NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE

DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

POSTGRADUATE PROGRAM

Computer, Telecommunications and Network Engineering

THESIS

Privacy on the Web:
Analysing Online Advertising Networks

Apostolos-Andreas H. Konstantinidis

Georgios G. Sargologos

SUPERVISOR: **Konstantinos Limniotis,** External Instructor

ATHENS

FEBRUARY 2021

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

Μηχανική Υπολογιστών, Τηλεπικοινωνιών και Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Privacy on the Web:
Analysing Online Advertising Networks

Απόστολος-Ανδρέας Η. Κωνσταντινίδης

Γεώργιος Γ. Σαργολόγος

Επιβλέπων: **Κωνσταντίνος Λιμνιώτης,** Διδάσκων εκτός Τμήματος

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2021

THESIS

Privacy on the Web:

Analysing Online Advertising Networks

Apostolos-Andreas H. Konstantinidis

A.M.: en319003

Georgios G. Sargologos

A.M.: en319007

SUPERVISOR: **Konstantinos Limniotis,** External Instructor

EXAMINATION COMMITEE **Lazaros Merakos,** Professor

Nikos Pasas, Professor

February 2021

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Privacy on the Web:

Analysing Online Advertising Networks

Απόστολος-Ανδρέας Ηρακλή Κωνσταντινίδης
Α.Μ.: en319003
Γεώργιος Γεωργίου Σαργολόγος
Α.Μ.: en319007

ΕΠΙΒΛΕΠΩΝ: **Κωσταντίνος Λιμνιώτης**, Διδάσκων εκτός Τμήματος

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ **Λάζαρος Μεράκος**, Καθηγητής
Νίκος Πασσάς, Καθηγητής

Φεβρουάριος 2021

ABSTRACT

This thesis mainly focuses on the big picture of the current stage of user data collection and protection and at the same time discuss scientific solutions that aim at protecting web users from numerous privacy threats which are caused by the online advertising industry. There is an extensive analysis about the infrastructure, the interactions among the elements, and the technologies enabling the delivery of ads and the use of personal data. The paper provides scientific details in order the reader to understand the present advertising ecosystem, and the privacy risks users are exposed to. There is a detailed discussion about the measures taken by European Union with a single purpose, to put in order the chaos that exists in the global system of interconnected computer networks called Internet. Except of the analysis of the GDPR, there is an investigation about the trackers that use our data while we access any domain and the results are provided precisely. Finally, there is a part, where some possible solutions of tracking users are discussed, after the fade of third party data due to the strict measures regarding users data collection, announced by the EU.

SUBJECT AREA: Privacy in Online Advertising

KEYWORDS: User data collection, online advertising, privacy, user data protection, advertising networks, trackers, malicious, GDPR, real time bidding

ΠΕΡΙΛΗΨΗ

Το θέμα της διπλωματικής αφορά ζητήματα ιδιωτικότητας στο Διαδίκτυο, με έμφαση την ιχνηλάτηση χρηστών για το σκοπό στοχευμένων διαφημίσεων. Θα μελετηθούν οι τεχνολογίες που χρησιμοποιούνται σε αυτήν την κατεύθυνση, και θα αξιολογηθεί η τρέχουσα κατάσταση υπό το πρίσμα του σχετικού νομικού πλαισίου, που είναι ο Γενικός Κανονισμός Προστασίας Δεδομένων(GDPR- 2016/679) της ΕΕ, αλλά και η e-Privacy οδηγία. Πιο συγκεκριμένα αναλύονται οι προσπάθειες των διαφημιστικών δικτύων, ώστε να λαμβάνουν όσο το δυνατόν περισσότερες προσωπικές πληροφορίες από τους χρήστες, οι οποίες χρησιμοποιούνται κυρίως για εμπορικούς σκοπούς, οι οποίες ωστόσο μπορεί και να υπερβαίνουν τη στοχευμένη διαφήμιση. Υπάρχει μια εκτεταμένη ανάλυση σχετικά με την υποδομή, τις αλληλεπιδράσεις μεταξύ των στοιχείων και τις τεχνολογίες που επιτρέπουν την προβολή διαφημίσεων και την ανταλλαγή προσωπικών δεδομένων. Τέλος, αναλύονται ορισμένες πιθανές λύσεις σχετικά με την παρακολούθηση των χρηστών, έπειτα από την κατάργηση των third party data λόγω των αυστηρών μέτρων σχετικά με τη συλλογή προσωπικών δεδομένων, που εφαρμόζει η ΕΕ.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Ιδιωτικότητα, Διαδικτυακή διαφήμιση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Διαδικτυακή διαφήμιση, παρακολούθηση χρήστη, ιδιωτικότητα,
Γενικός, Κανονισμός Προστασίας Προσωπικών Δεδομένων

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα εργασία αποτελεί διπλωματική εργασία στα πλαίσια του μεταπτυχιακού προγράμματος «Μηχανική Υπολογιστών, Τηλεπικοινωνιών και Δικτύων» με ειδίκευση την «Δικτύωση Υπολογιστών» του τμήματος Πληροφορικής και Τηλεπικοινωνιών του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών. Με την ολοκλήρωση της, θα θέλαμε να εκφράσουμε τις θερμές μας ευχαριστίες σε όλους όσους συνέβαλαν στην εκπόνησή της. Ευχαριστούμε θερμά τον επιβλέπων καθηγητή, κύριο Λιμνιώτη Κωνσταντίνο, για την εμπιστοσύνη που μας έδειξε από την πρώτη κιόλας στιγμή, αναθέτοντάς μας το θέμα αυτό, για την επιστημονική του καθοδήγηση, τις υποδείξεις του, την διαθεσιμότητα και συνεισφορά του, και την συνεχή του υποστήριξη από την αρχή μέχρι το τέλος. Ακόμη θα επιθυμούσαμε να εκφράσουμε την αμέριστη ευγνωμοσύνη μας στις οικογένειες μας για όλη τη στήριξη και τη συμπαράσταση, καθ' όλη τη διάρκεια των μεταπτυχιακών σπουδών μας.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

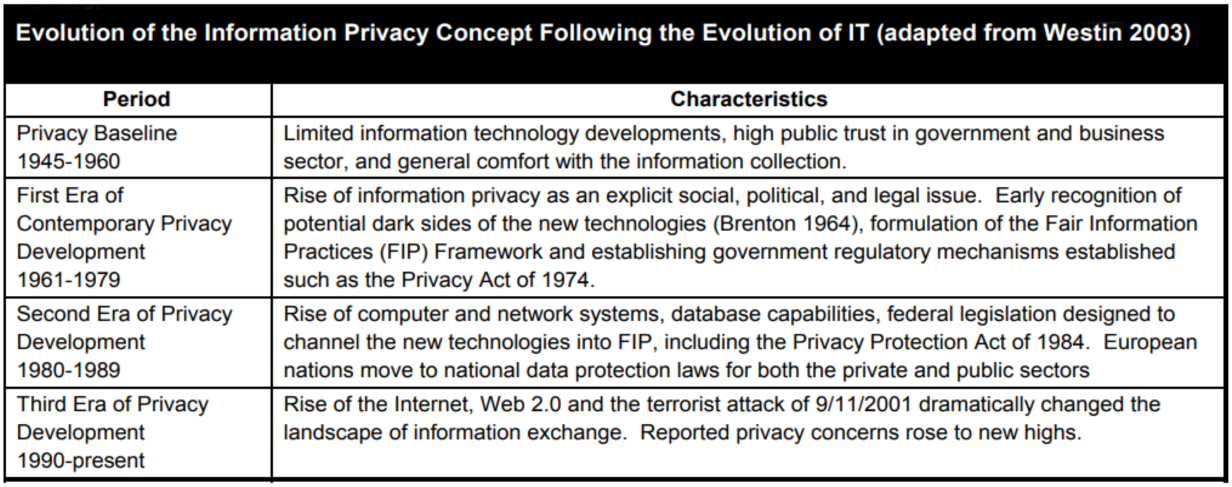
ΠΡΟΛΟΓΟΣ

Τα πλαίσια κάτω από τα οποία εκπονήθηκε η διπλωματική εργασία, μας έδωσαν την ευκαιρία να αναπτύξουμε τις ομαδικές και τήλε-ερευνητικές μας ικανότητες μιας και οι συνθήκες που επικρατούσαν παγκοσμίως λόγω COVID-19 δεν μας επέτρεψαν την δια ζώσης συνεργασία ή την πρόσβαση σε βιβλιοθήκες με φυσική υπόσταση. Η συγγραφή της παρούσας έρευνας έλαβε μέρος στη Σύρο και στην Αθήνα. Το διαδίκτυο παρ' όλα αυτά, και οι υπηρεσίες του, κατάφεραν να μας εξασφαλίσουν σε πολύ μεγάλο βαθμό την πρόσβαση στις πληροφορίες και στα επιστημονικά δεδομένα που επιθυμούσαμε προσφέροντας μας όλα τα απαραίτητα εφόδια που χρειαστήκαμε.

1. INTRODUCTION

1.1 Problem Statement

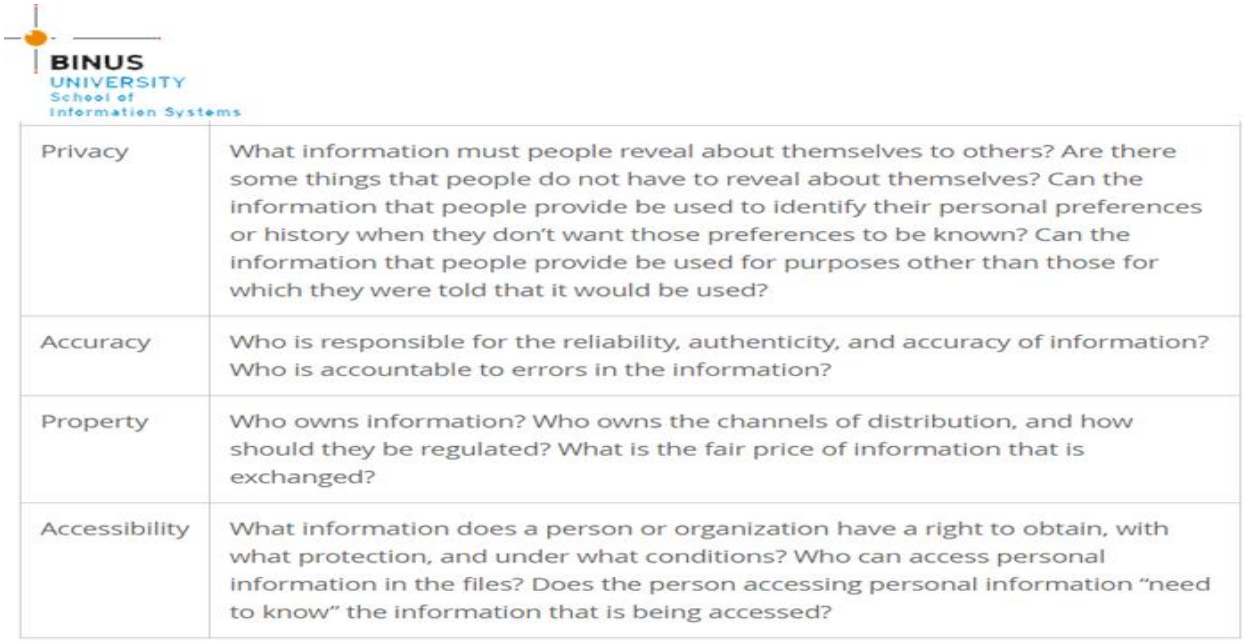
In 1890, many years before the world of digital data, telecommunications, social media and advanced algorithms, the privacy approach was defined as “The right to be let alone. The right to liberty secures the exercise of extensive civil privileges” by Warren and Brandeis[41].Historically, it is broadly known that the evolution of privacy approach and information privacy in particular, follows the evolution of information technology itself, as can be seen in **Figure 1**. Focusing on the earliest privacy guideline , around 1950, the collection of personal data was happening in the minimum degree or it was very well hidden as the free speech was not formed as it is today , and as a result there was faith and belief in government and business sectors. On the other hand, at the latest privacy guideline -the present, the privacy concerns are considered to be at the peak, as the business sector has made a huge progress and at the same time the political section seems to be corrupted at one percentage. The collection of personal information was firstly faced as an important issue in the middle 1970s. The collection and the analysis of user information is now of absolute and growing importance for both online and offline companies. The current technologies that enable the capture and analysis of large volumes of data and in near real time are known as Big Data. Using the Big Data approach, there were many improvements in online platforms and personalized services like Netflix, BO max, social networks, online shopping and generally in automation of daily life, like traffic guidance services. The frameworks for information privacy are plenty, but the one adjusted especially for information systems is the privacy, accuracy, property and accessibility (PAPA) framework by Mason as seen in **Figure 2.**[42],[43]



