Stateless? A Comparative Study of GDPR-like Policies Effect of Tracking in Web Applications

Mike Zeng, Iowa City West High School [Iowa City, IA]

Introduction

- As Internet becomes more widespread, ads have became extremely prominent. The advertising industry is valued at $232.7B (2023), expected to hit $500B+ in 2032.
- In particular, online advertising has prominence because of its ability to optimize what ads to serve on the fly to users by real time tracking.

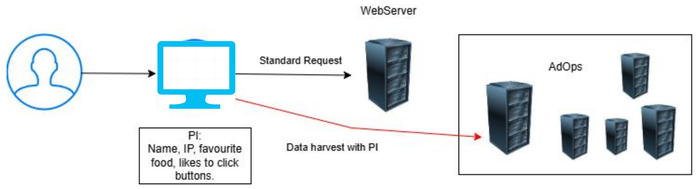


Figure 1: Ad Chain Diagram by Mike Zeng using draw.io

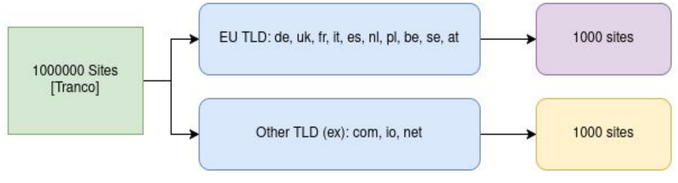
- However, tracking can be **highly invasive**, and as a result, **privacy laws** have been passed to **regulate sharing of personal data**.
- Prior work in GDPR/CCPA policy highlights how **compliance can be measured or how it affects websites**.
- In this study, I take a **different approach of measuring** GDPR effects. I aimed to create novel knowledge regarding tracking behavior of GDPR-enforced VS regular websites.

Research Questions

- RQ1: How do the cookie counts of websites in areas of **GDPR-enforcement** compare to those in **law-free states**?
- RQ2: How does **interacting with the website** affect tracking? What kind of **relationship** does it have?
- RQ3: How does **IP affect tracking**? Are websites using a slow-migration approach or complete migration?
- RQ4: Under GDPR, are cookie preferences **respected better or worse**? How do sites respect preferences based on IP?
- RQ5: Do cookie counts have a **relationship to script counts**? Is it causation or correlation?

Design

Tranco Split.



WebCrawler

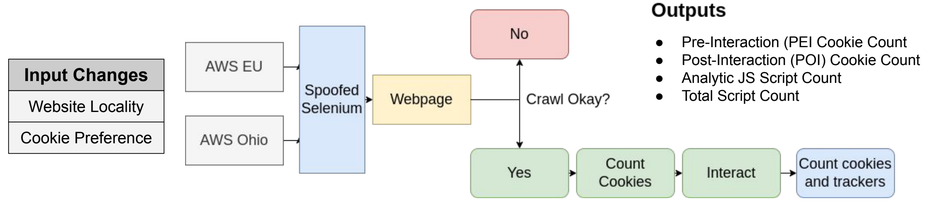


Figure 2: Design Flowcharts by Mike Zeng using draw.io

Results

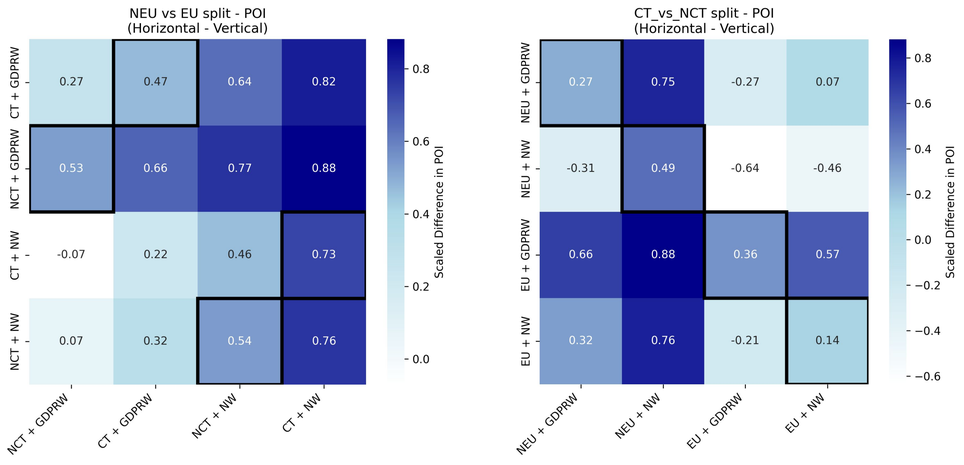
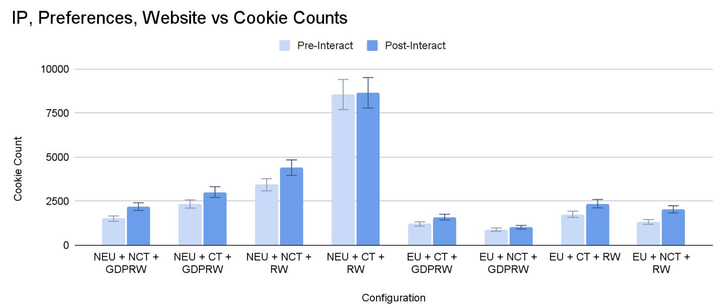


Figure 3.1: Results collected by Mike Zeng. All figures created by matplotlib.

Top Left: Scaled headmap difference between EU and Non-EU IP.

Top Right: Scaled difference between Cookie Tracking and No-Cookie Tracking.

Bottom-Left: Scaled difference between GDPR-enforced and non-GDPR website.



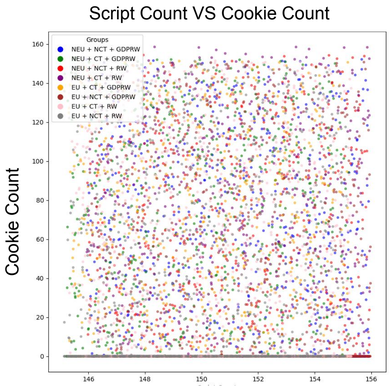


Figure 3.2: Results collected by Mike Zeng, matplotlib and Google sheets.

Top: Pre-Interact vs Post-Interact differences.

Bottom: Scatterplot of Script Count VS Cookie Count across groups.

Findings

- GDPR encourages websites to have lower cookie counts. In other words, it **discourages cookie usage as a tracking tool**.
- Interactions can trigger more tracking. Data suggests websites **still use behavior tracking**.
- **Slow migration** appears to be in place, as websites can change behavior **solely based on IP**.
- Cookies **preference is respected**, regardless of EU locality.
- Analytical scripts **are not being reduced** by GDPR, and data suggests scripts are **using some other technique**.

Future Work

In this study, I observed **key differences** between **websites hosted in the EU** compared to the world. In the future, will all websites in all countries trend to become like EU countries? There is clearly a **shift in the tracking methods** that websites in EU. What will the economic effects of that be for US websites if they do make the shift? There are many areas of this study that can be improved on. What will this study look like in 5 years? 10 years? What other metrics have been changing in websites? Overall, my research is just one piece of a larger context yet to be explored.