



Privacy Pioneer Goes Abroad:

Impact of User Location on Internet Privacy



Dominik Dadak, Harry Yu, Daniel Goldelman, Joe Champeau, Nate Levinson, Judeley Jean-Charles, Hamza Harkous, Sebastian Zimmeck

Abstract

Internet privacy has been a critical issue for users since its inception. When surfing the web, HTTP requests made to third parties have the potential to contain personal information without notifying the user. Our previous paper has shown that packaging machine learning and rule-based analyses into a browser extension is an effective way to convey data practices to users, as opposed to long, confusing, and abstract mentions of what *could* be taken from the user in privacy policies. However, the influence of a user's **physical location** on the types and quantities of collected data remains underexplored. Understanding the impact of **location-dependent privacy laws** will offer insight into any regional discrepancies that need addressing. **Our methodology consists of automating website visits to the top 500 global sites from virtual machines hosted in various target locations with Google Cloud [1], as well as the top 500 from our real location. We then analyze the different privacy practices that occur.** We hope that the results of our analysis will provide lawmakers with insight into areas of internet privacy legislation and regulation that need to be addressed.

What is Privacy Pioneer?

While browsing the internet, it is often unclear what data is being taken from the user. Privacy Pioneer [2] is a browser extension for Firefox developed to help shed light onto privacy violations in real time, utilizing both rule-based and machine learning approaches. Now that we have a tool for analyzing web traffic and identifying when data is being taken by the visited website and integrated third parties, we want to investigate whether the location that a user is visiting a website from has an impact on the data that is taken.

Here is an example of the output Privacy Pioneer would generate when a third party receives your data:

Third Parties

onetrust.com

Analytics

Description

- The URL that initiated this web request is known to practice Analytics.
- This request was found at: 1:43:25 PM on Wed, July 17, 2024