Evolution of the Information Privacy Concept Following the Evolution of IT (adapted from Westin 2003)

Period	Characteristics
Privacy Baseline 1945-1960	Limited information technology developments, high public trust in government and business sector, and general comfort with the information collection.
First Era of Contemporary Privacy Development 1961-1979	Rise of information privacy as an explicit social, political, and legal issue. Early recognition of potential dark sides of the new technologies (Brenton 1964), formulation of the Fair Information Practices (FIP) Framework and establishing government regulatory mechanisms established such as the Privacy Act of 1974.
Second Era of Privacy Development 1980-1989	Rise of computer and network systems, database capabilities, federal legislation designed to channel the new technologies into FIP, including the Privacy Protection Act of 1984. European nations move to national data protection laws for both the private and public sectors
Third Era of Privacy Development 1990-present	Rise of the Internet, Web 2.0 and the terrorist attack of 9/11/2001 dramatically changed the landscape of information exchange. Reported privacy concerns rose to new highs.

Figure 1 - Evolution of Privacy Concept [41]



BINUS UNIVERSITY School of Information Systems

Privacy	What information must people reveal about themselves to others? Are there some things that people do not have to reveal about themselves? Can the information that people provide be used to identify their personal preferences or history when they don't want those preferences to be known? Can the information that people provide be used for purposes other than those for which they were told that it would be used?
Accuracy	Who is responsible for the reliability, authenticity, and accuracy of information? Who is accountable to errors in the information?
Property	Who owns information? Who owns the channels of distribution, and how should they be regulated? What is the fair price of information that is exchanged?
Accessibility	What information does a person or organization have a right to obtain, with what protection, and under what conditions? Who can access personal information in the files? Does the person accessing personal information "need to know" the information that is being accessed?

Figure 2 - PAPA framework by Mason

1.2 Purpose

This thesis focuses on the privacy issues stemming from the behavioral advertising and the relevant technologies today. More precisely, a main goal is to provide a comprehensive description of the modern online advertising infrastructure, as well as the necessary information to explore users' evaluation of privacy issues in relation to online advertising. To this end, we describe the main actors of the advertising ecosystem, the technologies and interaction between them, the legal framework of operation and how the companies can face its restrictions. Another goal is letting the readers know about the measures imposed by the relevant legal framework, such as the General Data Protection Regulation and the e-Privacy Directive. Lastly, there is a detailed reference about possible privacy threats that users might face in the existing online advertising ecosystem and what the protections methods are used. Popular web sites are explored in terms of the usage of third-party advertising cookies, to evaluate the current status on their compliance with the relevant legal framework, as well as the possible privacy threats that arise.

1.3 Research Questions

1. Do users really receive the content they want free of charge in the ad-supported web?
2. What type of personal data gets leaked while accessing websites or using apps?
3. How can companies understand user preferences?
4. How trackers and advertisers track data from each user?
5. Are there any privacy concerns in online advertising, and how can they be encountered?

6. Is there a GDPR compliance?

7. Are there alternatives to the current ad ecosystem?
8. To what extent are the popular web sites compliant with the relevant legal framework with respect to creating profiling of users for advertising purposes

1.4 Thesis Overview

This Thesis is organized as follows:

In Chapter 1, there is an introduction about the importance of data privacy, and how it is linked to data security. Despite recent advances in data privacy legislation, consumer's privacy is regularly invaded or compromised by companies or governments. That has led some to argue that consumers have already lost the privacy war.

In Chapter 2, there is a lengthy discussion about online advertising, a form of marketing and advertising which uses the services of internet to deliver promotional marketing messages to consumers.

In Chapter 3, the parameters of online advertising, are being analyzed, as well as the technologies used to provide such services. The online marketing business has become innovative, by creating a myriad of new opportunities for advertisers to reach potential customers.

In Chapter 4, an investigation about the collection of user's personal data is made, considering the limitless collection by first- or third-party sources. There is an overview about the underlying privacy risks and the solutions that may mitigate them.

In Chapter 5, future changes for securing personal data and avoiding the risks are fully discussed. By 2022, the most popular internet browsers will have phased out third-party cookies. The main reason about this significant change and its effects can be stated as a decisive fact.

In Chapter 6, for the justification of the conclusions that was made in each chapter, software and online tools was used, proving the respect or the breach of each service to the General Data Protection Regulation (GDPR)

Finally, some concluding remarks are given in Section 7

2. BACKGROUND

2.1 Online Advertisement

According to Breuer, Brettel & Engelen (2011), online advertising was initially opened with simple banners, that later evolved to new advertising models and online channels with enhanced customers and advertiser interaction. Online advertising has become one of the most effective ways to expand business reach, find new customers, and increase revenue streams. According to Statista [3] spend in the Online Advertising market is projected to reach US$345,948m in 2020. Online advertising has been rapidly growing in the past of years and it is expected to reach US$982.82 billion by 2025, at a CAGR of 21.6% over the forecast period 2020 - 2025. Online advertising, also known as digital or internet advertising, is a form of marketing and advertising which uses the Internet to deliver promotional marketing messages to consumers [6].

Online advertising started in 1994 when Hotwired, a web magazine, sold a banner ad to AT&T and displayed the ad on its webpage [1]. The first online banner was sold based on the number of “impressions”, which is an advertising way to cost per 1000 impressions of the advertisement and often being referred to as CPM (cost per mille impressions). In 1996, Procter & Gamble changed the norm, and with a negotiated deal with Yahoo started compensating with a “Cost-per-click” model, commonly known as “CPC”. With the CPC model, Yahoo was being paid only when a customer was clicking the advertisement and navigating to the Procter and Gamble website. In ad networks advertisers largely adopt the cost-per-click (CPC) or cost-per-acquisition (CPA) pricing models where they only pay when a certain goal is achieved [4].

As with advertising generally, a key feature of online advertising is that consumers are being “paid” with content and services to receive advertising messages [2]. With online advertising, companies can benefit from:

- An increase in brand awareness by displaying their products to potential customers outside their networks
- A better understanding of your target audience by leveraging analytics data being collected in the ad platform
- Understanding real time what content helps them achieve their business goals and increase the return of investment (ROI)

It is wide known that until 2022 Google Chrome’s will stop supporting third-party cookies. This kind of change brings out several changes in digital advertising compared to what advertisers have experienced since the mid-2000s and specifically in the way they operate.

There are many different types of online advertising and the most important are:

- **Display Advertising**

Digital display advertising is graphic advertising on Internet websites, apps or social media through banners or other advertising formats made of text, images, flash, video, and audio [7] as we may see in **Figure 3**. The basic goal of display advertising is to attract customers of a website, social media or any other digital platform and make them perform a specific action. Display ads are often using text-based, image or video visuals that encourage the customer to click-through the advertisement. Display Ads are placed on relevant third-party websites that are part of a network the form of a visual banner. Google Display network is one of the most famous networks and has the potential to reach more than 90% of total internet users on a network of over 2 million websites, video content websites, blogs and apps. Display ads as mentioned before, first appeared in 1994 and many things have changed till today.

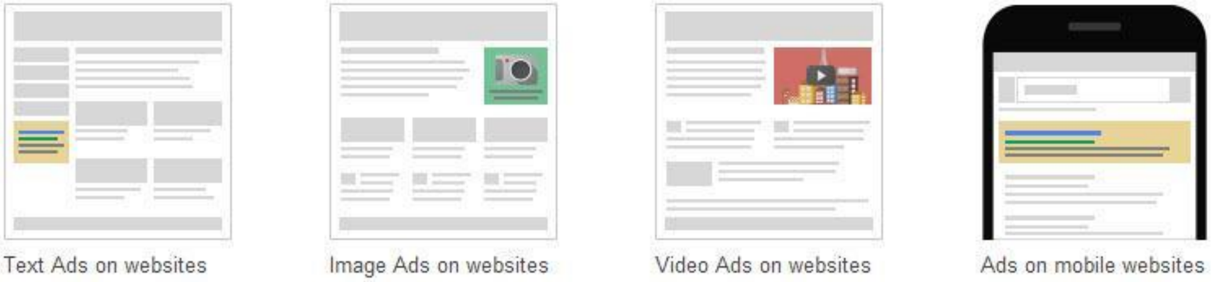


Figure 3 - Display Advertising Format

Despite the appearance that has changed, the frequency with which users interact with display advertising has been increased dramatically. Based on **Figure 4**, statistics show that an average user in the USA sees around 63 display ads per day.

