Tracking ecosystem in desktop vs. mobile platforms

Niharika Acharya
nuachary@ncsu.edu

Vaishnavi Doraiswamy
vdorais@ncsu.edu

Abhilasha Saini
asaini4@ncsu.edu

1 INTRODUCTION

The prevalence of trackers across different websites has been studied extensively and is well known. These trackers use the information of the user to monitor the activity of the user across different websites. There are also different tools and browser extensions in place to block such trackers. In the recent years due to the advancements in technology there have been a tremendous increase in the mobile apps for these websites. Almost everyone uses mobile apps for everything from connecting on social media to shopping and booking flight tickets. The behaviour of trackers across the mobile devices has not been extensively studied. This project aims to compare the tracking behaviour across mobile and desktop platforms. This paper explains our data collection methodology and the apps and extensions we have used to measure the trackers across desktops and mobile platforms.

We have also mentioned the dangerous leaks that we found when we analysed the data sent through these apps. We have also performed analysis on the response data and the cookie data extracted using OpenWPM. The data collection methodologies using ICSI Haystack's Lumen application, Ghostery and OpenWPM is mentioned in the proposed approach. The evaluation and finding on the number of trackers, third party requests and common trackers is further explained in the Results section.

2 BACKGROUND AND MOTIVATION

The users of the mobile platform have two mediums-mobile apps and the mobile web. There are many characteristics that only apply to mobile web such as device name, type, carrier network and browser. The user-agent string of the mobile user also contains the device model number of the mobile and hence is more diverse than for the user of the desktop. The users can be tracked using stateless and stateful mechanisms. Stateful tracking involves tracking users by means of cookies and syncing the user profiles across multiple apps and websites. The latter involves creating fingerprints of the user through browser and device fingerprinting such that the fingerprint of the user is unique and the user is identifiable.

Based on the GPS location of the user, the trackers are able to narrow down the users with much accuracy. Desktop users can be tracked on the other hand through their network IP address. Several tracking SDKs are readily available which have their own tracking code. On embedding these in the app, they can track the user and collect various metrics which is then sent to the server. The private information thus collected is used to create a unique fingerprint of the user. These are some of the significant ways in which the user data is exposed on the mobile and desktop ecosystem. This project aims to highlight the similarities and the key differences between the tracking ecosystems in mobile and desktop.

3 RELATED WORK

We have referenced the following related work for this project.

- Tracking the Trackers: Towards Understanding the Mobile Advertising and Tracking Ecosystem

  The interaction between the user installed mobile apps and the third party services is studied

through this paper. This paper conducts a study with the ICSI Haystack Lumen Privacy Monitor and gives insight on different trackers across the android app and identifies different third parties associated with advertising and tracking services.

- The Web's Sixth Sense: A Study of Scripts Accessing Smartphone Sensors

  This paper carries out extensive studies of smartphone sensor by extending OpenWPM privacy measurement tool to OpenWPM mobile. After studying the usage of sensor API on top 100k sites, this study also detected fingerprinting on mobile platform which was previously carried on desktop ecosystems.

- Online Tracking: A 1-million-site Measurement and Analysis

  Using OpenWPM web based measurement framework, this paper has analysed 1-million sites for different types of tracking prevalent in the desktop ecosystem. This study focuses on 15 metrics on each site including stateful and stateless tracking and measures the effectiveness of browser privacy tools.

4 PROPOSED APPROACH

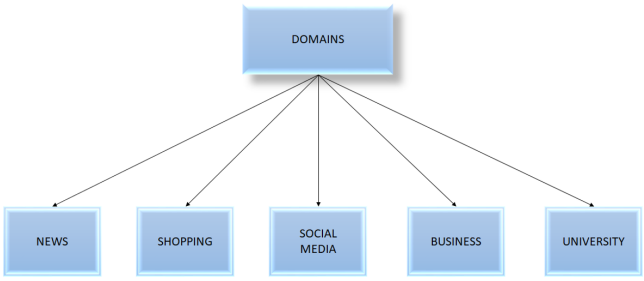

Figure 1: Domain Classification

We conducted our study by tracking websites from the following domains:

- News
- Shopping
- Business
- University
- Social Media

The main idea behind choosing this domain was based on the popularity of browsing and the security measures implemented. As news, social media, business, shopping are the most visited domains they are found to be more vulnerable to tracking compared to universities.

