

Methods of Monitoring and Collection of User Data on the Example of Facebook

Andreja Samčović^{1,*}

¹Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade, Serbia

Abstract

This paper presents a global overview on the tools and processes that are used every day to collect data on all platforms where the social networks are present. Furthermore, this work reviews the various channels of data application. The central theme is based on drawing the attention of users in all the ways in which the owners of the social networks monitor and record activities of the users. Mechanisms of data collection are described on the example of Facebook, depending on device and users' choice. Some problems of privacy and security are considered in Meta virtual reality.

Keywords

privacy, user data, information, social network, Facebook

1. Introduction

The human need for connection and communication has historically been one of the strongest motivations for development technology and society. In the modern information society, it is possible to exchange stories and experiences with people who are on the other side of the planet. One of the most popular channels of communication in the modern age is represented by social networks. The leader in the social network sector according to annual revenue and number of users is Facebook [1]. As a social network, Facebook is one of the largest generators of user data. The mechanisms and methods by which a company Facebook.com collects user data vary depending on a large number of different factors.

The first version of Facebook was open only to Harvard University students. The basic capability provided by Facebook is to collect and display information about the identity of all students at Harvard. The popularity that the network achieved in record time gave an additional wind behind the team, and after an initial period of slightly less than two years, Facebook was officially launched globally in 2006. The great interest in Facebook on a global scale is best described by the fact that it already reached the magical number of one million active users during the first year. Some of the key functionalities that have attracted users are primarily the ease of connecting with other users on the platform. The simplicity and effectiveness of the "add a friend" button is something that the competition had not applied until then.

The design philosophy is exactly what Facebook man-

aged to set apart from the competition. While other platforms focused on a robust, powerful system that gives full control to the user, and did not focus on ease of use and responsiveness of the pages. Facebook was focused on simplicity and uniformity of user profiles and experiences. In the beginning, Facebook's biggest competitor on the market was the social network Myspace [2].

The biggest difference in the approach of these two social networks is reflected in the issue of user identity online. On the Internet, before the time of Facebook, a degree of anonymity was assumed on Internet sites and social networks [3]. Generated usernames were used, where great emphasis was placed on the username not containing the user's real credentials.

Facebook invoked a tectonic change in the situation at the time by allowing people to present themselves with their first name, last name and profile picture, as well as additional information about themselves if desired. The design focused on transparency and de-anonymization of users at the time was very positively received by users. The biggest advantages are that users finally have the opportunity to present themselves on the Internet, and to make real, sincere friendships with people from all over the world.

2. Voluntary sharing of user data

While modern social networks have a wide range of technologically advanced mechanisms that monitor and record the behavior of users online at all times, not all responsibility can be pointed at the creators of the network. Perhaps a seemingly less technologically impressive, but far more elegant solution to obtain user information was to simply ask users to provide that information. An impressive number of psychologists, anthropologists and experts in human behavior and habits have worked day and night to create a platform that will convince people

BISEC'22: 13th International Conference on Business Information Security, December 03, 2022, Belgrade, Serbia

*Corresponding author.

✉ andrej@sf.bg.ac.rs (A. Samčović)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

and instill confidence in them that they can leave their personal data to an unknown third party. Facebook had a pioneering status in this regard because they tried, and years later it turned out that they succeeded, in changing the previously anonymous, impersonal nature of the Internet into an environment where people have confidence and desire to stand behind their views with their first name, last name, and their real picture.

Apart from presenting Facebook as a safe platform that will take care of the privacy of users and their data, another key factor in creating the entire user experience of the site is the big message that appears on the first page of the platform, when creating a new account [4]. The message emphasized in the foreground is "It's free and always will be". Nowadays, the attraction and influence of free goods and services on the individual have been thoroughly investigated and the mechanisms that are activated in the human brain when they see a free offer are used daily in the manipulation and persuasion of users on the Internet. Some of the information that users voluntarily provide to Facebook, and share publicly with their "friends" online are [5]:

- Name and surname of the user account;
- The e-mail address used to create an account - mandatory confirmation of the validity of the e-mail address;
- User's date of birth - necessary when registering an account;
- Gender of the user;
- Mobile phone number - required to confirm the user account;
- Synchronization of contacts from the mobile phone - access to all phone numbers and names of contacts on the mobile device;
- Synchronization of e-mail contacts - access to all addresses and contacts from the registered e-mail account;
- Data generated through social media posts;
- Data generated through correspondence with other online users;
- Linking accounts with sites that have the option to integrate with Facebook.

