# Bibliographic study

## *The analyses of the traffic generated by terminals*

**Members of the group :**

Clément LE GRUIEC

Nathan OLBORSKI

Hugo HOUILLON

Salma CHAHMI

**Tutors :**

Julien SAINT-MARTIN

Xavier LAGRANGE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

# Abstract

Nowadays, the use of smartphones is widespread. With the diversity and ease of installing mobile applications, people are able to benefit from subscription services, video calls, messaging, social networking apps and many other services. However, these mobile applications interfere with the network traffic, can access mobile data and are likely to violate users' privacy.

In this context, our project consists of developing a platform that visualises the traffic generated by mobile applications with as much transparency as possible.

This document is a state of art to identify and present the studies conducted on the traffic generated by the applications in the mobile ecosystem.

# Summary

# I. Introduction

Our smartphones are full of applications that provide various services. According to Cisco Annual Internet Report (2018-2023 ) white paper, over 70 percent of the global population will have mobile connectivity by 2023. The total number of global mobile subscribers will grow to 5.7 billion (71 percent of population) by 2023. It is also predicted that nearly 300 million mobile applications will be downloaded by 2023. Globally, 299.1 billion mobile applications will be downloaded by 2023. Social media, gaming and business applications will be the most popular downloads. These applications are likely to use the data of terminals, and make exchanges with the servers within encrypted packets.

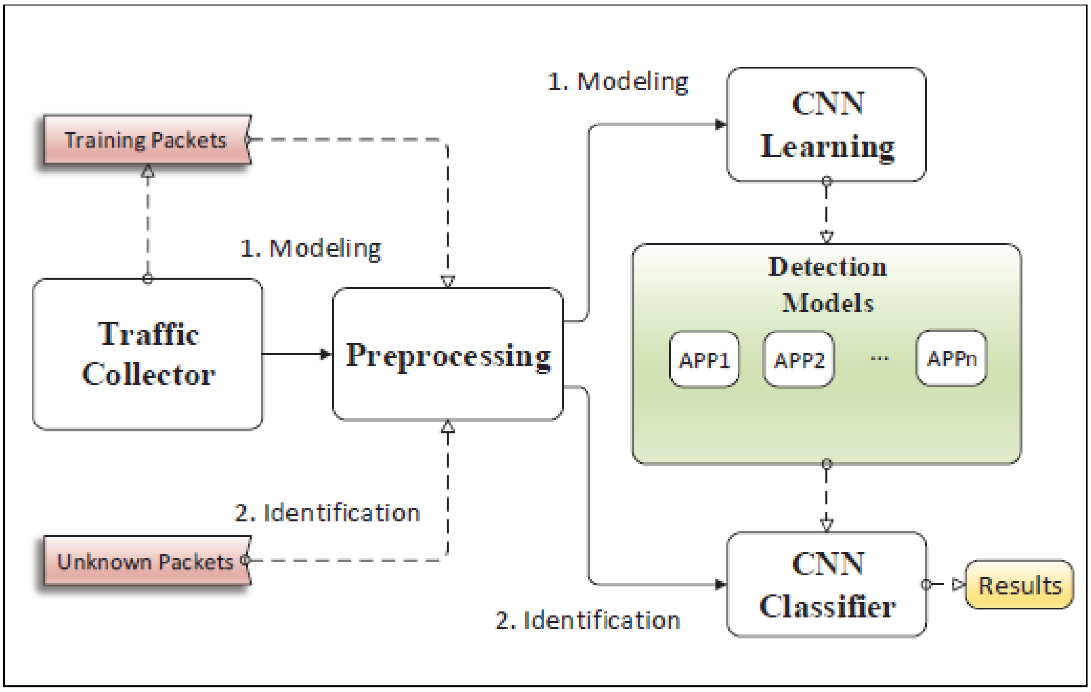
In this document we will focus on the traffic generated by these mobile applications over wireless and cellular networks based on previous studies.