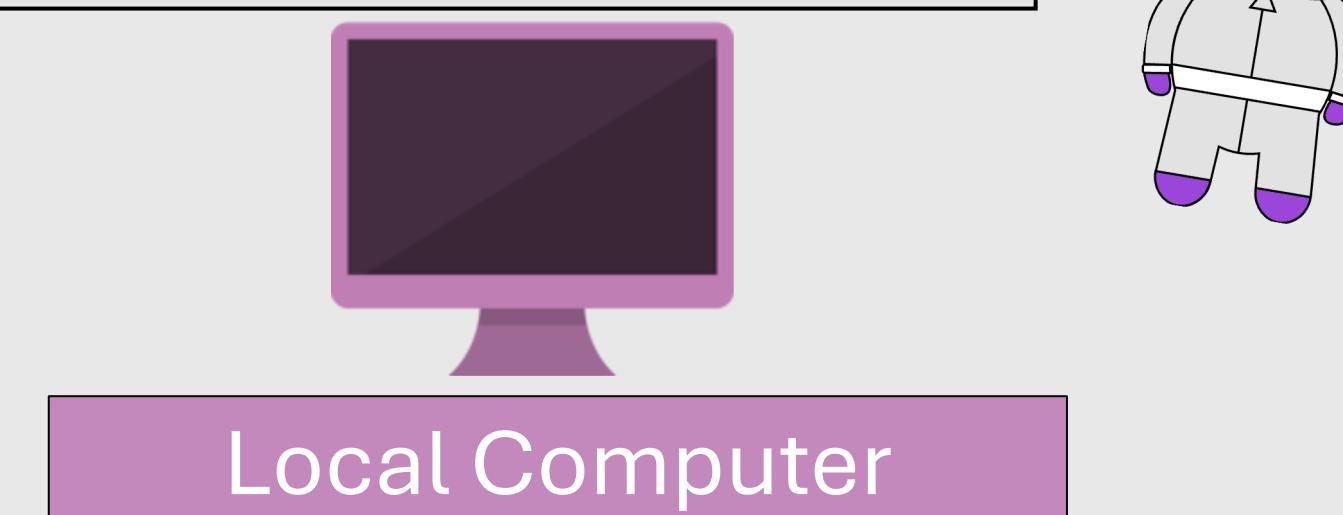
Request URL

<https://geolocation.onetrust.com/cookieconsentpub/v1/geo/location>

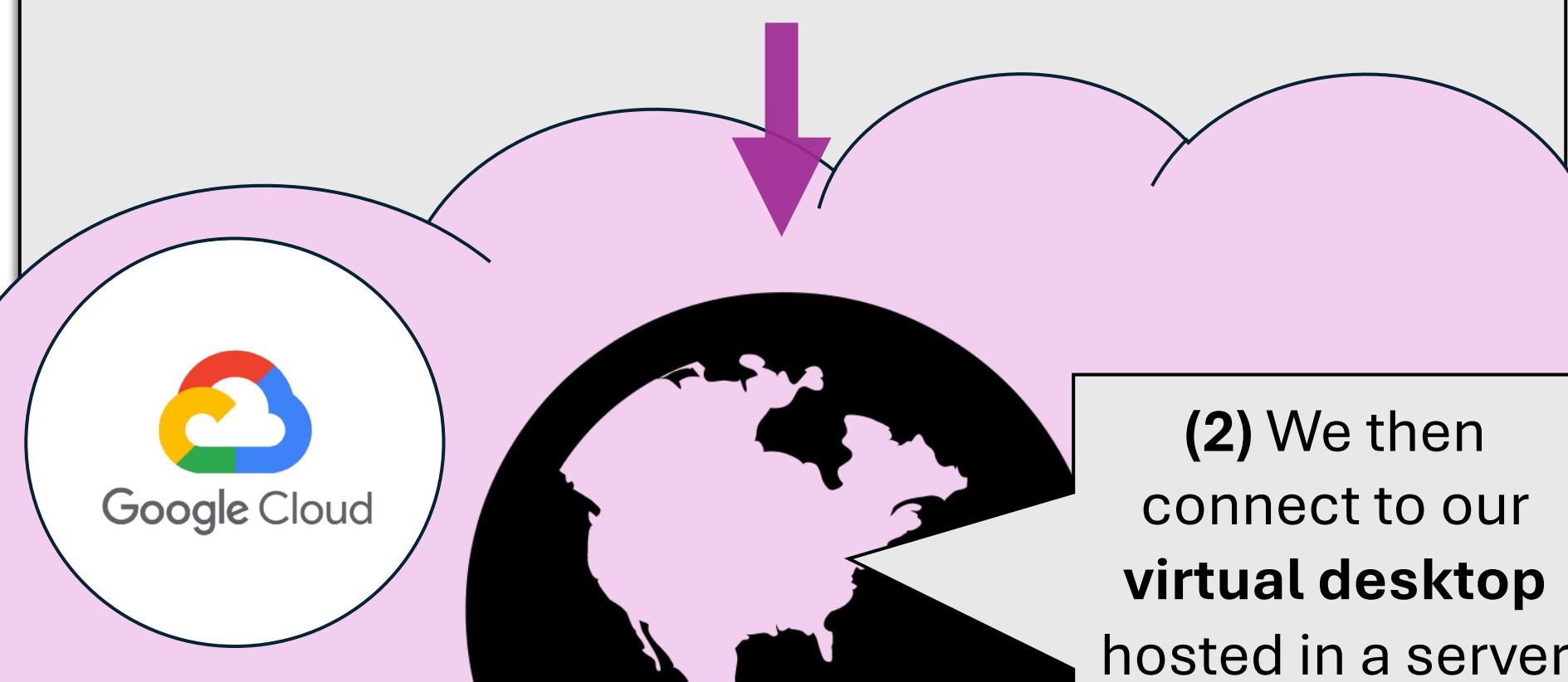
Setup

Since we can't be everywhere at once, how can we collect data from all over the world? We opted to use cloud computing and virtual machines. This method would allow us to host a computer from servers in different physical locations, thereby spoofing our location in a convincing manner. Using Google's Cloud Compute [1], we create identical virtual machines via machine images, with each machine being hosted in its corresponding server location.

(1) First, we create a **web crawler** – a program designed to automatically visit a list of websites. For every site we visit, **Privacy Pioneer** will collect data from those sites.



(used for monitoring the crawlers)



(3) Lastly, for each location, we run the crawler on our manually-curated list of sites and store our findings in a **database** for later analysis.



By comparing the data generated across different locations, we hope to inform lawmakers regarding where they should focus their legislative efforts.