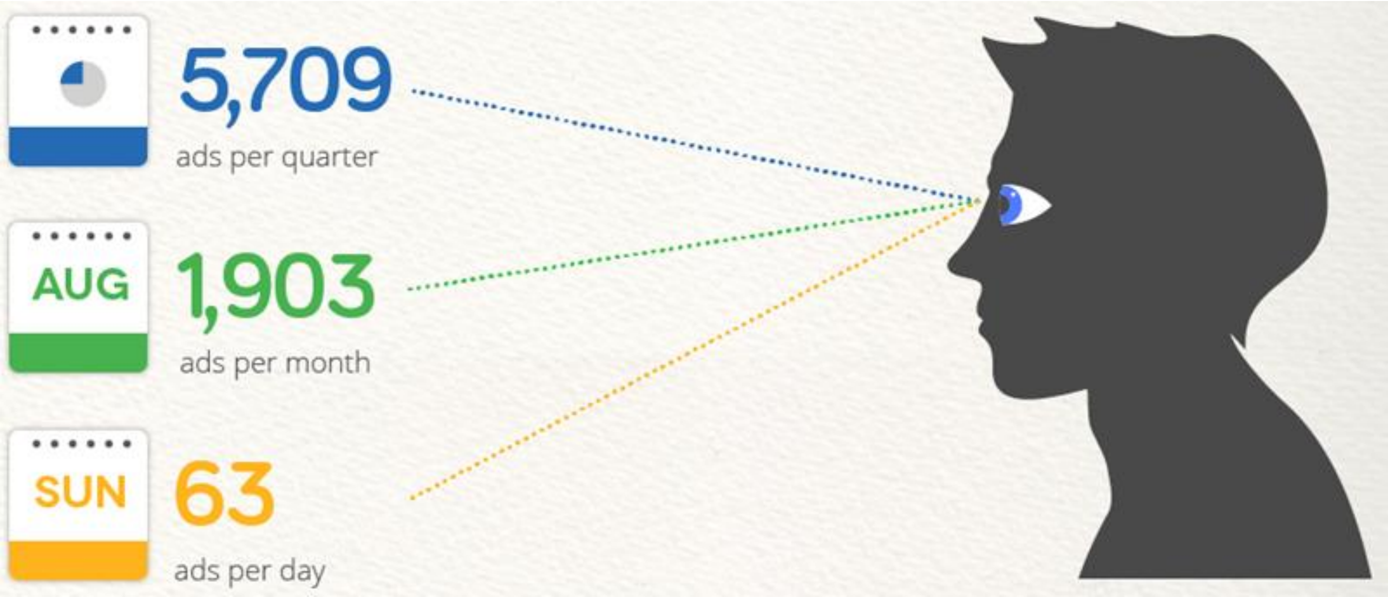
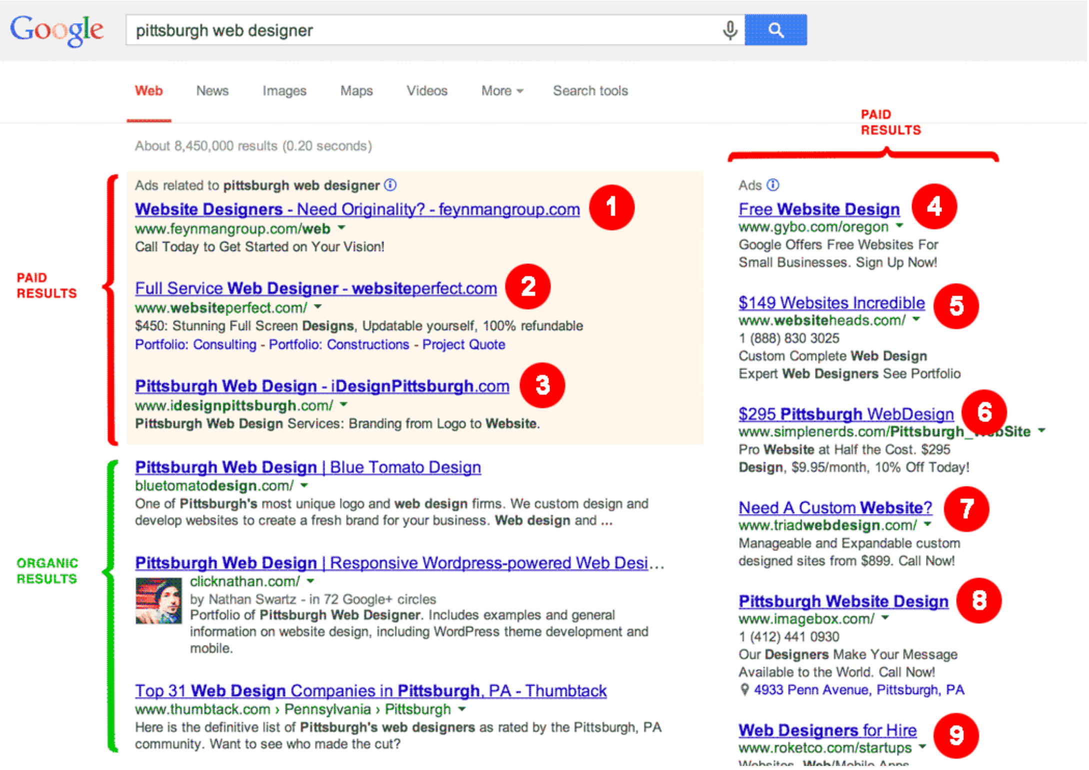


Figure 4 - Display advertising Us statistics for average impressions per day

An extension of GDPR, the upcoming privacy Regulation brings new more strict rules for every business that deals with any type of online communication service which uses collecting data technologies and participates in online advertising.

- **Paid Search Advertising**

Paid search is a form of digital marketing where search engines, such as Google and Bing allow advertisers to show ads on their search engine results pages (SERPs). Paid search ads appear when people are conducting a search online and will click to an ad that will be more relevant to what specifically they are looking for. Paid search results appear at the top of search results (SERP), and they feature a little box on the top left corner with the word 'Ad' as it is displayed in **Figure 5**. More specifically, once a user search is initiated, Ad platform searches through the pool of Ads advertisers and selects a seat of winners to display in the ad space of its search results page by a real time auction occurring in the background. The winners of the auction are chosen based on a set of different factors, such as the amount of keyword bids, the relevance of their keywords, quality of Landing page and the ad campaign itself.



pittsburgh web designer

About 8,450,000 results (0.20 seconds)

Ads related to **pittsburgh web designer**

Website Designers - Need Originality? - feynmangroup.com
www.feynmangroup.com/**web**
Call Today to Get Started on Your Vision!

Full Service **Web Designer** - **website**perfect.com
www.**website**perfect.com/
$450: Stunning Full Screen **Designs**, Updatable yourself, 100% refundable
Portfolio: Consulting - Portfolio: Constructions - Project Quote

Pittsburgh Web Design - i**DesignPittsburgh**.com
www.i**designpittsburgh**.com/
Pittsburgh Web Design Services: Branding from Logo to **Website**.

Pittsburgh Web Design | Blue Tomato Design
bluetomato**design**.com/
One of **Pittsburgh's** most unique logo and **web design** firms. We custom design and develop websites to create a fresh brand for your business. **Web design** and ...

Pittsburgh Web Design | Responsive Wordpress-powered Web Desi…
clicknathan.com/
by Nathan Swartz - in 72 Google+ circles
Portfolio of **Pittsburgh Web Designer**. Includes examples and general information on website design, including WordPress theme development and mobile.

Top 31 **Web Design** Companies in **Pittsburgh**, PA - Thumbtack
www.thumbtack.com › Pennsylvania › Pittsburgh
Here is the definitive list of **Pittsburgh's web designers** as rated by the Pittsburgh, PA community. Want to see who made the cut?

Ads

Free **Website Design**
www.gybo.com/oregon
Google Offers Free Websites For Small Businesses. Sign Up Now!

$149 Websites Incredible
www.**website**heads.com/
1 (888) 830 3025
Custom Complete **Web Design**
Expert **Web Designers** See Portfolio

$295 **Pittsburgh** WebDesign
www.simplenerds.com/**Pittsburgh_WebSite**
Pro **Website** at Half the Cost. $295 **Design**, $9.95/month, 10% Off Today!

Need A Custom **Website**?
www.triad**webdesign**.com/
Manageable and Expandable custom designed sites from $899. Call Now!

Pittsburgh Website Design
www.imagebox.com/
1 (412) 441 0930
Our **Designers** Make Your Message Available to the World. Call Now!
4933 Penn Avenue, Pittsburgh, PA

Web Designers for Hire
www.roketco.com/startups
Websites, **Web**/Mobile Apps

Figure 5- Paid Search results in Google

Paid search marketing affords businesses the opportunity to advertise within the sponsored listings of a search engine or a partner site by paying either each time their ad is clicked (pay per click) or less commonly, when their ad is displayed (CPM or cost per thousand) or when a phone contact is generated, which is 'pay per call' [8]. This makes it the most measurable marketing channel compared to the traditional forms of advertising.

Data is the fuel that powers the way that advertisers work and allows them to deliver targeted, relevant and personalized advertisements. Users are getting more and more concerned on the amount of data collected, where it is stored and how it is being used. New research shows that up to 30% of ad-blocker users feel that ads compromise their online privacy - with 2 in 3 worrying how companies are (or could) use their personal data [13]. In the upcoming years, with the

GDPR regulations and the restriction of third-party cookies collection, there will be significant shifts that paid search advertising works.

- **Social media Advertising**

Social media advertising is a form of online advertising that focuses on social media platforms to reach new customers, build brand awareness, increase sales and drive traffic to websites. According to [11] marketers are spending more on social media advertising; with social media ad spend amounting to more than $89 billion in 2019. Ad spending in the Social Media Advertising segment is projected to reach US$98,984m in 2020 and this is expected to show an annual growth rate of 6.5% till 2025 resulting in a market volume of US$135,384m by 2025 [12].

The most prominent social media platforms for advertising are Facebook, YouTube, Instagram, Twitter and LinkedIn. They all offer advertising options, but it is highly important to select the platform that will be more relevant to your audience. In general, all these social media platforms collect social data from users. Examples of this information are how users share, view and engage with content. Advertisers use them to target specific users based on parameters such as age, demographics and interests.

However, all these data processing activities can affect individuals' personal privacy rights. All brands that use social media for advertising collect a high volume of personal data, some of them are collected directly from the user (i.e. as a first party) or others using other sources (i.e. third party). These high volumes can be explained by:

- The high variety of social media platforms available,
- the high level of digital interaction between users and content,
- and all the available tools that social media uses (i.e. forums, podcasts, location-based services, instant messaging etc.) and make it a data rich tool.

And even though GDPR (**Figure 6**) has been in effect years now, to protect the privacy of the customers, many businesses are still not in full compliance. The positive impact of GDPR in social media is that users now have better control over, who gets access to their data and how they can use it. Advertisers can only collect, save, and use personal data for the specific purpose they have disclosed and must take measures to safeguard the data.

Figure 6- GDPR in social media

2.1.1 ONLINE ADVERTISEMENT

The main goal in online advertising, is the delivery of ads to users and the offer of having a large scale of options. Also, an absolute target is the financial growth of the players related to the procedures of displaying specific ads. This procedure is not centralized, and it needs a contribution of some components, as shown in **Figure 7**. According to [22], the first component is named demand-side platform (DSP) and it is a chunk of software used to purchase advertising, like banners, social and search ads in an automated way. It is mainly used by brands or representatives and allows advertisers to buy space, across a range of publisher sites, while helping them target individual user behavior, action, demographic, location, or any online activity. This ad space cost is set by publishers through ad exchanges, and DSPs evaluate these exchanges in order to automatically decide which impressions make sense for an advertiser to buy.

J. Estrada-Jiménez et al. / Computer Communications 100 (2017) 32–51

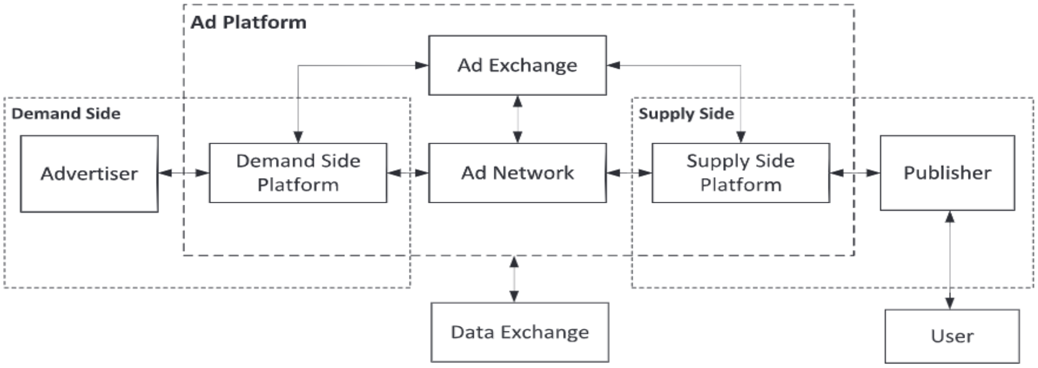


Figure 7- Basic Components of Online Advertising

The publisher, also called a supply source, is the only one, who can provide the opportunity and inventory which lets advertisers place ads in their domain. Specifically, it means that a publisher can be a website or an apian advertiser ,or the demand source, could also be an application such as a mobile game ,any online platform or a company who wants to deliver a message to a crowd massively. The publisher is the place that this message can be displayed, with a charge upon agreement between the two sides. There is also a chance to be an advertiser and a publisher at the same time. If an app is displaying its own advertisements while also hosting ads of other people, it is simultaneously an advertiser and a publisher.