# II. Characterizing the traffic generated by mobile applications

## A. Mobile application traffic identification

The number of mobile applications grows rapidly, and with this explosive growth comes many problematics concerning network management, privacy and other issues. Therefore, It seems necessary to identify the traffic generated by mobile applications, but it remains challenging. In fact, most of the applications use encryption for communication. Also, many of them use cloud services, which means that the name of the host can't uniquely identify the traffic of mobile applications.

One method to identify the traffic of mobile applications is to use deep learning employing Convolutional Neural Networks (CNN). The figure below visualizes the architecture of the approach :



This approach can be divided into two phases :

- Modeling phase that consists of training CNN detection models for application by extracting HTTP header fields in order to generate abstract signatures for each app.
- Identification of unknown packets based on the CNN detection models.

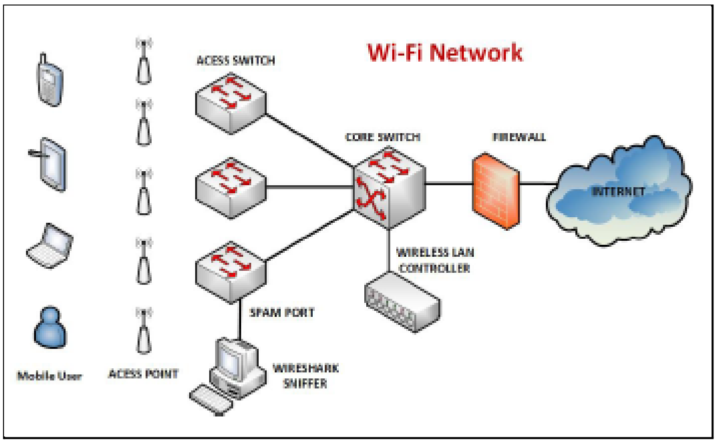
## B. Modeling mobile application traffic

Since the packets are exchanged encrypted, the analysis is based on two principal parameters, the packet size and the inter-arrival time that both have stochastic behavior (involves some randomness and uncertainty).

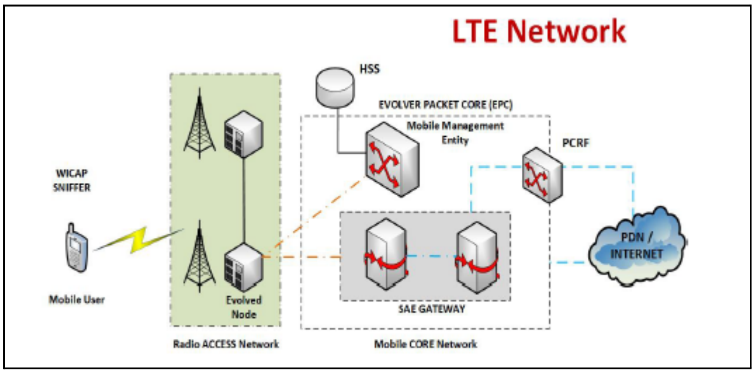
- Experience :

A study conducted by RISTI shows data collection classified by application according to packet size. It implements two scenarios :

