GDPR require consent:
  - for data usage that is not *strictly necessary*,
  - that is freely-given, unambiguous, specific, and informed,
  - explicitly and specifically lists processing purposes.

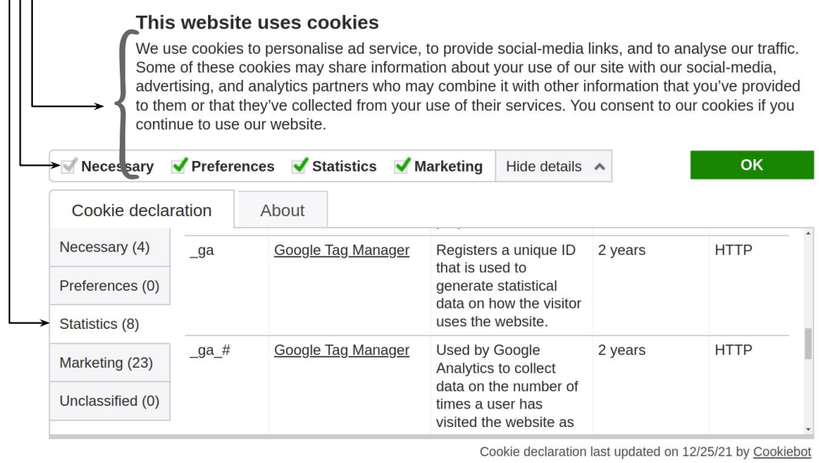

Fig. 1: Cookiebot consent banner that implements a majority of the regulation requirements. Ground truth data source for cookie categories in our study.

- Cookie consents are often noncompliant (from prior studies):
  - Dark patterns successfully nudge users towards consent.
  - GDPR and ePD rules ignored by up to 90% websites [2].
  - Consent is not followed by up to 50% websites [3].

We provide two browser extensions that classify cookies:

CookieBlock enforces user privacy independently of consent [4].

CookieAudit helps web developers fix cookie consent [5].

2 Methods

1. We crawl training data from 30k websites with three different 3rd-party consent providers (e.g., Cookiebot in Fig. 1). Websites declare cookies classified to 4 purposes: *necessary*, *functionality*, *analytics*, and *marketing*. We collected >300k cookies with their purposes.
2. We design 52 feature extraction methods: measuring cookie entropy, detecting dates, language strings, encodings, etc.
3. We train a machine learning model (XGBoost) with an accuracy of 87.2±0.23% in predicting cookie purpose. It reaches a performance comparable to expert analysis.
4. We use this model in CookieBlock and CookieAudit.
5. We report 8 potential privacy violation types from our cookie dataset, see Fig. 2. Shockingly, 94.7% of websites contain at least one of these violations.

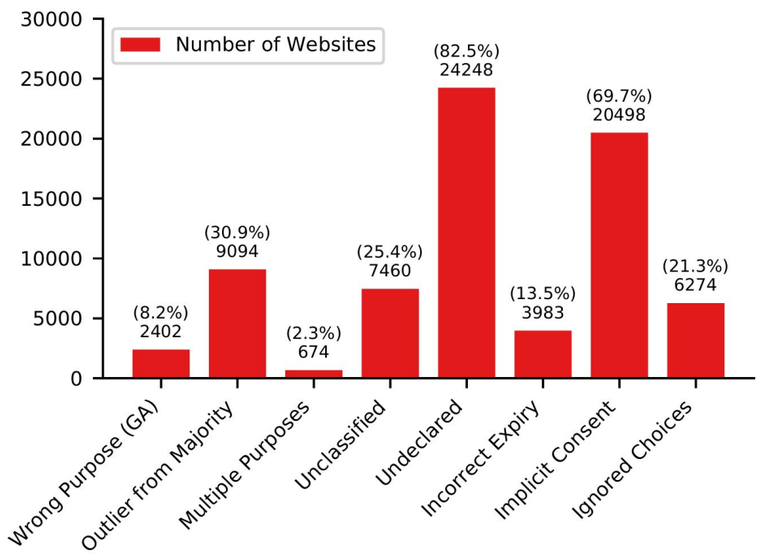

Fig. 2: Number of websites that show the respective type of violation. The first six are novel and have not been explored in prior work. For full details specification of these violation categories, see the publication [1].

3 CookieBlock and CookieAudit

- **CookieBlock** instantly removes cookies with user-rejected purposes, making the consent popups obsolete. These can be removed by e.g., uBlock Origin with annoyance filters [6].
- It prevents all of the potential violations from Fig. 2 and more. It works everywhere in the world, exporting EU privacy level.
- Our evaluation: 85% of websites work as intended, 7% have authentication issues, 8% minor issues (popups reappearing).
- Available to Chrome, Firefox, Edge, and Opera. ~8k users.
- User rating (browser stores, our feedback form): 4.1/5 stars.
- **CookieAudit** targets web developers and data protection agencies (enforcers), allowing users to identify potential violations and informing them how to address these.
- It detects consent (any type listed by annoyance filters [6]), used cookies, and reports potential violations from Fig. 2 and known problems with consent providers.

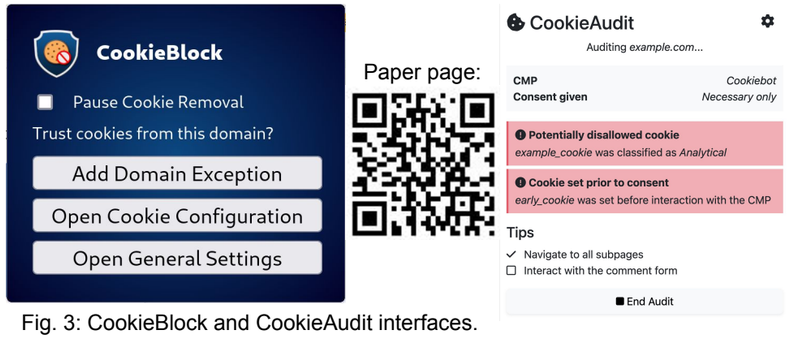

Fig. 3: CookieBlock and CookieAudit interfaces.

References

1. Bollinger, Dino, et al. "Automating cookie consent and GDPR violation detection." 31st USENIX Security Symposium (USENIX Security 22). USENIX Association, 2022.
2. Nouwens, Midas, et al. "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.
3. Trevisan, Martino, et al. "4 Years of EU Cookie Law: Results and Lessons Learned." Proc. Priv. Enhancing Technol. 2019.2 (2019): 126-145.
4. Bollinger, Dino, et al. "CookieBlock repository", online: https://github.com/dibollinger/CookieBlock
5. Zanga, Adrian, et al. "CookieAudit repository", online: https://github.com/Fredilein/CookieAudit
6. Fanboy, "EasyList Cookie filters", online: https://github.com/easylist/easylist/tree/master/easylist_cookie