At this point, a consideration about a possible connection of publishers and advertisers should be made. Sometimes, an advertiser does not have the ability to be informed about the existence of a publisher with an interesting audience or the vice versa. For example, imagine a person who sells his supercar at a reasonable price, but he lives in a small island with a few citizens. The solution for matching publishers and advertisers is the Advertising network.

Advertising networks or Ad networks, aggregate ad indexes from supply sources and pair them with demand sources searching for ad slots. The environment can either be just for a mobile network with participation of apps publishers and developers or for a web network or for both. Most ad networks offer a platform where advertisers can sign in and manage their campaign by setting parameters like the cost, they are willing to pay, the country and on what subject they are interested in. They can also have a view on statistical data and the success rate of their campaigns.

A categorization of online ad networks can be made into three types, depending on business strategies:

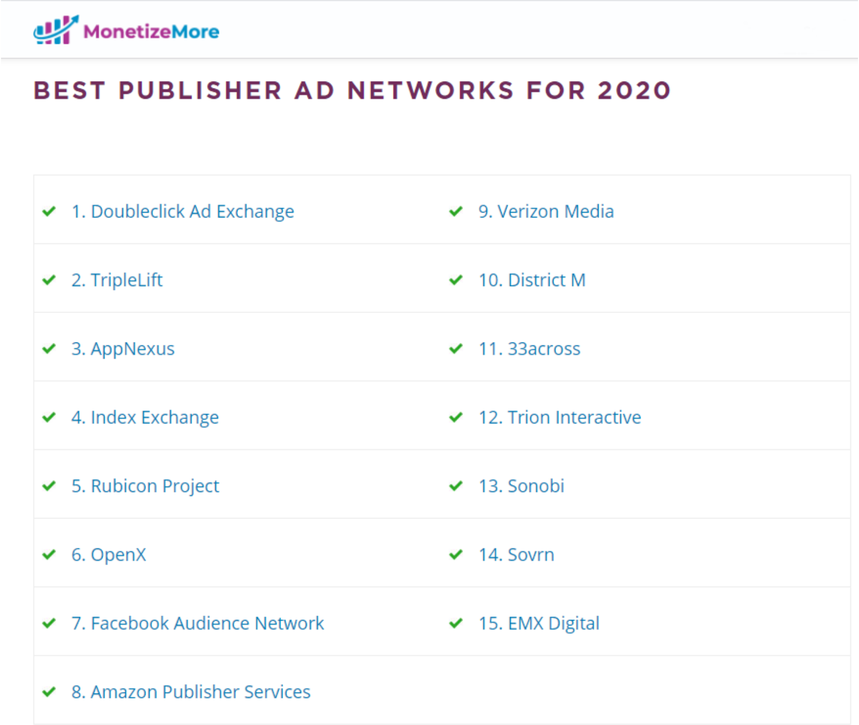
1. **Vertical networks:** There is full transparency between publishers and advertisers, as they are informed where exactly their ads will run. The promotion takes place at market prices with high quality of data and are mostly used by brand marketers. The financial model is mostly revenue share. Vertical Networks offer Run-Of-Site (ROS) advertising across specific Channels where a banner or other advertisement content can appear on any page or space of a specific website or they offer site-wide advertising options.

2. **Blind networks:** On this type of networks, a special price for advertising is offered to other marketers with exchange of other buyers who do not have restrictions or requests about the placement or value of their ads. The network usually runs campaigns as Run-Of-Network (RON) where ads may appear on any pages or spots of many sites. They achieve the low pricing through massive purchases of typically unsold and less valuable space, combined with ad targeting technology.

3. **Targeted networks:** The most trendy type, called next generation ad networks or ad networks 2.0. They mainly focus on specific targeting technologies such as behavioral or contextual, created in an ad server. They are also using information

from social graphs using connections in social networks and targeting network neighbors as stated in [26]. Targeted networks focus on using consumer clickstream data to increase the value of the purchased space.

The top-15 Ad networks as listed by Monetizemore for 2020 can be seen in **Figure 8:**



MonetizeMore

BEST PUBLISHER AD NETWORKS FOR 2020

✓ 1. Doubleclick Ad Exchange
✓ 2. TripleLift
✓ 3. AppNexus
✓ 4. Index Exchange
✓ 5. Rubicon Project
✓ 6. OpenX
✓ 7. Facebook Audience Network
✓ 8. Amazon Publisher Services
✓ 9. Verizon Media
✓ 10. District M
✓ 11. 33across
✓ 12. Trion Interactive
✓ 13. Sonobi
✓ 14. Sovrn
✓ 15. EMX Digital

Figure 8 - Top-15 Ad networks for 2020

Another necessary component is the supply-side platform. Supply-side platform (SSP) is a chunk of code or a software that allows a publisher to sell digital ads, through automated auctions. It is called like that, as people that use a SSP, supply the space for the ad to appear on, so it is used by publishers that produce content on a platform. This procedure has many things in common with DSP, but it is about the publisher side. SSPs allow a publisher to connect with a range of buyers by connecting their space to DSPs via an ad exchange. Publishers can set minimum price thresholds, create profitable agreements and define their own rules where advertisers or buyers can purchase their space.

Finally, the data management platform (DMP) is a chunk of software that collects first, second, and third-party data and then distributes them in a specific way, in order to help marketers gain the most profitable results. It's usually used for target grouping and for sorting audiences to improve marketing results. In practice it is used to collect and manage web browser cookies from user’s selection like interests or behaviors, and then send this information to DSP in order to make the final decision.

The next step is the data collection related to a particular topic, and the organization of them, by DMP. After the advertiser or the buyer decides what is the exact target group, an ad campaign will start, and transfer this amount of information into a DSP, which will connect to an SSP. The SSP offers a variety of choices to reach the desired audience at different prices, and the DSP will automatically figure out what is the most appropriate purchase to reach the specific group of people based on the information for the campaign contained in the DMP, at the best price. At the time DSP makes the buyer satisfied, with parameters like the price and space offered by the publisher through the SSP, the exchange takes place and the ad is purchased and delivered. This procedure can be seen in **Figure 9.**

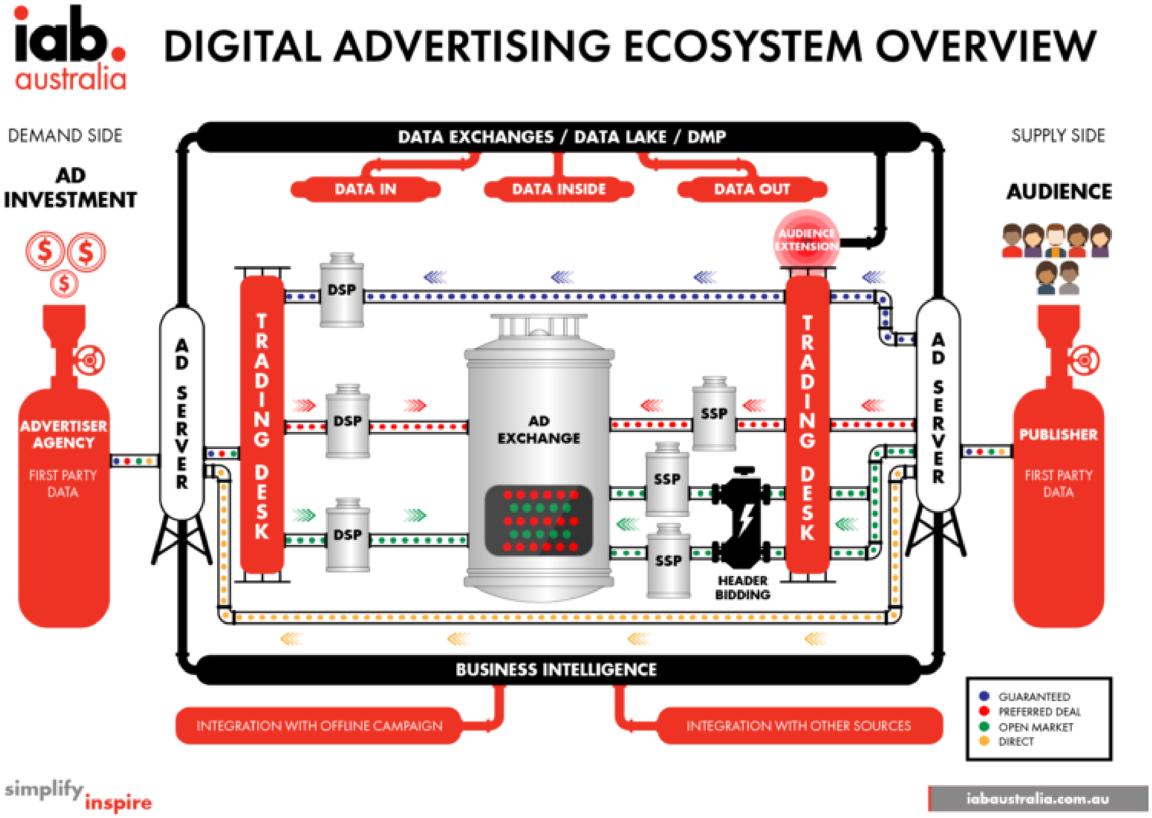


Figure 9 - Interactions between DSP, SSP and DMP

2.2 Personalized Advertising

Personalized advertising, also known as internet-based advertising is a powerful tool that improves advertising relevance for users and increases ROI for advertisers [16]. It is an evolving genre of advertising that creates personalized user experience, and therefore aims for better Customer Experience (CX), higher return of sales, broader reach and engagement Personalized advertising uses the type of cookies, that we will analyze in another section of this thesis, and more specifically the collection of personal's user data to target users and share more relevant advertising content and creates a better experience both for customers and advertisers. Because of the nature of personalized ads and the privacy issues associated with customer targeting, there are already some policy standards that all the known publishers such as Google, feature when they make use of personalized advertising functionalities. The need of policy standards occurs by the fact that customers are having concerns with the misuse of their sensitive personal information. As a result, when they interact with a personalized advertisement, they are often skeptical about how their personal data are flowing, which results in annoyance and decline in purchase intent because of the manipulation they feel.

According to the RSA Data Privacy and Security survey [20]:

- Data Privacy expectations are cultural. In European countries that GDPR came into effect, users' data privacy complaints have increased.
- Consumers fear losing control over their financial data (78%), security information (75%), identity information (70%), medical information (61%).
- 64% of US and 72% of UK respondents feel inclined to blame the company, in case of data leaks (or a hack).
- Customers are now more worried about the constant location and behavior tracking. They are concerned about their personal information being shared with third (or fourth) parties that could violate their privacy.

Examples of policy standards that publishers need to comply with are, European Union user consent and Google Ads policies for data collection and use [17], and they represent the commitment to uphold the highest policy standards for both parties. These policy standards define the way that advertisers can perform user data collection and how they can use it for personalized advertising. For example, Google, one of the most known advertisement platforms, won't allow any advertiser to make use of remarketing lists and similar audiences lists, collected with the personal data collection of customers, without asking for user's content.

On the other hand, there are some well-defined policy principles which protect users from an unorthodox ad targeting. Advertisers have to respect sensitive interest categories and not use them to achieve a better ad targeting or to promote advertiser's services or products. The next four principles include the main sensitive internet categories.

1. **Personal hardships**: Ads should not exploit the difficulties of users and for this reason we don't allow categories related to personal hardships. Such personal hardships include traumatic personal experiences, personal failings, treatments or health conditions.

2. **Identity and belief**: We want ads to reflect the user's interests but on the other hand we respect the identity and belief of users. Ads should provide a positive experience and not create stigmatization or prejudices. So, ads which are related to sexual orientation, political content, religious belief or race and ethnicity must be avoided.