- We use Wireshark to sniff packets over WiFi networks



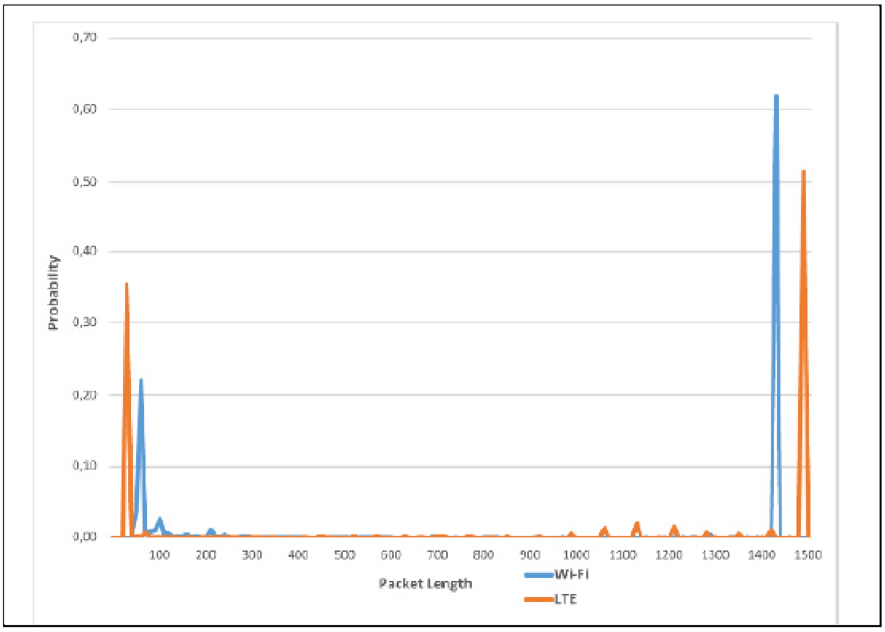
- We use of WICAP for an Android smartphone working over LTE cellular



To analyze the captured packets, we calculate the packet size probability considering intervals of 10 bytes for discrimination .

- Results :

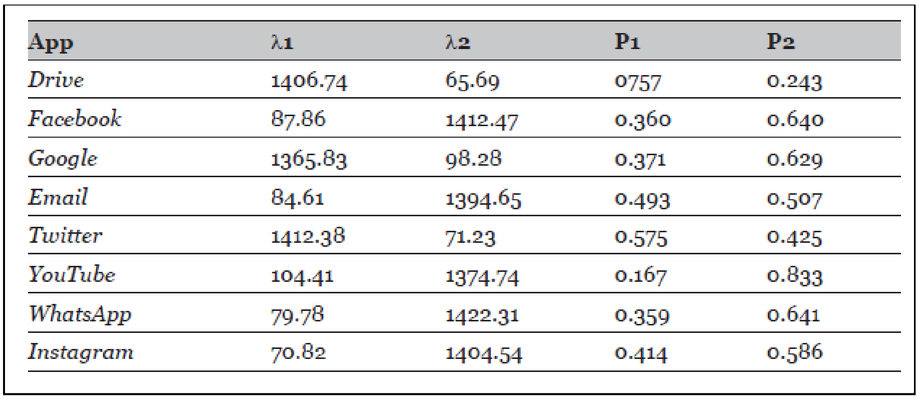
The figure below shows the packet size probability for Facebook. Similar results were obtained for other applications (GoogleDrive, WhatsApp, Youtube, etc.)



➔ The packet size is usually between 0 and 1500 bytes (Linked to MTU).
➔ Over WiFi networks : approximately 22% of packets with 60 bytes size, and 62% around 1430 bytes.
➔ Over LTE networks : 36% around 30 bytes, and 51% around 1490 bytes.
➔ There is a bimodal traffic distribution

- Conclusions :

We can estimate the real traffic by modelling the network traffic using a mixture Poisson distribution. We can represent some application packets patterns over WiFi, for example, with a mixture of two Poisson distributions with the bellow parameters :



| App | λ1 | λ2 | P1 | P2 |
|---|---|---|---|---|
| *Drive* | 1406.74 | 65.69 | 0757 | 0.243 |
| *Facebook* | 87.86 | 1412.47 | 0.360 | 0.640 |
| *Google* | 1365.83 | 98.28 | 0.371 | 0.629 |
| *Email* | 84.61 | 1394.65 | 0.493 | 0.507 |
| *Twitter* | 1412.38 | 71.23 | 0.575 | 0.425 |
| *YouTube* | 104.41 | 1374.74 | 0.167 | 0.833 |
| *WhatsApp* | 79.78 | 1422.31 | 0.359 | 0.641 |
| *Instagram* | 70.82 | 1404.54 | 0.414 | 0.586 |

# III. Characterizing background traffic activities

## A. Definition of background traffic

A sizable amount of the traffic is generated when the user is not actively interacting with the mobile phone. This can happen when the screen is off, or when the other applications are running in the foreground. It is defined as the background traffic, and it is ranging from one-third to two-fifths of traffic across wireless connections.
Background traffic can be subdivided into two categories :

- Streaming background traffic (Pandora, Spotify,...)
- Data-driven background traffic (Facebook, Gmail,...)