In the initial stages of the platform's development, given the technological limitations, the greater focus on obtaining user data depended directly on the user's willingness to leave that data. With the advancement of technology the integrations became more serious. The rudimentary and relatively high complexity of the Internet browser systems at the time made the initial steps in creating a user tracking ecosystem quite difficult, but with the great interest of companies as well as the coming wave of popularity of social networks, different networks started to create their own solutions to the current

problem. The foundation of the system that has grown into today's most comprehensive individual tracking system was based on technology that was created almost a decade before the creation of Facebook. In the initial stages, Facebook relied on internet cookie technology.

Today, it can be said with great certainty that there is no site on the Internet that does not use the functionality of internet cookies, at least at the lowest functional level. Facebook, of course, did not miss the opportunity to implement cookies on its social network. The complexity and thoroughness of the information that could be collected using cookies grew over time, and so did the amount of information [6]. Given that Facebook's focus was selling advertising space to the right clientele, it was clear from the start that the imperative was to gather as much information as possible. Apart from the information that users voluntarily left on the platform, which was only a fragment of the overall picture, of significantly greater interest to Facebook and client companies were the "invisible" behaviors of users on the social network. Internet cookies made exactly that possible. Some of the information that is still recorded and stored in Facebook's internet cookies, and later transferred to the company's servers for analysis and further processing, can be divided depending on the sphere of interest that needs to be processed [6].

In the early years of Facebook, the vast majority of users used Internet browsers to access Facebook. The information that was of interest at the time, and related to the device from which Facebook was accessed, was:

- Name and version of the user's internet browser;
- Browser window size;
- Device operating system;
- Devices connected to the computer;
- Device configuration, in the initial stages exclusively related to computers;
- Whether the access is via DS (Digital Subscriber Loop), dial-up, or one of the mobile Internet access;
- IP address of the device being accessed.

Cookies that are created with the aim of obtaining an overall technological picture of the user may not appear to be invasive or too much of a threat to the user's privacy, but they are not the only cookies used. The applications and potential abuses of information obtained through internet cookies are comparatively more important when it comes to users who do not have an online account, and when it comes to sites that do not have direct integration with Facebook.

Internet cookies alone could fulfill the requirements that were needed, but their big drawback is that users would very quickly get an insight into the true amount and type of data that Facebook collects from them. The

situation on the market at the time led Facebook to a situation where it was necessary to develop new solutions that would be oriented towards the future. Information about Facebook's tools used for in-house analysis is almost unavailable, inaccessible, and it is difficult to get a glimpse into the real system behind the aforementioned giant social network.

Having that in mind, it is important to emphasize that the entire functionality of Facebook's tools and mechanisms is still largely undiscovered, and for good reason. The information that is currently available to the public is enough to raise serious concerns among network users, not only for them, but also for people who do not have accounts on social networks. The technologies that Facebook itself publicly admits to using on its platforms are solutions that have grown over the years in their capabilities and the amount of data they can collect.

The mechanisms that are implemented have different names, and it is difficult to distinguish between them without looking "behind the curtain" into the protocols and the exact processes that these mechanisms execute. In professional literature, the terms tracker, follower, social plugins, embedded tracker, deeply integrated trackers, highly specialized trackers and many others are mentioned.

The list of data collected through trackers is not complete, but the knowledge and information that is available is enough to raise significant questions about the legality, necessity, ethical and moral correctness of data collection. It is important to emphasize that the systems are purposefully designed and created with the user in focus. A visual presentation of the collected data can be found in Figure 1 [6].

Some of the information that there is evidence of being collected through trackers are:

- Friends in social circles;
- Groups, pages, membership and activities therein;
- All interaction with posts and other members;
- Monitoring of "liked" activities;
- Monitoring activities related to events posted online;
- Collection and monitoring of data on photos and video content;
- Information about time spent on profiles and contents;
- Information about published content;
- Assessments of the riskiness of user behavior;
- Tracking cursor and keyboard activity on the site.

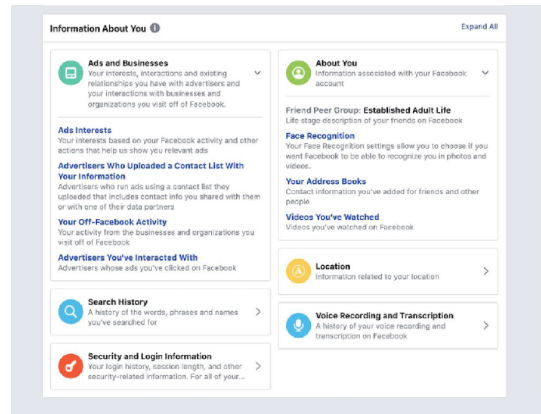


Figure 1: Visual presentation of collected data.