3. **Sexual interests**: A very sensitive category is the sexuality of a person. Sexual interests are inherently private and must be respected. For this reason, one of the prohibited categories are those that reflect the sexuality and preferences of the user. For example, products for birth control and non-family safe content.

4. **Access to opportunities**: Personalized ads should reflect the interests of someone and not to specific audiences. Access to social and economic opportunities is fundamental for individual well-being, social status and quality of life. Some of the prohibited categories are ads for jobs, for credit cards, loans and individual houses for sale or rental.

2.2.1 ONLINE ADVERTISEMENT

2.2.1.1 ON-SITE BANNER

Display advertising or banner advertising uses banner ads (graphic or text), that appear in specifically designated areas of a website or social media platform (Facebook, Instagram, Twitter etc.). Digital display advertising comes in a variety of different forms, but at its core, it revolves around the same principle. [27]

Figure 10 – First Banner ad

The first online banner appeared in 1994 (**Figure 10**), it was the text “Have you ever clicked your mouse right HERE? YOU WILL” and was published on HotWired.com for a campaign by AT&T. The banner stayed there for four months and 44% of those who saw the ad click on it. The number is huge if someone considers that the average click-through rate on display advertising today is about 2%.

A web banner or banner ad is a form of advertising on the World Wide Web delivered by an ad server. This form of online advertising entails embedding an advertisement into a web page (**Figure 11**). It is intended to attract traffic to a website by linking to the website of the advertiser. In many cases, banners are delivered by a central ad server. [28]

Banners vary depending on the type and their size. The type can be any media element that can travel the message to the targeted audience such as video, image or HTML5 animation. Apart from that we come across banners in various sizes and shapes.

SAVE UP TO
40%
before 16th October
Where PR, Marketing & Political Campaign Pros Collide
REGISTER NOW
CAMPAIGNS & MARKETING SUMMIT
CONVERGENCE OF MARKETING, TECHNOLOGY & CAMPAIGNS

Figure 11 – Variety of Banners

The way that an online banner works is very similar to a physical banner outside a shop which would make you enter the shop for a special discount or a new product, an online banner has a single goal, to generate traffic to the website that is being advertised. The main roles of banner ad are the following:

1. **Increasing customer traffic**: The banner ads encourage the visitors to enter a specific site by clicking on it. By this action of the user the ad is paid in many ways such as CTRs, time spent on the site, impressions etc. A successful banner is the one that makes a good targeting of the users and as a result a big percentage of them click on it.

2. **Sell a product**: Banner ads encourage the user to buy advertised products. So, for being effective a banner should be in the right position and the most important is to respond to the visitor's wishes. A good example here is Airbnb which is a success and reach of Google display ads. Airbnb started expanding some years ago by searching relevant websites with its content to target specific audiences and put ads out there.

3. **Grab the customer's attention**: A banner ad should be attractive as the main goal is to capture someone's attention. For this reason, the advertiser needs to study the target audience market and as a result to create a banner that fits in the specific area. It's up to the advertiser if he needs an animated bouncy ad or a classy font on a neat background.

4. **Announcing discounts and offers**: Another way for business to increase revenue is by making discounts and offers. Let's say it's Black Friday and a shop makes some good offers for a variety of products. The best way to spread the news and inform the customers is to design Black Friday themed banner ads and publish them out.

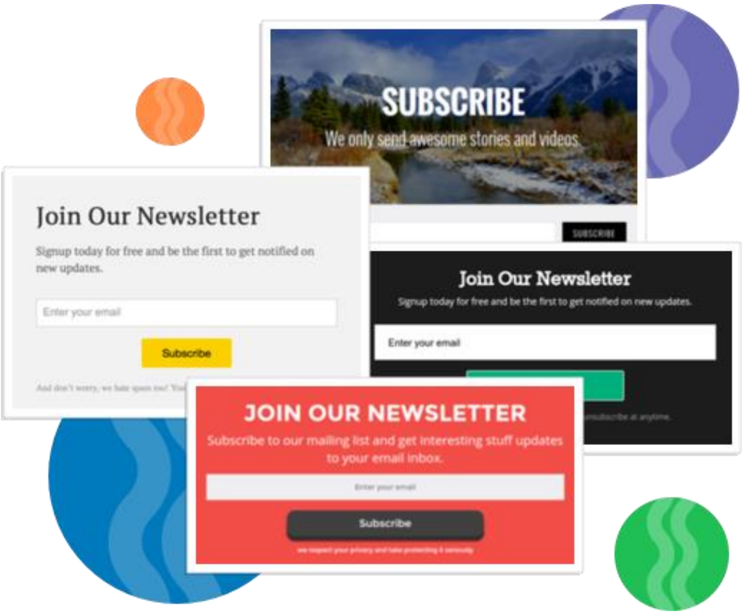
Today, on-site banners are based on real-time bidding (RTB) mechanism, also known as programmatic bidding, that allows advertisers to buy real-time ad impressions while the banner ad is loading. Scientifically, the further demand for automation, integration and optimization in RTB opens new research opportunities in the fields such as Information Retrieval (IR), Data Mining (DM), Machine Learning (ML), and Economics [44]. In addition to that, another technology that is becoming known is **t**argeted banner advertising**.** As the name implies, targeted means creating focused banner Ads to reach only a specified audience, by selecting attributes such as demographics, interests and audience behavior as a criterion. For this type of advertising, the cost model that is being used is Cost per Mile (CPM) rather than Cost per Click (CPC).

2.2.1.2 POP UPS

A popup banner is a banner that appears suddenly in the foreground of the visual interface, usually in a clearly delimited window. Popup banners are called this because they “pop up” on the page, interrupting the user with a promotional message. They are usually used for offering discounts and for promoting new products or services, but also for prompting users to click on the banner for more information on a service or product. Popups can also be used for opt-in forms. [30]

There are many different types of Pop-up Ads that somebody can use. It is important to highlight that every pop-up type serves different purposes. Hence, choosing the right one is the key for its success. The most popular ones are the following:

- **Exit-intent pop-up**: This type of pop-up ads is triggered when a user tries to leave the website. Are ideals for a last offer to users who are ready to exit the website.
- **Time-based pop-up**: Pop-ups in this category triggered after a specific time that someone stays on a website. Are ideals because it is considered that the user to waste time on the website is interested in its content, so is the right time for an offer or to subscribe to the website's newsletter.
- **Scroll-triggered pop-ups**: In this category the pop-up ad appears when someone scrolls some percentage of the webpage. The case is pretty like a time-based pop-up.
- **Click-triggered pop-ups**: Pop-ups that are triggered when someone clicks on a link or a button. Perfect because the ad acts when the user is interested in a product or service.



We only send awesome stories and videos

Signup today for free and be the first to get notified on new updates.

Signup today for free and be the first to get notified on new updates.

Subscribe to our mailing list and get interesting stuff updates to your email inbox.

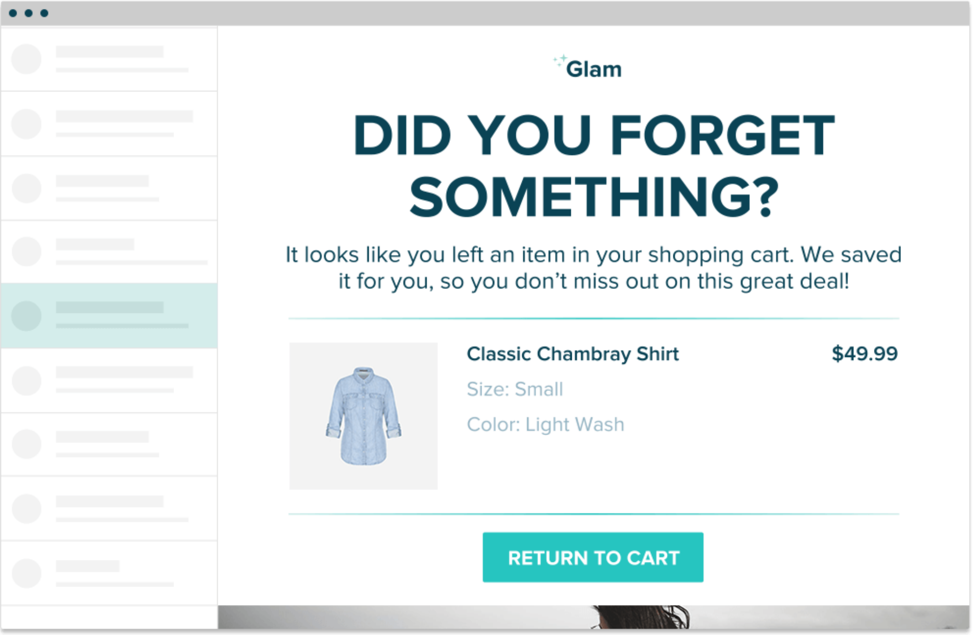
we respect your privacy and take protecting it seriously

Figure 12 – Example of Pop up ads

Recent research has shown that users feel more bothered from pop-up ads than any other type of online advertising. A percentage of 73% of internet users have automatically disapproved of pop-up ads, while most of them don't even see them as all major browsers have built-in pop-up blockers.

2.2.1.3 PERSONALIZED EMAILS

Whenever any person offers a gift to somebody else for a special day like birthday or anniversary, it really shows cherish and care. On the other hand, if a person offers a gift to his best friend that does not match his interest, it will make the sense that they don't really know each other so well. The same thing can be said for companies about personalized marketing. The term “personalization” has now become a very precious medium where marketing can make the audience feel as a unique unit who gets a personal treatment.



DID YOU FORGET SOMETHING?

It looks like you left an item in your shopping cart. We saved it for you, so you don’t miss out on this great deal!

Classic Chambray Shirt $49.99

Size: Small

Color: Light Wash

Figure 13 – Personalized Emails example

Email personalization (**Figure 13**) procedure is a marketing technique that uses personal information of subscribers in order to produce targeted emails. The handling of each user is individual, targeting to the customer’s needs and increases email marketing efficiency at a great level. Personalization is a dominant tool for email marketing, as it allows marketers to send relevant and targeted emails depending on the user’s data that they voluntarily transferred to marketers, at the time they subscribed to email newsletters. A variety of personalized email techniques are used today, such as mass emails with personalized subject lines, behavior-based trigger emails, reactivation email, personalized discounts emails on special days or celebrations. A detection of behavior patterns and discovery of correlations between customers can boost the techniques described before.