## B. Utility of background traffic

Many applications run in the background when the user is not actively interacting with them, and that is for multiple reasons :

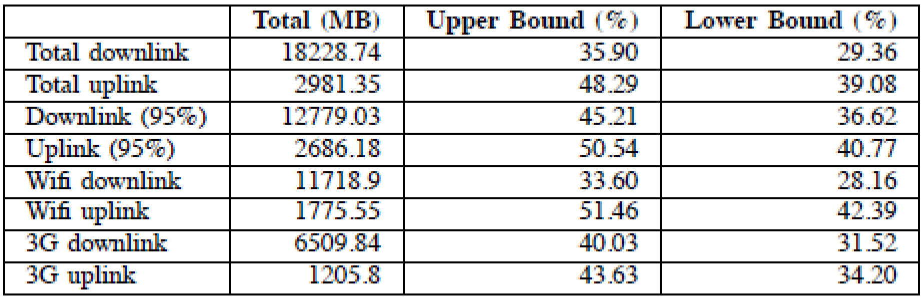
- To make a refresh or updates ( Updates for news or weather )
- To synchronize with cloud service, and get notifications ( Facebook or Gmail notifications )
- To support non-touch based interactions ( Listening to Spotify )

These background activities provide applications with connectivity and fast response time which allow a better user experience.

## C. Analysis of background traffic

A NetSense smartphone study conducted at the University of Notre Dame provides a dataset gathered from students (N=110) who were actively using their smartphone (a base floor of 2Mb/day on average) across a seven week period . To simplify, we consider the background traffic generated when the screen is off.
The table below shows the average weekly traffic statistics for downlink and uplink over 3G and WiFi networks.

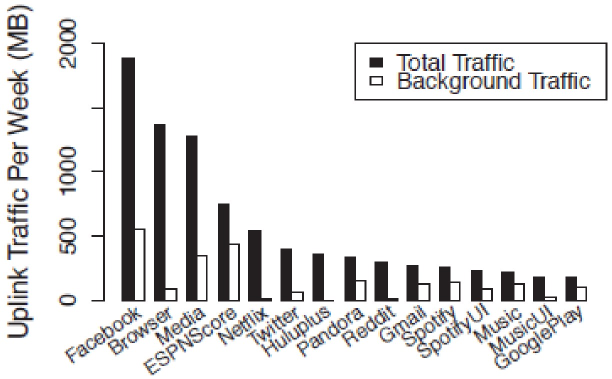
We note that the upper bound corresponds to treating edge cases as background traffic, while the lower bound represents it as  foreground traffic.

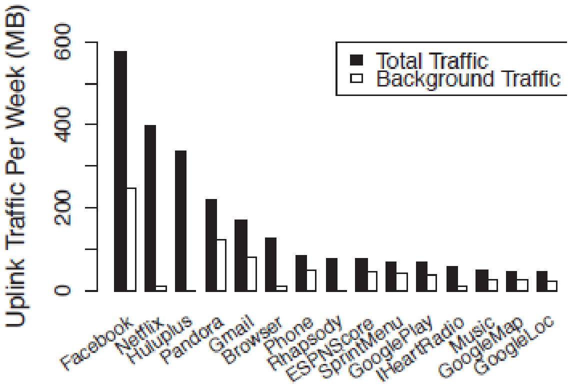


| | Total (MB) | Upper Bound (%) | Lower Bound (%) |
|---|---|---|---|
| Total downlink | 18228.74 | 35.90 | 29.36 |
| Total uplink | 2981.35 | 48.29 | 39.08 |
| Downlink (95%) | 12779.03 | 45.21 | 36.62 |
| Uplink (95%) | 2686.18 | 50.54 | 40.77 |
| Wifi downlink | 11718.9 | 33.60 | 28.16 |
| Wifi uplink | 1775.55 | 51.46 | 42.39 |
| 3G downlink | 6509.84 | 40.03 | 31.52 |
| 3G uplink | 1205.8 | 43.63 | 34.20 |