To analyse the tracking patterns on desktop browsing and mobile browsing we took the following approach:

- Platform used to carry out the study:

  Mobile and Desktop browsing was studied using the Ghostery Platform. To analyse the http request and response data from the Android apps and study the tracking behaviour, we used ICSI Haystack's Lumen Privacy Monitor. To study the request and response metadata and cookie data we used OpenWPM which is an open source web privacy measurement framework.

- Data collection and analysis

  For analysis on the number of trackers we installed Ghostery browser on mobile and Ghostery extension on the Chrome browser on desktop. We visited 30 websites manually from Ghostery browser on mobile and Chrome browser. We collected information regarding the number of trackers, unique trackers to desktop and mobile. We wanted to understand how the trackers in apps are different from those in browser and desktops. Also we wanted to find if there any common trackers in Android apps and desktops. We installed ICSI Haystack Lumen Privacy Monitor on our Android phones. We also installed TLS certificate on our phones so that Lumen can intercept TLS traffic. We visited the apps through Lumen and analysed the data it collected regarding third party requests, trackers and leak of sensitive information if any. We used OpenWPM framework to collect data regarding request data, response data and cookie data for 150 websites with 30 websites in each of the 5 domains mentioned above. OpenWPM

crawls the websites in an automated way using Selenium browser automation and stores the results in crawl-data.sqlte database. We analysed following data from crawl-data.sqlite:

1. site_visits
2. http_requests
3. https_responses
4. javascript_cookies

We developed python scripts to extract the data from the database crawl-data.sqlite. To find the number of third party requests and blocked domains, we used adblockparser python library and easylist. We used adblockparser with easylist rules against the request and response data in crawl-data.sqlite to get the blocked requests per visit of a site.

- Result evaluation
  Our first step was to get a count of the websites that are marked as trackers on desktop and mobile. Followed by that, we tried to identify the tracked websites unique to each and the ones that were common to both. Another thing we analysed was the tracking patterns on browsing the mobile applications of the websites. We used Lumen application by downloading it on mobile device and by manually downloading and visiting the 30 applications on it. Also, using Open WPM we analysed the third party requests and the average number of blocked sites per domains and average no. of third party cookies set per visit of the websites in each domain. We then plotted the network diagram to understand the common trackers found by Lumen and OpenWPM.

5 EVALUATIONS AND RESULTS

5.1 Ghostery: Browser(Mobile Web) and Extension(Desktop)

An analysis on the following was done by browsing the websites on Mobile vs Desktop.

- Number of trackers trying to track while visiting the website.
- Number of trackers that were being blocked

Ghostery Categorises the trackers into the following domains :

- Advertising
- Site Analytics
- Social Media
- Essential Trackers
- Customer interactions

It was observed that Advertising, Social Media and Site analytics are the most common tracking categories observed. An identification of the trackers blacklist unique to Mobile and Desktop was also studied.