According to Send Pulse [19], a very successful multichannel marketing platform, the main reasons for any business to use personalized emails, are described below:

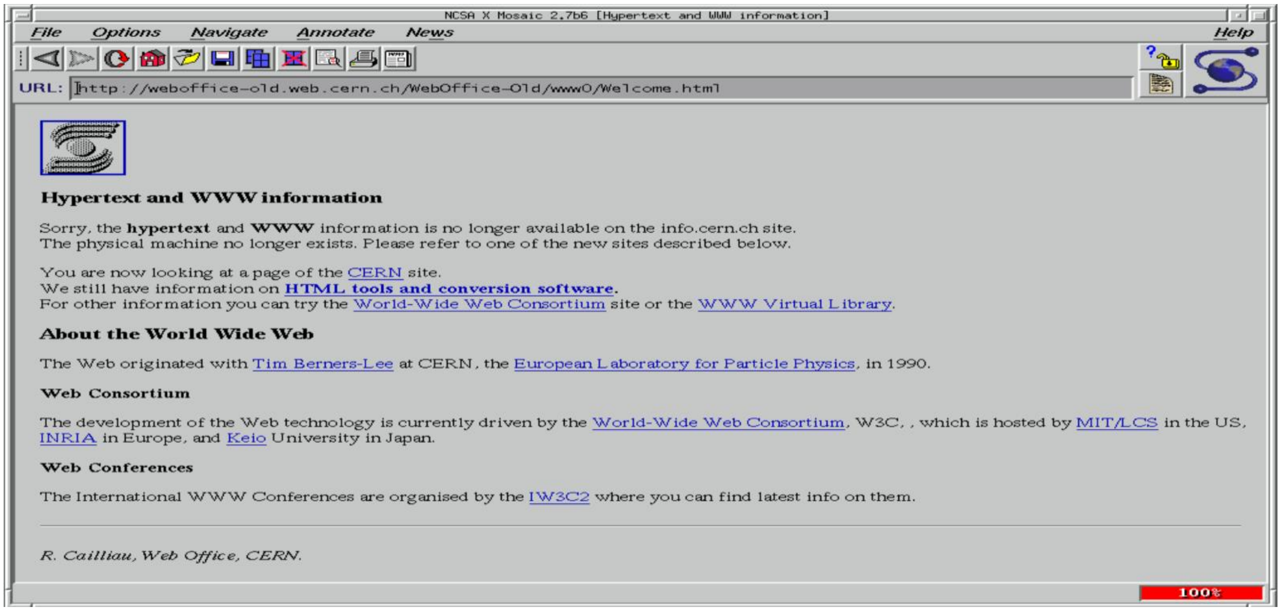
- Return on Investment (ROI) raise. Personalization of email campaigns at a top level, can increase the earnings up to 15%.
- Results in better engagement. According to Accenture, 91% of people are more likely to buy from companies that send accurate personalized emails. A survey showed that inviting people to an event using invitations that contains a picture of their city, the chances of successful visits are way high.
- Improves reputation. A high level of engagement that personalization offers to email marketing, contributes to fewer spam complaints and lower unsubscribe rates. It has a positive impact on emails deliverability and ensures a better sender reputation.
- Motivates users to share personal data. 83% of consumers are willing to share their data with brands if that enables personalization, according to Accenture study. In exchange, brands should be transparent and fair about the way they use this information by informing their audience at specific time periods.
- Open rate boost. The open rate is mainly determined by the subject line. Personalized subject lines lead to 50% more opens, and this technique is easy to implement. There should be a name field in the subscription form, where a {{name}} variable will be added to the subject line while creating an email campaign.
- Easy to create relevant content. Brands who have succeeded in collecting the appropriate amount of information about their audience, demonstrate their products and services more attractively, projecting them to the right audience. For example, a person who seems to watch the NBA, could be easily attracted by a special offer about basketball stuff.
- Meets subscriber’s expectations. When people subscribe and provide their personal data, they expect a personalized approach in return. Following that approach literally, will provide confidence and relief to the audience while in any other case provides the feeling of unreliability.

Marketers, after collecting enough information about their audience, are taking advantage of personalization in order to maximize their influence into their customers. As reported in [18], the plethora of information in a visual environment lowers the processing capability of the human brain, causing selective focus of human attention to specific points of visual fields. In psychological science, researchers have shown that small numbers are associated with faster left-hand responses, and larger numbers with faster right-hand responses, the phenomenon which is also termed as SNARC effect (Spatial-Numerical Association of Response Codes; Dehaene, Bossini, and Giraux 1993).Exploiting this phenomenon ,marketers are confident that consumers who scan through the email newsletter will pay more visual attention to 'left' region as compared to 'right' region, thereby triggering their responses more to the links placed in left region. Combining these two concepts together, in an email newsletter between the two divisions of left region into top-left and bottom-left, they suppose that the links placed in the top-left region will trigger a lot of click responses. However, as consumers' attention shifts to the right region of email newsletter, their responses will not be totally clear. As for the right side, between the two divisions of the top-right and bottom-right, there is an expectation that the links placed in the bottom-right, are likely to be more responsive. Overall, links placed in the left region of an email newsletter are supposed to be more clicked than those placed in the right region.

2.3 Cookies and HTTP

Back in 1993, where the World Wide Web (WWW) was officially announced in public. The first web browser supported, was Mosaic and it could integrate multimedia such as text and graphics. It was developed at the National Center for Supercomputing Applications (NCSA) at University of Illinois's. Mosaic (**Figure 14**) was not able to support a state mechanism. Applications' demand for an environment that can support state, led to the first publicly available version of the Netscape Navigator web browser in September 1994 which supported a stateful mechanism called cookies. The man responsible for the development and mainly for the success of these web browsers is Lou Montulli. According to Montulli, cookies are named after the computer science term "magic cookie,", which is "something passed between routines or programs that enables sender and receiver to have a silent interaction." Since then, a lot of web browsers have been developed, and the most popular of them are Mozilla Firefox, Google Chrome, Safari, Opera, Internet Explorer, all of them based on the stateful plan of Lou Montulli.

Clearly, cookies are small text files that contain important pieces of information about online users. Whenever a user visits a new website, the browser creates cookies and saves them locally at his personal computer. If the user returns to the website at any time, cookies will help it recognize various things, such as what content the user viewed, and which pages had been accessed. The whole Internet would be messy and chaotic without cookies. We couldn't have the chance to add products to online shopping carts, we would have to type our login credentials every time we visit a website or application, and we would have to change the default language of multilingual websites on each of our visits. It is the only way for web-based applications to maintain status in the stateless HTTP protocol.


NCSA X Mosaic 2.7b6 [Hypertext and WWW information]
File
Options
Navigate
Annotate
News
Help
URL: http://weboffice-old.web.cern.ch/WebOffice-Old/www0/Welcome.html
100%

Hypertext and WWW information

Sorry, the **hypertext** and **WWW** information is no longer available on the info.cern.ch site.
The physical machine no longer exists. Please refer to one of the new sites described below.

You are now looking at a page of the CERN site.
We still have information on **HTML tools and conversion software**.
For other information you can try the World-Wide Web Consortium site or the WWW Virtual Library.

About the World Wide Web

The Web originated with Tim Berners-Lee at CERN, the European Laboratory for Particle Physics, in 1990.

Web Consortium

The development of the Web technology is currently driven by the World-Wide Web Consortium, W3C, , which is hosted by MIT/LCS in the US, INRIA in Europe, and Keio University in Japan.

Web Conferences

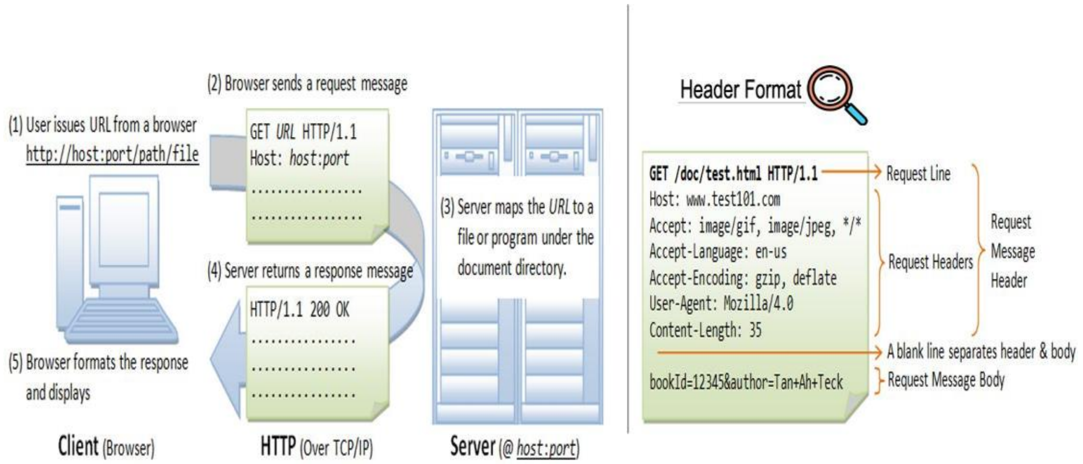
The International WWW Conferences are organised by the IW3C2 where you can find latest info on them.

R. Cailliau, Web Office, CERN.

Figure 14 – Mosaic Web Browser

The Hypertext Transfer Protocol as described in [21] provides the foundation for the Web, through the cookies which are a substance to HTTP. Every time a user clicks on a link in a web browser, the browser typically connects to the web server identified by the Uniform Resource Locator (URL) embedded in the link and sends it a request message, to which the server sends a response message. Then, after receiving the response, the browser disconnects from the server. Because the client makes a new connection for each request, the server treats each request as though it were the first one it had received from that client. Each request is treated completely independently of any previous one and that’s the reason why we consider it as “stateless:”. This situation makes it easier to build web browsers and servers, but it makes some web applications harder to write. For example, it would have been much harder to create the now-ubiquitous web shopping applications if they could not keep track of what’s in your shopping basket. HTTP requests -responses include three parts, as shown in **Figure 15**:

- (1) a request / response line.
- (2) request /response headers, which provide meta-information.
- (3) the request / response entity itself. The header meta-information provides both control information for HTTP and information about the entity being transferred. Information about cookies gets conveyed in such headers.


(1) User issues URL from a browser
http://host:port/path/file
(2) Browser sends a request message
GET URL HTTP/1.1
Host: host:port
(3) Server maps the URL to a file or program under the document directory.
(4) Server returns a response message
HTTP/1.1 200 OK
(5) Browser formats the response and displays
Client (Browser)
HTTP (Over TCP/IP)
Server (@ host:port)
Header Format
Request Line
Request Headers
Request Message Header
A blank line separates header & body
Request Message Body

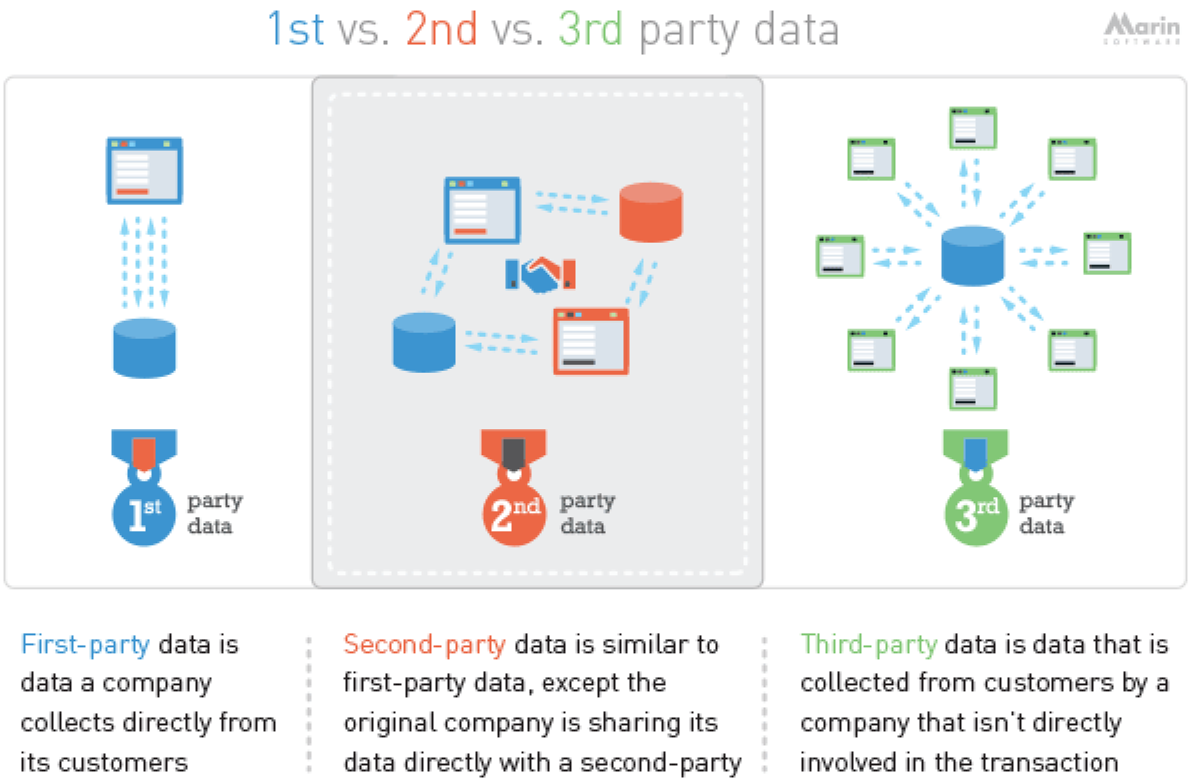
```
GET /doc/test.html HTTP/1.1
Host: www.test101.com
Accept: image/gif, image/jpeg, */*
Accept-Language: en-us
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0
Content-Length: 35

bookId=12345&author=Tan+Ah+Teck
```