As shown in the table, the background represents nearly one-third for the downlink, and two-fifths for the uplink which is a non-negligible portion of traffic.
We also notice that WiFi has less downlink traffic, but more uplink traffic in comparison to 3G.

Other statistics represent the traffic generated by top 15 applications :



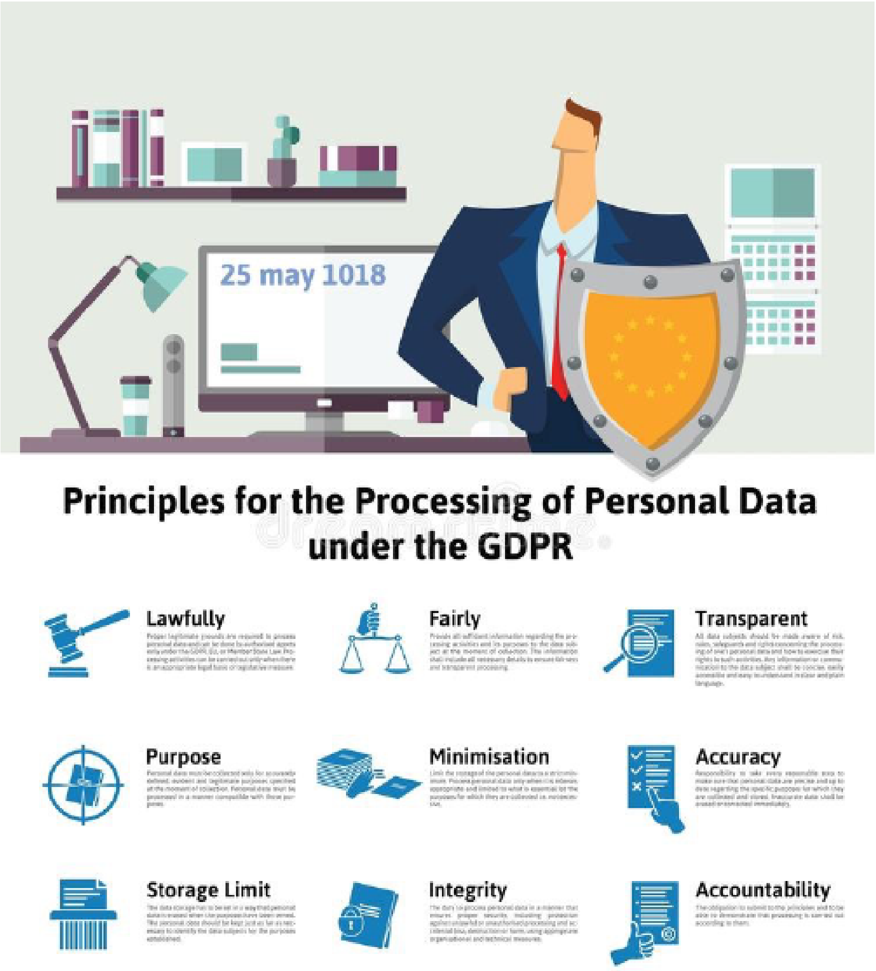


The amount of background traffic differs from one application to another. For example, Facebook exhibits a high level of background traffic while Netflix is a foreground-driven application.