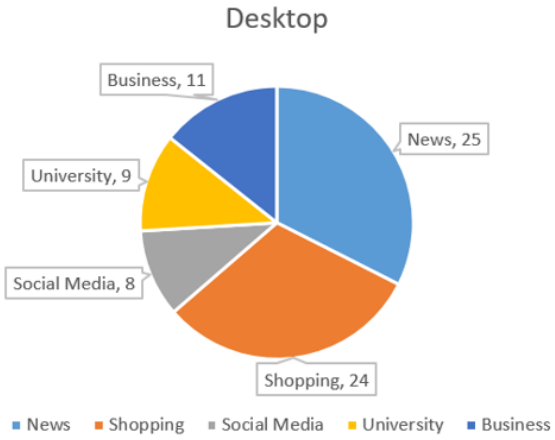

Figure 2: Sites blocked by Desktop

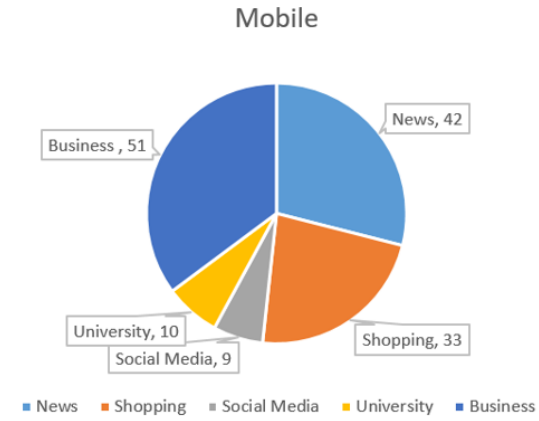

Figure 3: Sites blocked by Mobile

From the figure above we observe that, the number of trackers that mobile blocks is more compared to the desktop blocking. Websites that fall in the blacklist of both mobile and desktop were seen to be blocked by mobile but not desktop. Also, mobile's tracking and blocking list in every domain was observed to be more compared to the ones on desktop.

Few of popular websites and categories that are blocked by mobile usually and are just tracked by desktop but not essentially blocked are as follows:

- Advertising Trackers: facebook custom audience, demandbase, facebook pixel, beeswax
- Site Analytics : Score Card , Google analytics, Research Beacon
- Social Media : LinkedIn Widget , Facebook social plugin

5.2 Understanding the mobile tracking through Lumen Privacy Monitor

Lumen intercepts the traffic that the app sends to the server and monitors the requests to find the tracking domain, common third party and tracking domains across apps and also the percentage of traffic wasted on advertisements.

Using Lumen we did a comparative study across 30 apps across the 5 domains mentioned above to find the following information:

- No. of trackers with respect to domain type.
- Most common trackers across apps.
- Dangerous leaks in apps.

5.2.1 Understanding the number of trackers with respect to domains.

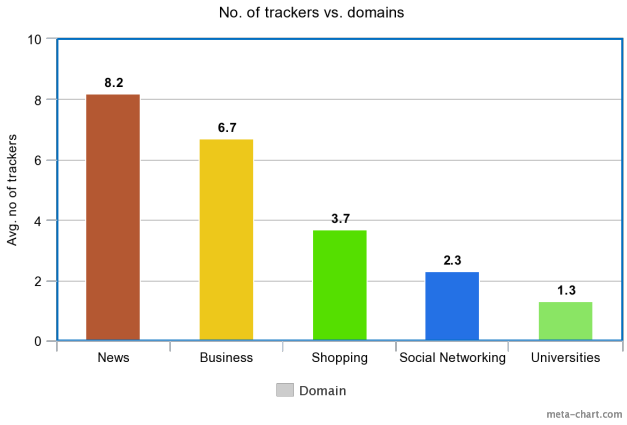

Figure 4: Average number of trackers vs. domains

From the above plot, we can see that news domain has the highest number of trackers that are present when one visits the websites. There are about 8 trackers when one visits any app in the news category. The number of trackers decreases for business, shopping and social networking. It is the least for the apps in the Universities domain.

5.2.2 Understanding the common trackers across apps.

Through Lumen we analysed the third party and tracker data to understand the number of common trackers across all the apps. Understanding the common trackers will help gauge how the same user's profile is matched across multiple apps. In the below figure we can see the top common trackers across multiple apps.

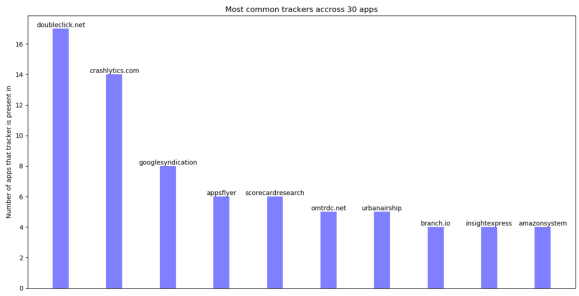

Figure 5: Most common trackers in 30 apps

The above plot shows the most common trackers found vs. the no. of apps that the tracker is present in.