Figure 15 – HTTP Request format

As for the response part, the server sends random information, named “cookie,” in a specific format of a response header. This random information could be anything like a user identifier, a database key, information about the browser of a user or whatever the server needs in order to maintain a current status. From the client side, there is a return of the cookie information in the appropriate format in a Cookie header, located in one of the requests made to the same server. The server may choose to include a new cookie with its responses, which would replace the old one. There is also an agreement between a server and client as the server trusts the client in order to save its state and return it on the next visit. About the storage of cookies, the browser is responsible to store locally only those cookies it receives from a server it has visited. Although, sometimes the browser has the privilege to visit servers without letting users know about it and store cookies locally on user’s computer (i.e. subscription details, user actions, CRM, feedback or social media)

2.3.1 1ST, 2ND, 3RD PARTY DATA

Different situations are demanding a variety of data types. Companies have some options between three main types of audience data, first party, second party and third party, as presented in **Figure 16**. Each type has unique benefits but also some disadvantages.



First-party data is data a company collects directly from its customers

Second-party data is similar to first-party data, except the original company is sharing its data directly with a second-party

Third-party data is data that is collected from customers by a company that isn't directly involved in the transaction

Figure 16 – First party, Second party, and Third-Party Data Definitions

Valuable information that administrators can collect from their own sources, is called first party data. As stated in [10], it could be any information related with users or customers from both online and offline sources, such as:

- Behavioral data from online interactions like clicks, views, comments, purchases that can be collected by using website tags. This information provides wisdom about visitors of the online properties, regarding content or products they're interested in, letting administrators to adjust their plans and strategies to the audience's preferences.

- Subscription data: Especially for e-blogs or content sites, subscribers are an integral part of the business model. In addition, companies that sell other products but have an email list, subscriptions are a crucial part of a marketing effort. An analysis on the subscribers of a content, can provide many information about the types of people who are interested in a company and the content.

- Social data: The collection of information about who follows somebody on social media and who likes, shares or comments particular posts, can provide useful details of who your audience is. A mining of information about the audience's interests and preferences can easily be done by evaluating their profiles, posts and pages they follow. The actual content of the messages and comments can also provide information about what people think about a content or products.

- In-store purchase data: Apart from online data about sales, in-store purchases can be merged, for a realistic approach. The mixture of offline sales data with e-commerce sales data, can grant administrators or brands a better understanding of the purchase behaviors of their customers. It provides insights about products that are most popular and what is the trendiest type of them.

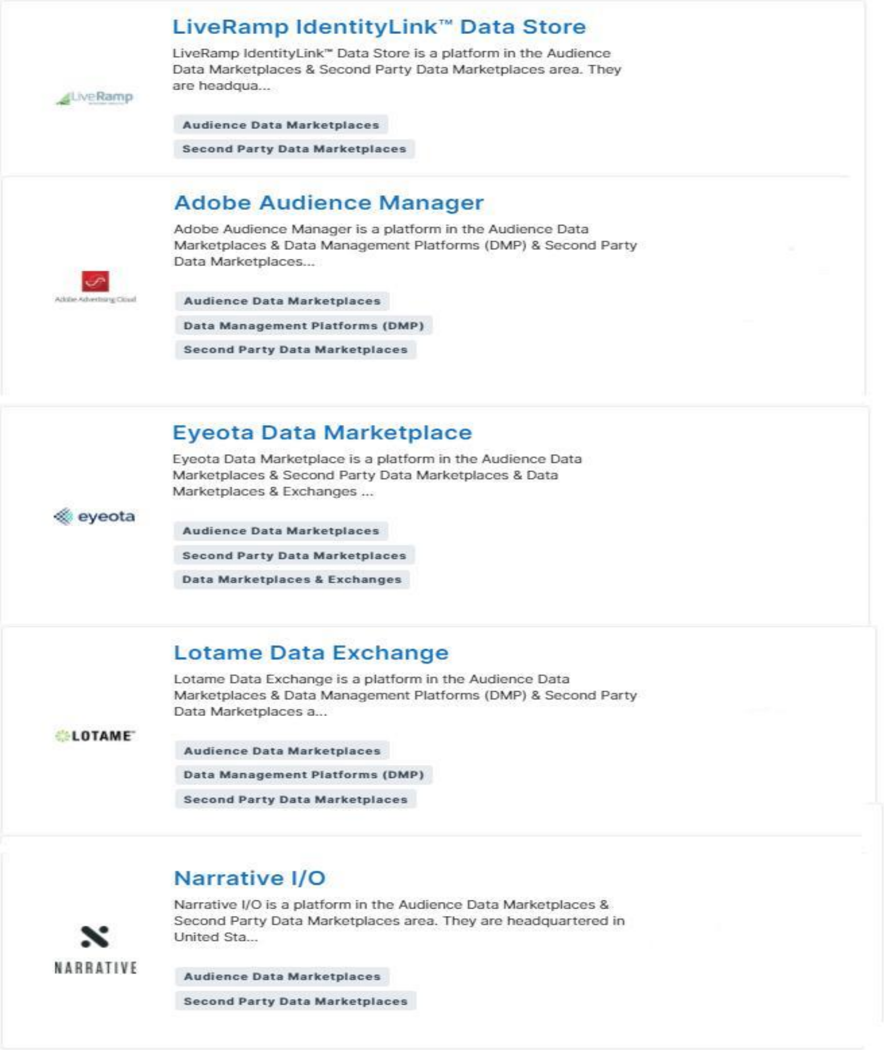
- Customer feedback data: Asking customers for feedback through survey papers, online forms, reviews, emails, phone calls or comments on the website, can inform owners about the pros and cons of each operation. Both positive and negative feedback is valuable, as they reflect if something is working well and what has to be improved as it is collected from different groups of customers.

Clearly, first party data arrive straight from the audience, and it is mainly considered as a very valuable source. Except from its value, they are also served at no cost, making it cost-effective. Mainly, this type of data can provide personalized experience and collect high-quality data with the confidence of compliance on privacy rules as all user consents will be made through the owner's application or website. Their usage is also important for dividing and categorizing users into groups while administrators can focus on managing their content to the needs of their audience, in order to maximize the effectiveness. Another benefit is that they are supported by all browsers and can be blocked or deleted by the user. The collection and management of those data is way easy, with the use of a data management platform.

Every company knows that first party data is a true wealth and should be a priority for brands who are willing to monetize and improve their Return on Investment (FrF) ratio during their marketing efforts. Return on Investment (ROI) is a performance score used to express the efficiency of an investment. It tries to directly measure the amount of return of an individual investment, relative to the investment's cost. In order to calculate ROI, the benefit of an investment is divided by the primary cost. The result is expressed as a percentage.

On the other hand, second-party data is mainly first-party data from another owner or administrator who offers them for a charge, which have been collected directly from their own audience. Like first party data, they could be information from activity on websites, social media, online purchase history, surveys or mobile application data. In order to obtain the second-party data, the interested person must have a direct interaction with the company or organization who collected the data. After this procedure, the buyer and seller agree on terms of a simple contract defining their own details and then exchange the data. This method provides flexibility and adaptability, while ensuring the collection of high-quality data and at the same time offering relationships with other companies. The sellers can be easily found with a simple search in a second party data marketplace, where thousands of types of data are being offered.

Second party data marketplace is an online platform where organizations or content owners upload entries regarding a large amount of data collected by them. Any interested company or person can request to buy any set of available data instead of searching the proper type by his own. In that environment, there is also the option for somebody who wishes to monetize his valuable collected data, by finding buyers and contacting them for a deal that meets both parties' requirements. Some popular brands offering their second party data in an online marketplace named database, can be seen in **Figure 17.**



LiveRamp IdentityLink™ Data Store

LiveRamp IdentityLink™ Data Store is a platform in the Audience Data Marketplaces & Second Party Data Marketplaces area. They are headqua...

Audience Data Marketplaces

Second Party Data Marketplaces

Adobe Audience Manager

Adobe Audience Manager is a platform in the Audience Data Marketplaces & Data Management Platforms (DMP) & Second Party Data Marketplaces...

Audience Data Marketplaces

Data Management Platforms (DMP)

Second Party Data Marketplaces

Eyeota Data Marketplace

Eyeota Data Marketplace is a platform in the Audience Data Marketplaces & Second Party Data Marketplaces & Data Marketplaces & Exchanges ...

Audience Data Marketplaces

Second Party Data Marketplaces

Data Marketplaces & Exchanges

Lotame Data Exchange

Lotame Data Exchange is a platform in the Audience Data Marketplaces & Data Management Platforms (DMP) & Second Party Data Marketplaces a...

Audience Data Marketplaces

Data Management Platforms (DMP)

Second Party Data Marketplaces

Narrative I/O

Narrative I/O is a platform in the Audience Data Marketplaces & Second Party Data Marketplaces area. They are headquartered in United Sta...

Audience Data Marketplaces

Second Party Data Marketplaces

Figure 17 – List of 2nd data marketplaces according to datarade.ai

Second party data is a very powerful tool for marketers, but it is mainly accessible when there is no competition between the two parties involved. Second-party data offer the chance to target prospective customers who are likely to be interested in the advertisers' products or content. There should be a trust and transparency about the source in order to make sure that the provided data are valuable and accurate to the buyer's needs, in order to help him target his ads in the maximum degree.

In the very early era of the web, content was produced and stored by a single person, or a group of persons belonging to a team. They could manage their content depending on their customers' needs, as reflected by their personal data. This type of data has been named first party as mentioned earlier in this section. Marketing players perceived that using and analysing customers' data can maximize their profit and decided to start buying data collected from other companies, calling them second party data. As the web evolved and was accessible from many new sources also known as Internet of Things ,marketers recognised a chance to collect an extensive amount of data ,by buying a mixture of users' data that have been exchanged between a very large amount of sources. This type of data, known as third party data, is collected from a variety of websites or platforms and then aggregated by DMP. It is the most trendy and worthwhile data, and at the same time the main material for brands or marketers. More specifically, third-party data is information collected by an entity that doesn't have a direct relationship with consumers. By aggregating data from different sources, DMPs are able to create audience profiles containing information on users' web interactions and behaviours, which are then used to categorize them into specific consumer types, called segment data. Third party data are purchased and sold automatically using programmatic advertising, which means it happens rapidly and on a large scale.

There are two common ways which third party data are being collected. The first method is known as cookies matching or cookies syncing and it is described on section 3.2.Briefly, interacted parties are exchanging their cookies , in order to identify common users, collect or update the data about them and handle them with particular ways. The second method can be achieved by raw data [15], a type of data that have not been processed by any machine or a human, not properly structured. Mainly it is some part of code, like user cookie for example, which doesn't bring much information, but when this data is integrated with appropriate user profiles, it is really helpful for marketers or business analysts. The integration is possible within the data provider, by using a DMP combined with AI algorithms in order to match raw data with 3rd party data profiles available on the platform.