# IV. Privacy within mobile applications

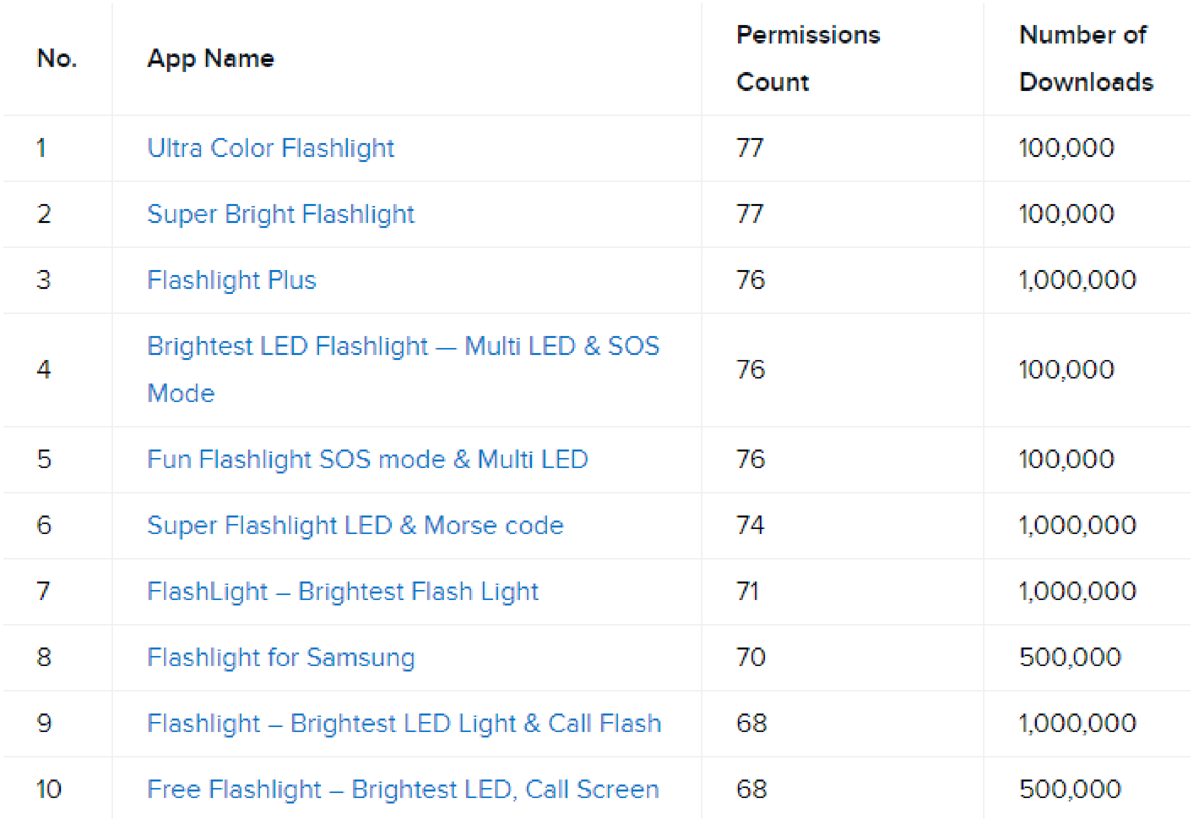
## A. Concept of privacy

Smartphones are full of applications that provide users with different services.These applications require some data from terminals for their good functioning; however they can abuse and exploit much more data than needed, such as phone location, contacts, camera, WiFi network lists and other data. With the abuse of data by some mobile applications, or possible leakage of information, privacy is increasingly at risk.
We can define privacy as the ability of users to choose when, how, and to what extent their personal information can be accessed. According to RGPD, there are some principles that need to be respected concerning the collection and usage of personal data. We can summarize that in the figure below :