We can see from the above plot that doubleclick.net is the most common tracker and it is present in 17 of the 30 apps. Similarly, crashlytics.com is present in 14 of the 30 apps.

Looking at this plot we get an idea of how one tracker that is present on multiple apps can match different profiles of the same user and create a unique user profile across all the apps.

5.2.3 *Understanding the Dangerous Leaks in Apps.*

As mentioned previously, the leak of private information such as private IP or browser and hardware properties are used to create a unique fingerprint of the user. We wanted to understand what are some of the the dangerous leaks of private information through apps.

Information Leaked	App name
Private IP	Pandora
Timezone	CNBC
Brand	CNBC, NYTimes
Device Model	BBC News, CNBC, eBay, Facebook, Indeed, LinkedIn, NYTimes, Pandora, USA TODAY, Washington Post, WSJ
Operator Name	CNBC, WashingtonPost
Radio version	CNBC
Build Fingerprint	BBC News, CNBC, Facebook, Indeed, LinkedIn, NYTimes, Pandora, USA TODAY, Washington Post, WSJ

Figure 6: Leak of private information through Apps

The above figure gives the different private information leaked through apps.
Device Model and Build Fingerprint of the device was leaked in many of the apps such as BBC News, CNBC, Indeed etc. Private IP was leaked through Pandora. Some apps such as CNBC and WashingtonPost also leak Operator Name.

5.3 Evaluations using OpenWPM

Using OpenWPM we collected the request, response and cookie metadata for 30 websites in each domain i.e 150 total sites. The analysis using OpenWPM is divided into 3 categories.

- Third party requests for each domain.
- Blocked sites analysis for domains.
- Third party cookie analysis for domains.

5.3.1 *Third party request for each domain.*

Using OpenWPM we analysed the http request metadata to find the number of third party requests for a site. We also analysed the trend in third party requests according to the different domains.
The below plot shows the number of third party requests with respect to the domains.

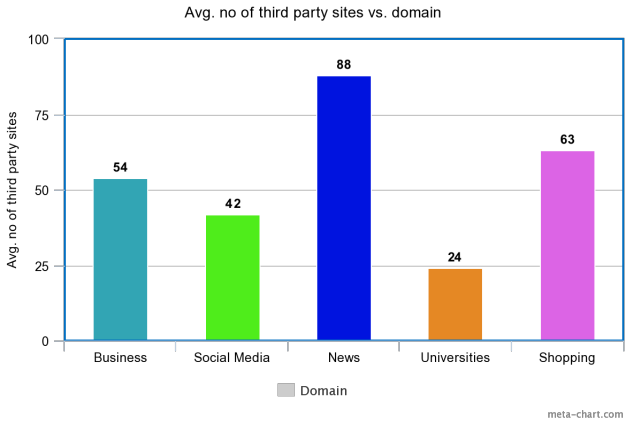

Figure 7: Third party requests vs. domains

The plot displays the average third party requests for each domain. As we can see in the above figure, the average third party requests made by any website in the news domain is 88. It goes on decreasing for shopping and business domains. The average third party requests are the lowest at 24 for Universities domains.

5.3.2 *Blocked sites analysis for domains.*

Using OpenWPM, we analysed the request and the response metadata to analyse the blocked domains. We used adblockparser and easylist to analyse which of the requests will be blocked. The below graph shows the average number of requests that will be blocked when we visit the site than belongs to that domain.

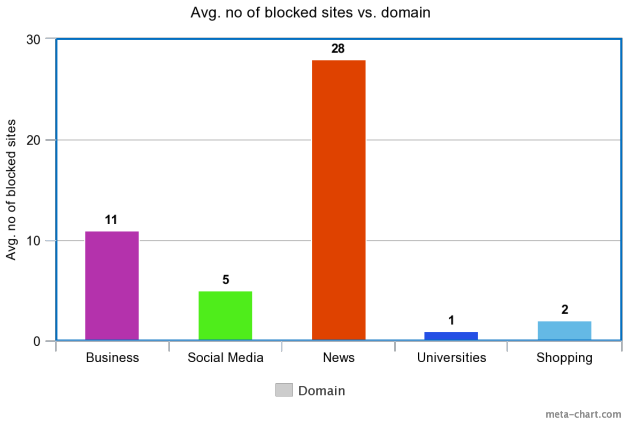

Figure 8: No. of blocked sites vs. domains

From the above plot we can see that the number of average blocked domains when we visit any website in the news domain is as high as 28. This number goes on decreasing for Business and Social Media and is least for Universities. The average number of blocked domain on University websites is the minimum at 1.