3. Collection of user data on mobile devices

The first decade of the third millennium was marked by the rapid development of Internet technologies and companies specialized in doing business in the new era. Facebook has developed systems for tracking and collecting data on user computers to a level where it comfortably took over an ever-larger part of the market from competing marketing companies. A big change in the field of information technology was brought about by the popularization of portable devices.

Thanks to the development of new platforms for accessing the Internet, there have been changes in the behavior and habits of the users themselves. Sites that until then were primarily designed for access from desktop computers had had to be adapted to new formats and user requirements. The large heterogeneity of devices from the technological side put the internet platforms at the time in a problem, because it became necessary for the user experience on all devices to be the best possible. It became clear that adapting the site for all devices is simply not profitable, and that it is necessary to find a new solution for the new era of devices. The direction the industry turned was to switch to specialized stand-alone applications instead of the previous focus on accessing platforms via internet browsers, where the site was the focus, and the central point of user interaction with many sites. In order to maintain their popularity, social networks have been forced to engage in the struggle to create the best user application for iOS, Android and other operating systems, primarily designed for portable devices.

Facebook initially had difficulty creating a successful app, and they ran into a lot of problems during initial development. Some of the biggest problems they

had to overcome were related to the limitations of the processing power of mobile devices. At the beginning of the second decade, mobile devices had relatively modest chips, which did not provide amazing performance, but the priority was on economy and extended battery life, because it was necessary to use energy to power new touch screen screens of larger dimensions. The lower processor power compared to desktop computers made it quite difficult to implement a monitoring and collection system, but it was not the only limitation. The memory on portable devices was very limited and relatively slow, so it was necessary to keep the amount of data at the lowest level, so that users would not be forced into a situation where they have to delete the application in order to be able to use their device. For both factors, the company managed to make a compromise and find a solution to adapt to the new situation. The factor they have not been able to deal with in the best way is the limitation of data flow.

Namely, on a global level, access to fast broadband Internet for mobile devices has not yet been solved in the best possible way. Data transfer was slow, partly due to poor network infrastructure, partly due to device performance, and expenses for the user because service providers had limited capacity and network coverage. Facebook then found itself in a difficult situation because it faced the choice of reducing the amount of data that it will exchange with users or to continue collecting and researching a completely new market of data and passing that cost to the user. The choice from Facebook's point of view was clear, the priority was on profit and they continued with their policy of aggressively collecting user data. A fact that Facebook saw earlier than its competition was the potential of new information it could collect from users' mobile devices.

Facebook noticed that there was a whole new dimension of data that it could be collected, analyzed and applied thanks to new platforms and technology. The biggest difference compared to desktop devices is mobility. Until then, computers were mostly fixed in one place and used from the same place, which did not give companies the opportunity to monitor the movement of users, because they accessed social networks only through the most frequent fixed devices. Information about the location and movement of users has become a kind of "gold mine" for collecting user data. The application of collected information about the user's location has become a priority, due to its great analytical and monetary value. One example of collected information is shown in Figure 2.

A significant difference between desktop and mobile devices is that mobile devices have a microphone, nowadays more than one, which is always active. Facebook's app had the ability to access the devices' microphones as