## B. Privacy issues

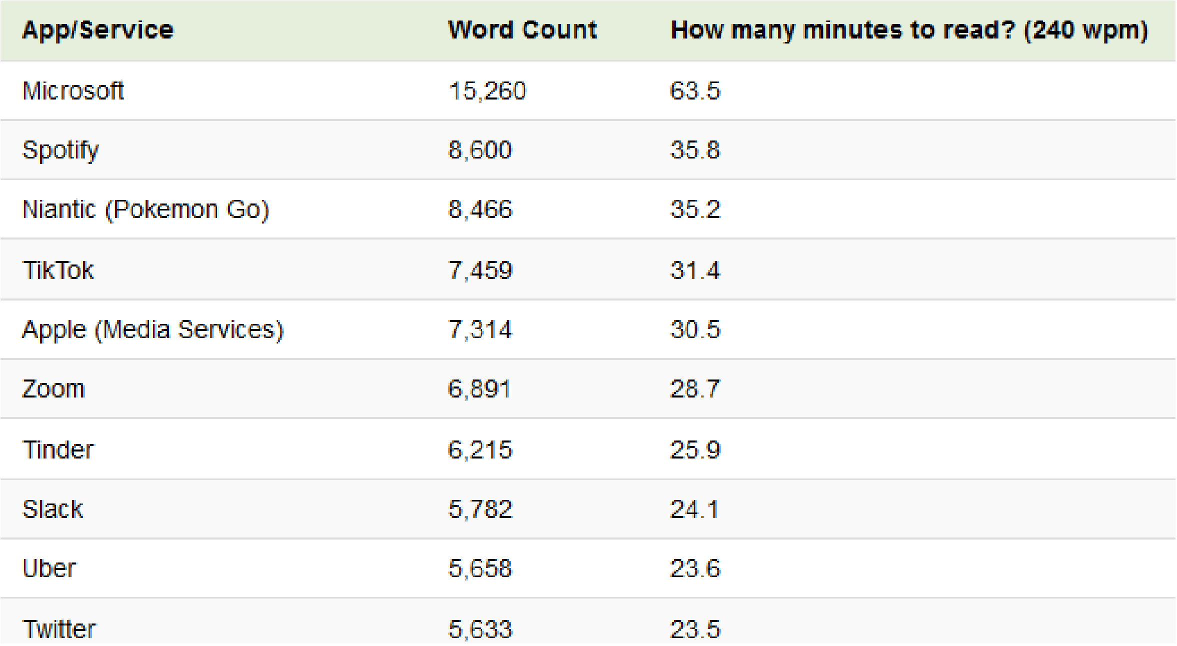
There are many issues concerning the privacy of users; what data applications are collecting and how they are processing personal information. We can take as an example permissions demanded by applications. Permissions give applications privileges to access some specific data. That may be required for the good functioning of applications. For example, Google Maps needs to access the geographical localisation of the terminal. However, it is not always the case. Some applications have access to unnecessary data. In a report made by an Avast researcher, hundreds of android flashlight applications ask for many permissions that, in the vast majority of cases, aren't justified. The average number of permissions is 25, with two applications asking for 77 permissions in total as we can see in the table below :



| No. | App Name | Permissions Count | Number of Downloads |
|---|---|---|---|
| 1 | Ultra Color Flashlight | 77 | 100,000 |
| 2 | Super Bright Flashlight | 77 | 100,000 |
| 3 | Flashlight Plus | 76 | 1,000,000 |
| 4 | Brightest LED Flashlight — Multi LED & SOS Mode | 76 | 100,000 |
| 5 | Fun Flashlight SOS mode & Multi LED | 76 | 100,000 |
| 6 | Super Flashlight LED & Morse code | 74 | 1,000,000 |
| 7 | FlashLight – Brightest Flash Light | 71 | 1,000,000 |
| 8 | Flashlight for Samsung | 70 | 500,000 |
| 9 | Flashlight – Brightest LED Light & Call Flash | 68 | 1,000,000 |
| 10 | Free Flashlight – Brightest LED, Call Screen | 68 | 500,000 |

These permissions concern recording audio, reading contact, and some other dangerous accesses like the ability to kill background processes, place phone calls, handle SMS messages, or trigger downloads without notifying the user. Many of these applications are used by malwares, and the data collected could be weaponized against users. This is not the only way of exploiting data. Many applications buy the data collected to some third-party services or organizations. They can be stored in databases for user patterns in order to target products to the ideal audience at lower costs than the traditional marketing.