5.4 Third party cookies set vs domain

We analysed the cookie metadata collected by OpenWPM to understand the third party cookie behaviour.

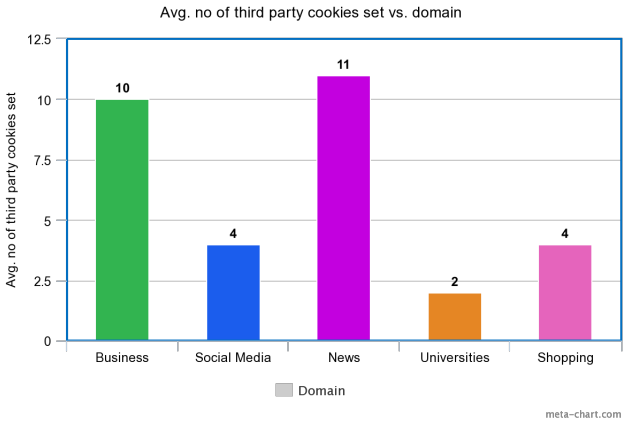

Figure 9: Avg. no. of third party cookies set per domain

The above figure explains the average number of third party cookies set on each site visited with respect to domain. As we can see the sites in the news domain have a maximum of average third party cookies set at 11. Shopping, Social Media have fewer third party cookies set and it is the lowest for University domain. On analysing the expiry date of the cookies we found that many of these cookies were set to expire far in the future or indefinitely long. That is the users data is stored for long periods of time and is used to track the user over time.

5.5 Network diagram to understand the common trackers

To understand the common trackers from the OpenWPM crawl we plotted the network diagram of the common trackers.

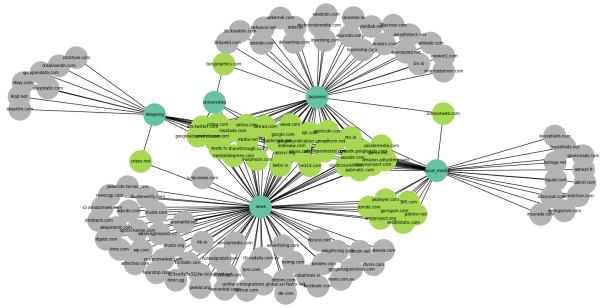

Figure 10: Network diagram of the trackers in OpenWPM on 150 sites

In the above network diagram, the dark green nodes represent the five domains and all the other nodes represent the trackers found on the 150 sites. The nodes colored in light green represent the common trackers across all the domains.

To understand the common trackers found between 30 sites analysed using Lumen on android apps and the websites of the same crawled using OpenWPM, we constructed a similar network diagram.

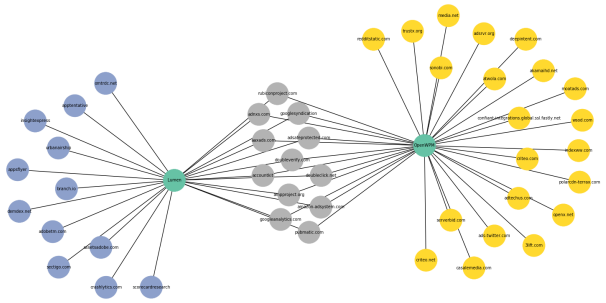

Figure 11: Common trackers in mobile app and desktop across 30 websites

The above figure showcases the common trackers across Lumen and OpenWPM crawl on mobile apps and websites. The blue nodes in the diagram represent the trackers unique to the Android apps and the yellow nodes represent the trackers unique to websites. The grey nodes represent the common trackers across the apps analysed through Lumen and websites analysed through OpenWPM.

