CookieBlock & CookieAudit: Fixing Cookie Consent with ML

Citation

Bollinger, Dino, Karel Kubicek, Carlos Cotrini, David Basin. "Automating cookie consent and GDPR violation detection." 31st USENIX Security Symposium (USENIX Security 22). USENIX Association, 2022.

Link to the published (official) version of the paper

https://www.usenix.org/conference/usenixsecurity22/presentation/bollinger

Abstract

The European Union's General Data Protection Regulation (*GDPR*) requires websites to inform users about personal data collection and request consent for cookies. Yet the majority of websites do not give users any choices, and others attempt to deceive them into accepting all cookies. We document the severity of this situation through an analysis of potential GDPR violations in cookie banners in almost 30k websites. We identify six novel violation types, such as incorrect category assignments and misleading expiration times, and we find at least one potential violation in a surprising 94.7% of the analyzed websites.

We address this issue by giving users the power to protect their privacy. We develop a browser extension, called CookieBlock, that uses machine learning to enforce GDPR cookie consent at the client. It automatically categorizes cookies by usage purpose using only the information provided in the cookie itself. At a mean validation accuracy of 84.4%, our model attains a prediction quality competitive with expert knowledge in the field. Additionally, our approach differs from prior work by not relying on the cooperation of websites themselves. We empirically evaluate CookieBlock on a set of 100 randomly sampled websites, on which it filters roughly 90% of the privacy-invasive cookies without significantly impairing website functionality.

CookieAudit: new contribution not published in the original work

With CookieAudit, we want to help improve websites' overall privacy compliance, improving the web for those users that do not use CookieBlock. CookieAudit targets two groups - web developers and data protection agencies.

We believe that the non-compliance reported in our study primarily stems from a lack of awareness rather than malicious intent. We therefore want to provide web developers with the browser extension CookieAudit that helps them to detect privacy violations and guides them through the process of resolving them.

At the same time, CookieAudit is also intended to help data protection agencies enforce regulations at scale. A consistent and thorough enforcement is currently hindered by the vast number of domains to be audited, and the comparatively low capacity and funding of these agencies. This forces authorities to focus on major offenders, and allows small to medium-sized domains to ignore the requirements. CookieAudit should ease the process of identifying privacy violations by collecting evidence in a semi-automated procedure, thus aiding authorities in identifying privacy violations more quickly and easily. As a side-effect, a more consistent enforcement should incentivize a stricter adherence to regulations by any website, regardless of size.

Functionality

Users are expected to manually browse the scanned website with CookieAudit active. The extension provides a set of browsing actions, which are recommended in the form of a “To Do list”. These actions include "reject/accept all cookies", "register/login", and "open random subpages". While the user is performing these actions, CookieAudit automatically detects the presence of a cookie consent notice using annoyance removal filters [1]. It also detects whether the website is using one of the top 20 consent management platforms. Lastly, all cookies that are observed during browsing are assigned to usage purposes by the model from CookieBlock. If the website uses a consent management platform that explicitly assigns cookies to purposes in the consent notice, we also match these purposes to the cookies observed on the website.

Once the user is finished with browsing, they can generate a compliance report. We report all potential violations from our study and potential violations caused by design choices of the top 20 consent management platforms (this list is manually curated). Next we report all observed cookies and their classification based on our model. Lastly, we provide recommendations to address observed potential violations as well as recommendations of reducing private data usage in general.

CookieAudit is open source [2], so anyone can propose changes to our report generation.

Note that the described functionality of CookieAudit is a work in progress as of June 27th. The addition of CookieAudit caused a change in the authors list.

References

[1] Fanboy, "EasyList Cookie filters", online: https://github.com/easylist/easylist/tree/master/easylist_cookie

[2] Zanga, Adrian, et al. "CookieAudit repository", online: https://github.com/Fredilein/CookieAudit