It is true that users take a part of responsibility when accepting the terms of services provided by applications. Actually, 97% of people between 18 and 34 years old agree to the conditions without reading them. But, at the same time, terms of conditions are made in a way that discourages users reading them. They use a legal language with complex sentences, and require a non negligible amount of time to be read as it is represented in the table below ( with an average speed of 240 wpm ) :



| App/Service | Word Count | How many minutes to read? (240 wpm) |
|---|---|---|
| Microsoft | 15,260 | 63.5 |
| Spotify | 8,600 | 35.8 |
| Niantic (Pokemon Go) | 8,466 | 35.2 |
| TikTok | 7,459 | 31.4 |
| Apple (Media Services) | 7,314 | 30.5 |
| Zoom | 6,891 | 28.7 |
| Tinder | 6,215 | 25.9 |
| Slack | 5,782 | 24.1 |
| Uber | 5,658 | 23.6 |
| Twitter | 5,633 | 23.5 |

## V. Conclusion

To sum up, this paper focused on modelizing the traffic generated by mobile applications, and some of the challenges they face. Many researches have been made in this context. This bibliographic study gives a brief summary of these studies that will be potentially beneficial in our project.

## VI. References

- https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

- https://ieeexplore.ieee.org/abstract/document/7846960?casa_token=oDggBsIe-o4AAAAA:8HH35RUFRRDLMq4x-yvw3d68s9fo1XF2G8j-fk7LYEHNSG_c85F7wApgle_jFqsbKrmCAtjw

- https://ieeexplore.ieee.org/abstract/document/6911870/?casa_token=vOOtBN9kFmAAAAAA:d1PL2YdSVIvPbkMjWg-ioDN3qhfAy2uPxU4yBtMmdFvp8-v2AIhi2faHJn4Irf-UyJdlFCSk

- https://ieeexplore.ieee.org/abstract/document/8706824/?casa_token=pP6rE7oynrYAAAAA:NMf6YxpPMJ1jr0azlHIf3tgdN80yxJOCnScXy8bcMAGNRrqqwQgPYJbLmQCAq8xP4EsTns56

- https://link.springer.com/chapter/10.1007/978-3-662-54970-4_22

- https://dl.acm.org/doi/pdf/10.1145/2413247.2413286?casa_token=L-7OA9dLDvYAAAAA:JMMtP6EwN3juvwdqeByNJWhIL0N6ZbxeMuu4sHVsjZVoJ7PW5FHifJrCveJ_Ay24p3V_nC59Oals

- https://cris.unibo.it/retrieve/handle/11585/753253/789043/privacy_furini.pdf

- https://www.researchgate.net/publication/254008749_Periodic_Transfers_in_Mobile_Applications_Network-wide_Origin_Impact_and_Optimization

- https://dl.acm.org/doi/abs/10.1145/2789168.2790107?casa_token=vWYWYO0s1ZEAAAAA:N0CoG4X3oJuF31KLAtHuZLdE1t_r-GhIDuAdH0nMZ3sP-dswE0VoRzuuGoxOvD_qiZn0P_UzU3DZ

- https://search.proquest.com/openview/4cd2ca7eae004969ed20cec02c8f104e/1?pq-origsite=gscholar&cbl=1006393&casa_token=nOH7CHeL6gcAAAAA:n80VU6MzoWPpVKPiAmlsqN-5dwN4CNFIcbSod5ZlWEdeOnFrrPET4Py7_b4TSfZ-vSuR3tKamDw

- https://ieeexplore.ieee.org/abstract/document/6243128/

- https://ieeexplore.ieee.org/abstract/document/6757895

- https://www.visualcapitalist.com/terms-of-service-visualizing-the-length-of-internet-agreements/