Understanding Website Compliance with Web Cookie Privacy Laws

Maxwell Lin, Karen Wang, Helen Wu
Mentors: Dr. Xiaowei Yang, Shihan Lin

Introduction

Websites use cookies for diverse purposes, such as authentication, advertisement, and behavior tracing. This leads to potential vulnerabilities of websites as well as possible violations of users' privacy. The motivation behind this research project is twofold:
- Users should have the right to control how their information is used on the Internet.
- Websites must adhere to privacy laws such as GDPR.

Research Questions

1. What kinds of cookie banners are being provided to users?
2. When users select an option on a cookie banner, are their decisions being respected by the website?

Results

• During a crawl of the landing pages of 255 sites with cookie banners, 100 out of the 255 domains violate GDPR by retaining tracking cookies even after the "Reject" button is clicked.
• 1475 unique tracking cookies were detected across all 255 sites during a normal run.
• 1108 unique tracking cookies were detected across all 255 sites after rejecting all cookies.

Methodology

Web Crawling
- Tracking cookies (also known as "targeting cookies" or "advertising cookies") pose greatest privacy threat as they collect user data to transmit to 3rd parties for commercial purposes.

Selenium-based web crawler intercepts HTTP requests

First run: Crawl each site (and its inner pages if specified) to load cookies into browser.

“Normal” run: Crawl each site (and its inner pages if specified), saving session data into HAR file.

Interact with cookie banners by modifying "referer" header to simulate user navigation of inner pages of a website.

Custom CMPs
- Many websites implement cookie notices using a Consent Management Provider (CMP).
- OneTrust has the largest market share of CMPs.
- By decoding, modifying, and reencoding the OneTrust consent cookie, we can simulate more granular consent options (e.g., only necessary and functionality cookies).

Distribution of CMPs Used by Websites

90% of the 255 sites crawled used the OneTrust CMP.

Related Works


Scan to view GitHub repository

View Code