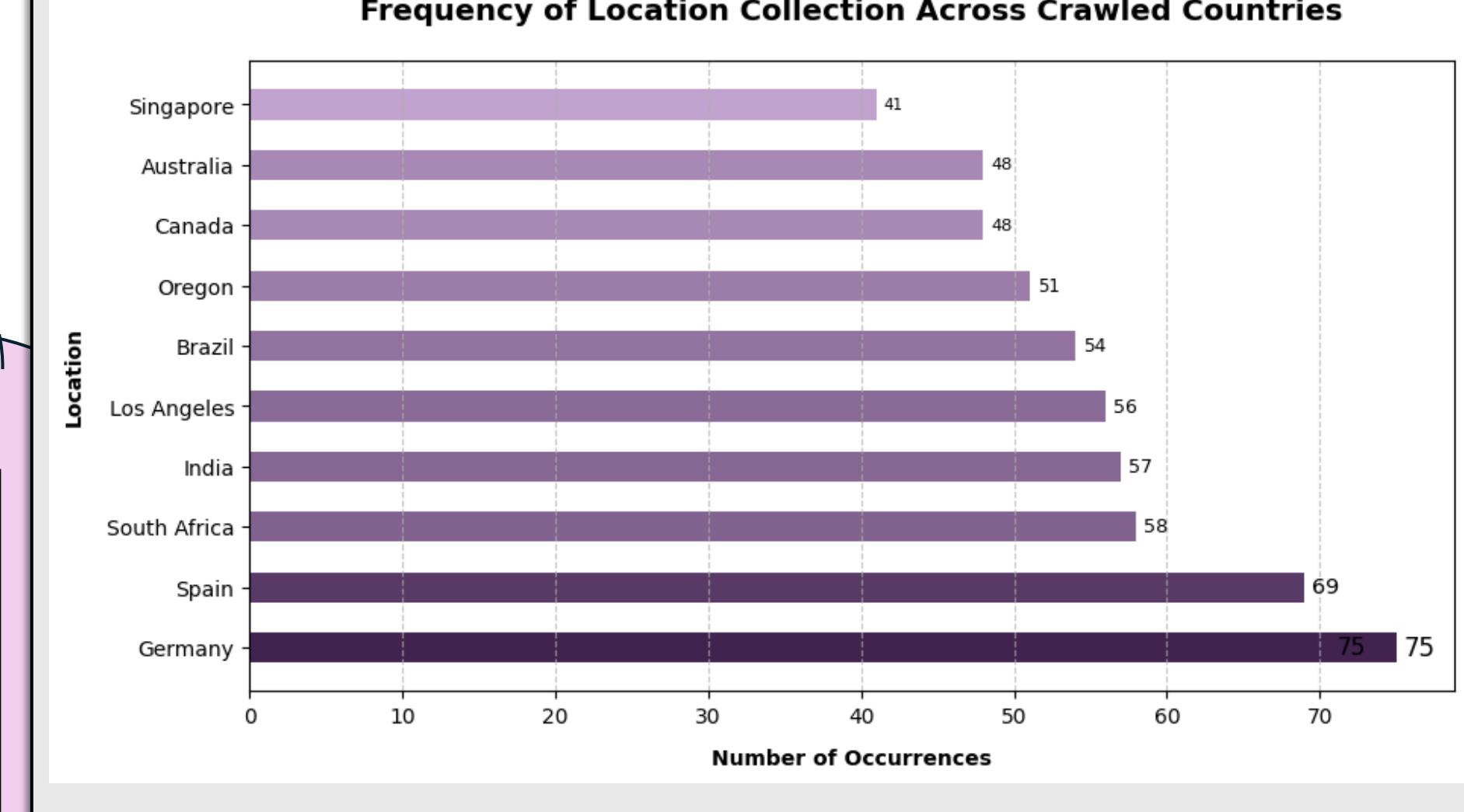
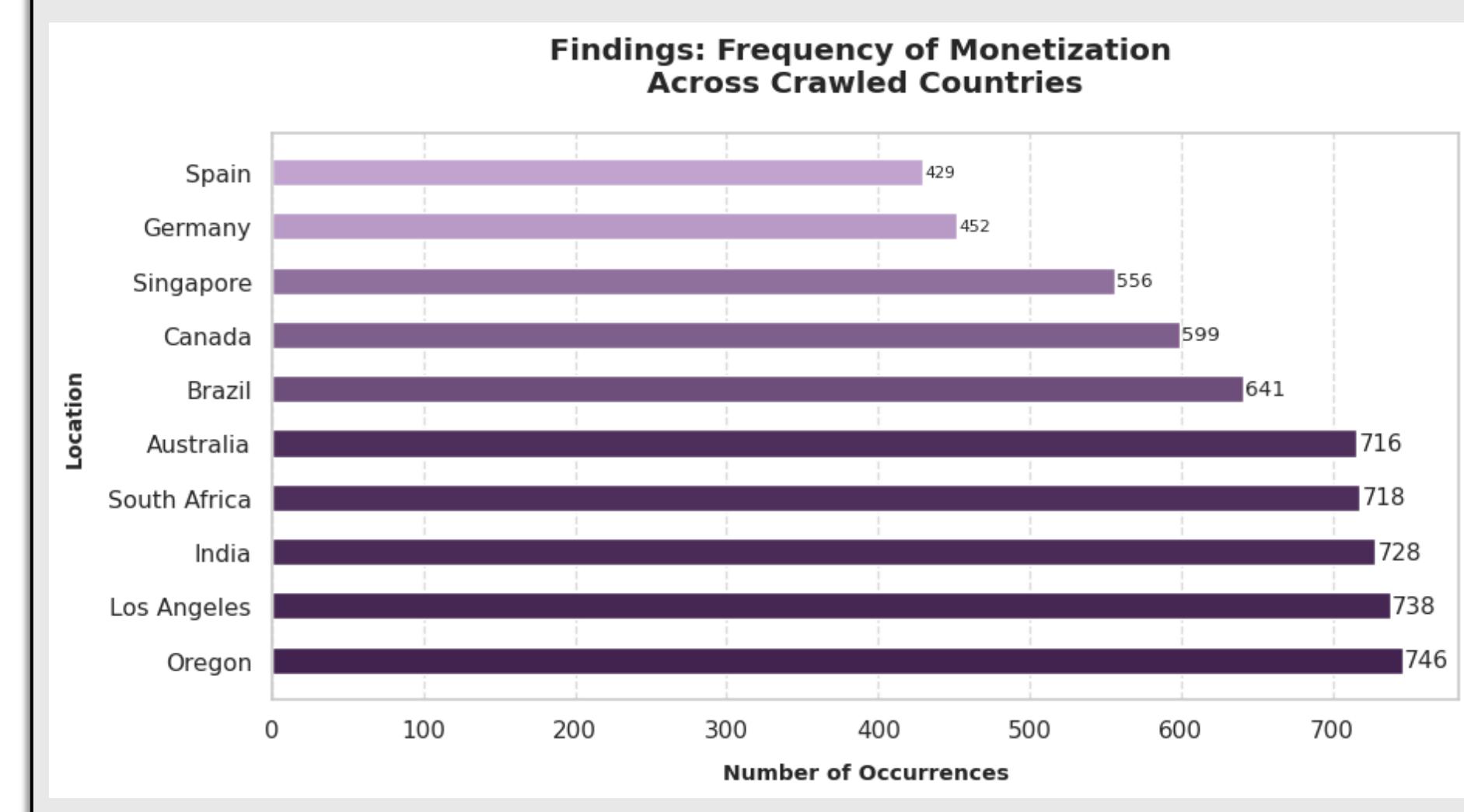
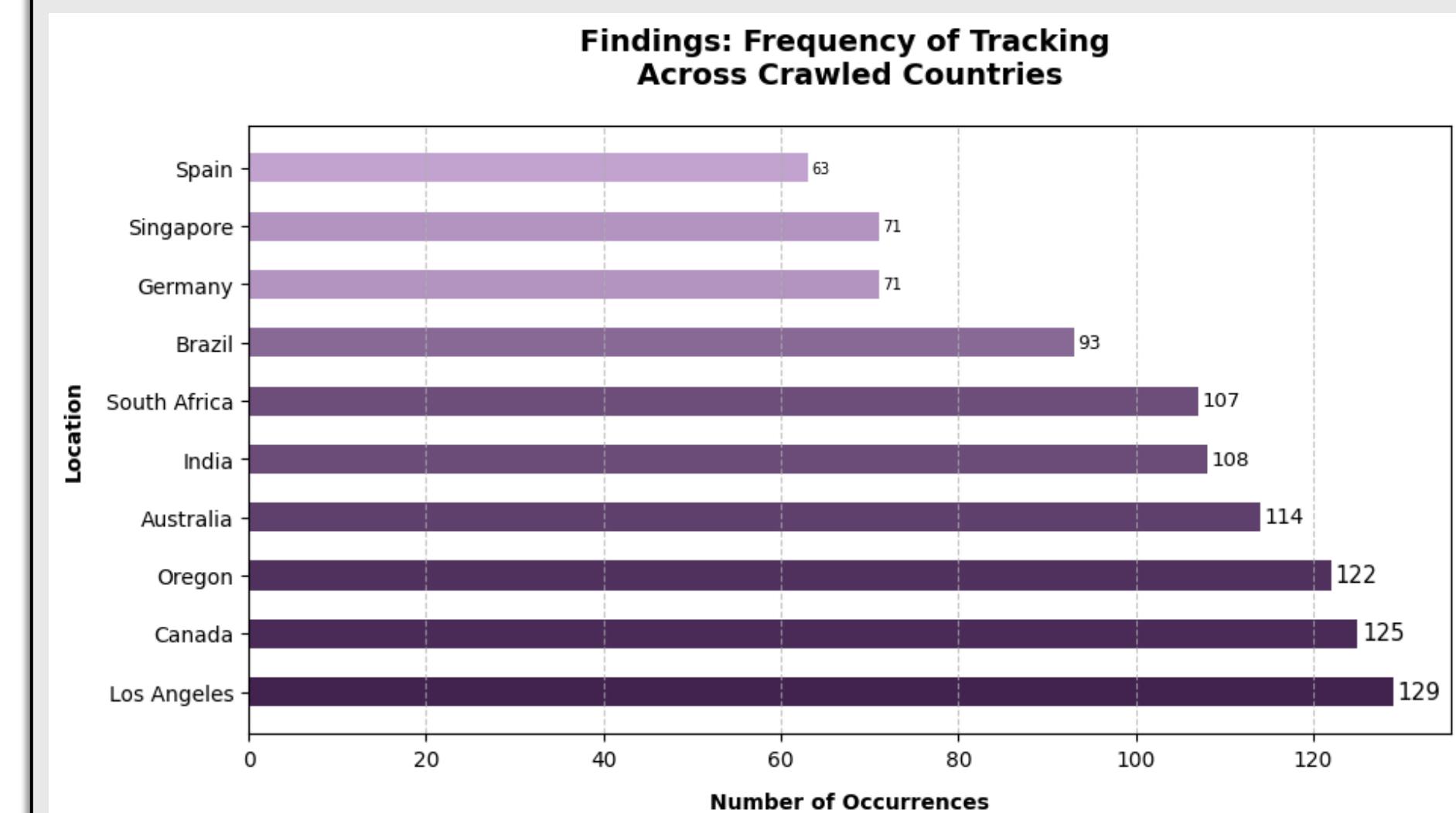
Locations

These are the locations we decided to crawl. We wanted to have a geographically diverse set of countries to get a full spread of different privacy laws.

Australia - Sydney
Brazil - São Paulo
Canada - Toronto
Germany - Berlin
India - Delhi
Singapore - Singapore

South Africa - Johannesburg
South Korea - Seoul
Spain - Madrid
USA - Oregon
USA - Los Angeles

Initial Results



Based on an initial test set of 60 sites per country, we can already see variations between locations. For the tracking and monetization categories (such as advertisers, social networks, etc.), the countries governed by the GDPR (Spain and Germany) have fewer instances overall, while North America takes the lead. However, Spain and Germany seem to be taking location data more often than others.

We hope to perform a full crawl in the coming weeks.

References

- <https://cloud.google.com/about/locations/>
- <https://github.com/privacy-tech-lab/privacy-pioneer>

Acknowledgements

We are grateful to Google for their support of this research via the Research Scholar Program, especially, Nina Taft and the whole Applied Privacy Research group. We also thank Wesleyan University, its Department of Mathematics and Computer Science, and the Anil Fernando Endowment for their additional support. Conclusions reached or positions taken are our own and not necessarily those of our supporters, its trustees, officers, or staff.