6 LIMITATIONS AND FUTURE WORK

The limitation that we faced in our study was that Lumen does not use the same blacklist as does Ghostery. Hence we could not do a comparative study between Android apps through Lumen and websites through Ghostery. Lumen only works for Android ecosystem and there is a lack of apps for iOS Operating system that would monitor the trackers and block them. Hence a comparative study between different operating systems was also not feasible.In the future we would like to increase the total number of websites that we crawled. Also automating the manual process of using Lumen on each Android app should also be automated. We would also like to do a comparative study on more parameters to better understand the differences in the tracking ecosystem.

7 CONCLUSIONS AND DISCUSSIONS

We analysed the data sent to the server from the Android apps using ICSI Haystack's Lumen Privacy Monitor for Android. Through this we found that many of the Android apps have many trackers common among them. These trackers are used to match user profiles across multiple apps and create a unique profile for a user. We also saw that many of these apps leak sensitive private information such as device model and build fingerprint and even private IP in some cases that can be used to create a unique user profile by means of fingerprinting. On comparing the third party sites and the sites blocked through Ghostery we find that the number of trackers on mobile is more than that on desktops. Many of the trackers that are found on mobile are not found on desktop. We also carried on an analysis on the request, response and cookie metadat using OpenWPM and comparing it against easylist using adblockparser. We found that the average number of trackers is far more in news domains and goes on decreasing for business and universities and it is the lowest for University domains. Also news domains have the most number of third party cookies set with many of them with expiry date indefinite. We also plotted network diagrams to understand the common trackers in mobile apps and desktop browsers.

REFERENCES

[1] Steven Englehardt, Chris Eubank, Peter Zimmerman, Dillon Reisman, Arvind Narayanan *OpenWPM: An automated platform for web privacy measurement.* March 15, 2015

[2] Anupam Das,Gunes Acar, Nikita Borisov, Amogh Pradeep *The Web's Sixth Sense: A Study of Scripts Accessing Smartphone Sensors.*

[3] Reuben Binns,Ulrik Lyngs,Max Van Kleek,Jun Zhao,Timothy Libert,Nigel Shadbolt *Third Party Tracking in the Mobile Ecosystem.* March 15, 2015

[4] Steven Englehardt, Arvind Narayanan *"Online Tracking: A 1-million-site Measurement and Analysis"* 2016

[5] Bjoern Greif *"Cookies and fingerprinting: tracking methods clearly explained" https://www.ghostery.com/blog/ghostery-news/cookies-fingerprinting-co-tracking-methods-clearly-explained/.* March 6, 2018

[6] Sam Macbeth *"Tracking the Trackers: Analysing the global tracking landscape with GhostRank"* July 2017

[7] Apps,trackers,privacy,and regulators: a global study of the mobile tracking ecosystem

https://blog.acolyer.org/2018/03/05/apps-trackers-privacy-and-regulators-a-global-study-of-the-mobile-tracking-ecosystem/
[8] Mobile Technology Tracking Methods other than cookies *https://www.allaboutcookies.org/mobile/mobile-tracking.html*
[9] Mobile Tracking: How it works and why it's different *https://www.trustarc.com/developer/?p=86*
[10] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, Claudia Diaz *The Web Never Forgets: Persistent Tracking Mechanisms in the Wild*

8 INDIVIDUAL CONTRIBUTIONS

Task	Estimated Date	Team member
Research on the platforms to carry out the study	10/01	Abhilasha, Vaishnavi, Niharika
Installing Lumen, OpenWPM and setting up the environment	10/15	Abhilasha, Niharika
Running OpenWPM scripts on first 30 websites	11/1	Niharika
Primary analysis using Ghostery browser and extension	11/1	Vaishnavi
Primary analysis using by installing Lumen app	11/1	Abhilasha
Detailed analysis on rest of the websites	11/15	Abhilasha, Vaishnavi , Niharika
Python scripts to analyse the collected data	12/1	Abhilasha, Vaishnavi
Drawing conclusions and completion of evaluation	12/3	Niharika
Completing the first draft of report and presentation	12/05	Abhilasha, Niharika, Vaishnavi
Final report completion	12/09	Abhilasha, Niharika, Vaishnavi