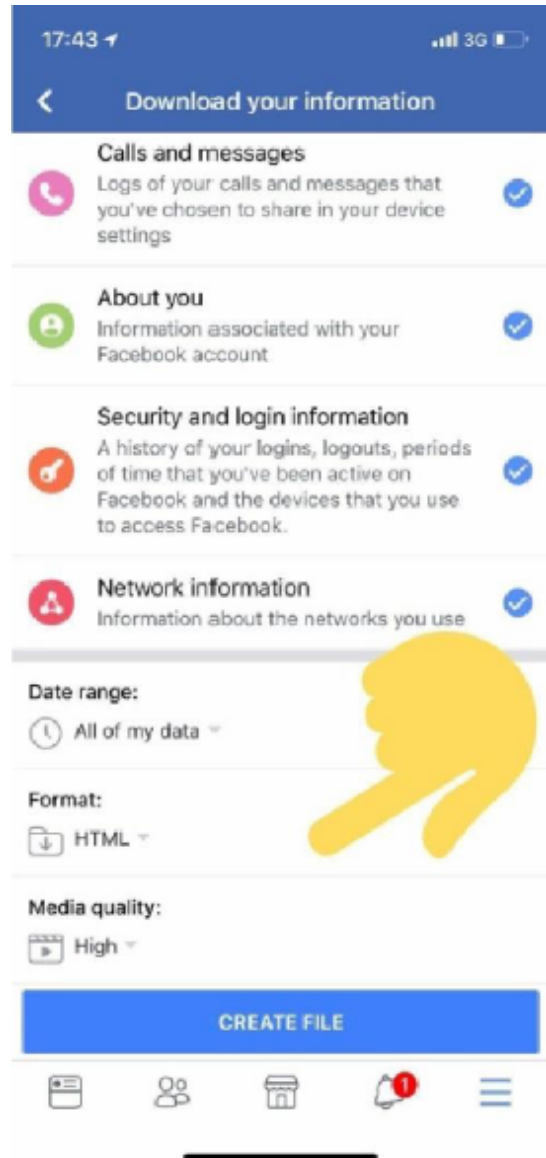


Figure 2: Collected information on mobile device.

well as the camera. While users thought that the app only listens when voice messages are being recorded, it turns out that once the user allows access to the microphone, the app has the ability to have constant access, both when the app is active and when it's not, and when the phone is active, and when is in standby mode. Facebook suffered a great outcry from users, with a large number of people leaving the network, but in the end the fact still went unnoticed by most users.

The camera on mobile devices is also in the same situation. Without the user's knowledge and confirmation, the app can access the camera. Almost all metrics that were used to collect data on desktop computers have been adapted for mobile applications. One of the new sets of information that has become of much greater interest is the set of information about the use of the device itself. It primarily refers to the time spent with the screen, and the way and time the user spends on the device, which was one of the tricks up the sleeve for Facebook's further business.

The app required permission to write data to the device's local storage to function properly, but Facebook went a step further and enabled scanning and manipulation of files stored on the device. While there is currently no publicly available evidence of malicious manipulation of data on the devices, it is publicly available that Facebook reviews and checks the files on the phone with the aim of analyzing all other applications that are installed on the mobile device. Numerous procedures and investigations have been launched into the legality of tracking users within these limits. Facebook's smart moves at the dawn of the smartphone era have allowed the company to further distance itself from the competition in terms of revenue, popularity and technology. Users got new functionalities and global connectivity with all other users, and in principle for free, but what is the real price that users pay?

4. Collection of user data through virtual reality systems

The next revolution that has raised the global telecommunications and Internet industry to a higher level is the development and progress in the field of virtual reality (VR) technology [7]. In the middle of the previous decade, emphasis was placed on creating virtual environments for socializing and communicating with other people, with the aim of bringing people who are on opposite ends of the world closer together digitally. As before, Facebook was among the first to see the potential value and perspective of this idea and technology and joined in the earliest stages of creation.

The confidence invested in this project is best described by Facebook's purchase of Oculus Rift for \$2 billion, then a pioneer in the design and development of software and hardware for virtual reality. As has been established so far, new technologies have brought with them new challenges that must be overcome. The differences compared to the previous changes are enormous, and the biggest ones are reflected in the very scope and ambition of the project.

The idea, which has been prepared, represents the unification of all Facebook networks of applications and



Figure 3: Metaverse scene.

services under one name - Meta, and the virtual reality project is available to users - Metaverse [8]. One example of Metaverse scene is on Figure 3. After almost a decade of development, the project is only on the verge of testing with a limited number of users. Facebook did take over the production of hardware and software in cooperation with Oculus Rift, but due to the circumstances of the global crisis and the pandemic, the project was forced to slow down, and the public launch was postponed. The aspects of the project that have not been delayed relate to the hardware components that were already with users under the umbrella of Oculus Rift, while Facebook contributed its software environment and support.

For optimal use of virtual reality services, it is necessary to adjust living space in a way that will allow the easiest use and interaction with the digital environment. The recommendation of companies that market services based on virtual reality create the best user experience is based on the creation of an optimal environment for monitoring individual behavior. In order to register movement in the space of virtual reality, it is necessary to install cameras in living space that will monitor the location and change of location of the user while using the platform. When configuring the system for the first time, it is necessary to calibrate the system to recognize the individual and adequately adjust other aspects of the system to best adapt to the user's inputs.

For the optimal operation of Meta, and virtual reality software, it is necessary to collect and process a considerable number of different data from a large number of input devices located in the hardware system. During the actual use of the network, the system monitors and records data that until the recent past were difficult to imagine, at least on a physical and medical level. Data collected in more physically demanding activities that can be experienced through virtual reality are of great value to the entire health, pharmaceutical and medical insurance industry. Data on fitness, physical activity and vital data such as blood pressure and percentage of oxygen in the blood represent very specific data about users,

because the way in which they are processed and then distributed to the users themselves represents a complex system [7].

Through the use of virtual reality technology, all users agree that their interaction is actively monitored and analyzed. Through the terms of use of Meta, it is stated that everything that happens in the Metaverse, i.e. in the space of virtual reality, can be collected and analyzed by Facebook. The way users consume virtual reality is significantly more active and present than social networks. Interaction with other users is based on active conversation, not messages. Although that functionality also exists, the emphasis is placed on an approximately authentic resemblance to reality. Users have the option to create their own digital character based on their own appearance and to enter into virtual chatrooms, games or creative activities, while on the other hand, business users are offered the option of creating virtual meetings and joint work activities. Users must be aware that every online exchange is under the direct scrutiny of Facebook and will be used to create user profiles and maximize the profits that users will generate for the company.

The conclusion that can be drawn from the knowledge of Facebook's way of doing business on the virtual reality project is as impressive as it is worrying. So far, no technology has been invasive at the level of virtual reality, and in a way it represents a benchmark, and on the other hand, a kind of pinnacle in the creation of technology for monitoring and collecting user activity and data that was not possible before. Virtual reality is still a relatively young technology. The consequences of its existence will not be fully known for some time, but if previous events and the experiences of early adopter users are preludes to future developments and potential problems, attacks on user privacy will be increasingly aggressive and invasive.

5. Conclusion

The value of user data and information on the market shows a growing tendency, and therefore the need to collect an increasing number of them. Through its many years of work, Facebook has developed mechanisms and methods that have brought the company to a convincing leadership position in terms of the value and amount of collected data. The company followed technological trends and marketed its platform on almost all possible devices that support the functionalities necessary for the correct operation of the network.

Historically, Facebook was initially designed as a platform for desktop computers, then the focus was on smart devices, while in the last few years the priority has shifted to virtual reality technology. Facebook's presence on a wide range of different platforms enables the collection of over 4 petabytes of data on a daily basis. User data

is obtained by applying various solutions, from internet cookies, integrated trackers, tracking activities on websites, tracking contacts, sensors and memory on mobile devices, all the way to home mapping, constantly on microphones with voice recognition and cameras with user face recognition.

While the sanctions and fines the company suffered were comparatively mild in relation to the damage they caused, some users gave up using the network, while the vast majority continued to spend time on Facebook. Legislation regarding the protection of user data has been tightened in recent years, but company still left with too much room for abuse. Most of the responsibility lies with users and their desire and effort to try to protect their privacy from the prying eyes of big companies.

The development of tools to protect user privacy has gained popularity in the past few years, mostly due to the publication of numerous lapses made by companies dealing with social networks. In the center of attention are Internet browsers and their options that hide at least part of the activity from trackers, while on mobile devices monitoring is limited through access permissions to various device modules. Despite all the protections, the best way for the user not to share his data with companies is to neither create nor access different social networks. In the modern age, every user has the right whether he wants to access the network or not, and thus bear all the consequences of his decisions.

Bearing in mind the facts presented, it is necessary to educate users, familiarize them with the ways in which potentially dangerous data about them is collected, and point them to some of the solutions that will protect their privacy while using the Internet. On the other hand, it is necessary to put pressure on the legislators to make the business rules of companies based on the collection and processing of user data, by tightening the penalties and sanctions in cases of "data leakage" or unauthorized access to user data.

Acknowledgment

This work is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia [projects TR32025, TR32048].

References

- [1] Datareportal, Facebook Statistics and Trends, 2022. URL: <https://datareportal.com/essential-facebook-stats/>.
- [2] A. Hartung, How Facebook beat MySpace, Forbes. Retrieved from (2011). URL: <https://www.forbes.com/sites/adamhartung/2011/01/14/why-facebook-beatmyspace/?sh=6790af0d147e/>.

- [3] T. Aichner, M. Grünfelder, O. Maurer, D. Jegeni, Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019, *Cyberpsychology, behavior, and social networking* 24 (2021) 215–222.
- [4] Facebook, Manage Your Privacy, 2022. URL: <https://www.facebook.com/about/basics/manage-your-privacy/>.
- [5] Facebook Help, Cookies on Facebook, 2022. URL: <https://www.facebook.com/help/336858938174917/>.
- [6] S. P. Singh, 32 Types of User Data Facebook Collects; This is How You Can See Yours, 2021. URL: <https://gadgetstouse.com/blog/2021/01/20/32-types-of-data-facebook-collects-this-is-how-you-can-see-yours/>.
- [7] E. Dick, Balancing user privacy and innovation in augmented and virtual reality, Technical Report, Information Technology and Innovation Foundation, 2021.
- [8] J. P. Calbalhin, Facebook user's data security and awareness: A literature review, *Journal of Academic Research* 3 (2018) 1–13.