The downside is that these data are not exclusive and as they are publicly available, some competitors could have access to the same data. On the other hand, they provide information about users that would never be accessible with any other way, and it also happens on a large scale. Also, the combination of first- and third-party data is a very powerful option as they help marketers target new potential customers and learn more about the people that are interested in their products or services. They are also helpful for demographic, behavioral and contextual targeting as they are massive and from distributed sources. In general, the main difference between 1st, 2nd and 3rd party data is the trade-off between quality and reach, as shown in **Figure 18.**

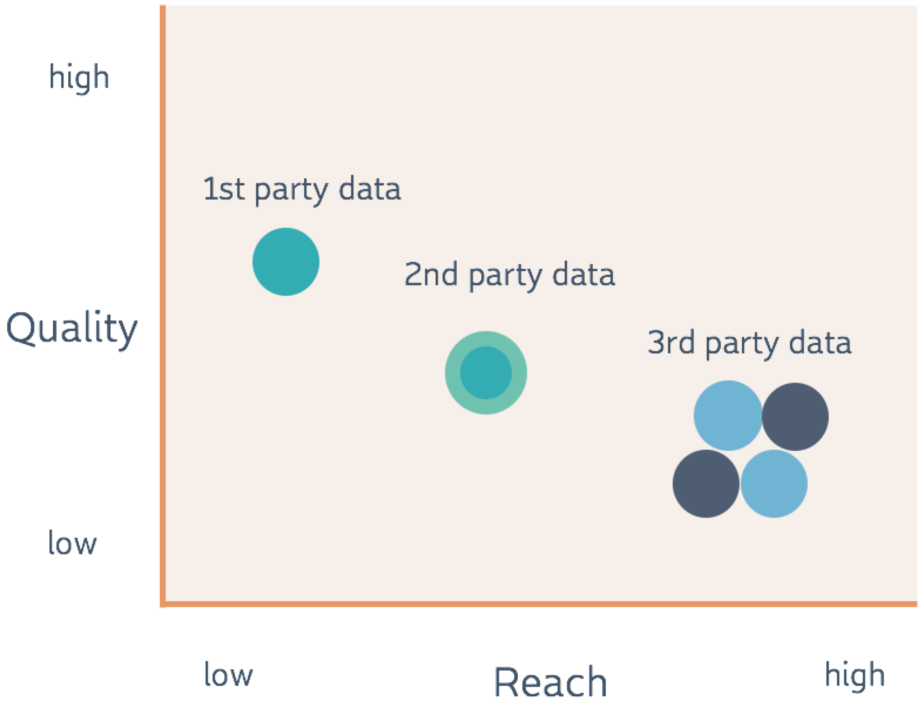


Figure 18 – Differences of 1st, 2nd, and 3rd party data

As reported in [9],[14] third party services, are also facing vulnerabilities, including cross-site scripting and cross-site request forgery that enable an unauthorized and unrelated “third-party” website to retrieve information or perform actions on the “first-party” website that the user has only consented with. User’s web browsing history is directly linked to personal information. Any page a user visits, can expose location, interests, purchases, employment status, sexual orientation, financial status or similar things. Examining individual page visits is very enough to make conclusions about a user while analyzing patterns of activity provides even more assumptions. Every time a first-party page embeds third-party content, the third-party website is getting aware of the URL of the first-party page through an HTTP referrer. If the page embeds a script tag from a third party, it will also learn the web page’s title from document’s title. Some first parties will voluntarily transmit even more information. Collection of sensitive personal data is a real threat and not an imaginary anxiety. A few years ago, Jonathan R. Mayer and John C. Mitchell from Stanford university, discovered that an advertising network, EpicMarketplace, had publicly exposed some of its segment data, offering a brief look on what third-party trackers seek to learn about users. These data segments included pregnancy issues, financial worries and possible ways to face them. The team from Stanford also found that the free online dating website Occupied was sending data to a specific provider, about how often a user drinks, smokes, and does drugs.

Today, users are demanding greater privacy, which includes transparency, choice and control over how their data is used. That’s why Google has announced that it will phase out third-party cookies on Chrome browsers by 2022.The way that advertisers and marketers could survive and still have effective techniques to target optimally their users, is described in Chapter 5.

2.4 User Tracking

One of the most important terms in online advertising is the ROI (return on investment), that means that advertising agencies try to find ways to make their ads even more effective. A lot has changed since the mid-1960s. Advertisers start to have access to a wealth of granular ad tracking data for every single campaign they run. So much data, in fact, that most marketers need to spend significant time sifting through dashboards to determine which points matter to their bottom line. With ad tracking, marketers now have the possibility to measure, test and examine ads based on how users interact with them. Nowadays, advertisers have the ability to measure everything from clicks and views, even impressions and behavior across multiple websites and sessions.

The term Ad tracking refers to the process of collecting data and specific information from users that have to do with an online advertising campaign. To achieve this goal, advertisers adopt several methods such as, cookies, tracking URLs and tracking pixels. As a first step you must determine the metrics you want to track for your campaign, then you have to find the best ad tracking method for your purposes (**Figure 19**).

Figure 19 – Different ways for user tracking

- **Cookies** are the driving force for analyzing user behavior on a website across multiple sessions of activity. Advertisers are looking for specific consent from users while they are tracking their activity using their cookies. User's experience can be customized only when explicit consent is given. From an ad tracking perspective, cookies are responsible for the most ad tracking campaigns. Using cookies is easy to build a user profile considering someone's web activity and habits. Advertisers by owning this profile can display ads that are highly relevant to the user's interests. Finally, there are several other information available to capture using cookies such as user's location, language and browser configuration.

  The method of cookies seems ideal for serving ads aligned with their activity on the web or with products they've demonstrated an interest in. By using cookies, users can enjoy a personalized experience on a website based on their previous interactions.

- **Tracking URLs** is the option in which we add a tracking token, also called a UTM parameter to the end of a normal page URL. Here is an example, the first one is the regular one without using a tracking token while on the second URL we use a tracking token to the end (in bold).

  Regular page URL:

  http://www.yourwebsite.com/your-landing-page/

  Page URL with a tracking token:

  http://www.yourwebsite.com/your-landing-page/?utm_campaign=test-campaign&utm_source=email

  Using this extra element in the second URL, advertisers can get information about where users click the link. This action is quite useful in case an advertising campaign runs on multiple websites. The exact procedure of how this UTM parameter works is the following. When a user clicks on a URL with a UTM parameter added to the end, it essentially sends a signal back to your ad tracking tool that the URL was clicked. The "source=_____" bit of the tracking token can provide information about where the user clicked the link. Similarly, the "campaign=_____" bit can be used to signal to your tracking tool that the link should be bucketed as part of a campaign. [33]

The option of tracking URLs seems ideal for a marketer who runs a PPC campaign, using another website for his advertisement or sending an email and wants to measure the resonance of clicks and leads of his work.

- **Tracking pixels** is another popular way to track users. This method uses a transparent pixel image 1x1 that you can paste into ads, e-mails, and webpages. The main job for this pixel is to send a signal back to your tracking tool every time that pixel loads. With that technique the system knows that a user viewed the page and collects data about users and their activity. On the other side, the company must keep only the data concerning its campaign and to discard the rest. When tracking pixels used in a legitimate way, they are able to provide clear information about how many people click on an ad versus how many views it. Apart from that they can be helpful if you want to optimize your ads and orient them in a specific audience.

  Tracking pixels are ideal if you want to analyse the success of every step of your conversion path. They can help you optimize each step of your user journey giving your insight into how they are interacting with your ads.

The main purpose of ad tracking is generally to provide a measure of the combined effect of the media weight or spending level, the effectiveness of the media buys or targeting, and the quality of the advertising executions or creative. Advertisers use the results of ad tracking to estimate the return on investment (ROI) of advertising, and to refine advertising plans. Sometimes, tracking data are used to provide inputs to Marketing Mix Models which marketing science statisticians build to estimate the role of advertising, as compared to pricing, distribution and other marketplace variables on sales of the brand. [38]

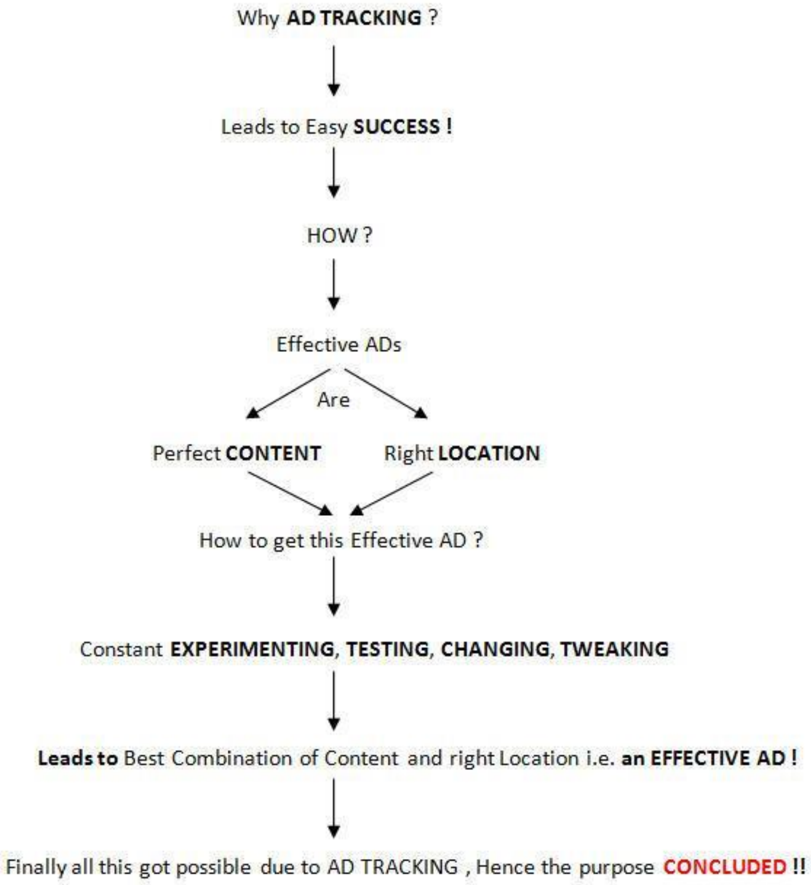

Why AD TRACKING ?
Leads to Easy SUCCESS !
HOW ?
Effective ADs
Are
Perfect CONTENT
Right LOCATION
How to get this Effective AD ?
Constant EXPERIMENTING, TESTING, CHANGING, TWEAKING
Leads to Best Combination of Content and right Location i.e. an EFFECTIVE AD !
Finally all this got possible due to AD TRACKING , Hence the purpose CONCLUDED !!

Figure 20 – Purpose of Ad Tracking [37]

With the help of data that you tracked you can achieve an effective Ad which targets the right audience. So for being certain if a campaign brings the desired result you need to track your visitors on your website, to count the number of them who came through a particular ad, track what they are doing while staying on your site, how many of them convert into buyers and finally your ROI.

The advantages of ad tracking are clear. They allow advertisers to have a better view of performance of their ad campaigns while understanding their audience. The main benefits of ad tracking are the following:

- **Understand your audience:** Ad tracking gives the opportunity to advertisers to understand their audience better. So, with the data collected, they can answer questions such as: Where the users are coming from? What are their interests? And finally, which keywords do they use to find the business? Advertisers usually search answers for these questions, as they are the first step if they want to increase their revenue.

- **Optimize your campaigns:** Knowing what customers want makes it easier for advertisers to improve their campaign performance, optimize to increase customer reach and shift money spent to real profit.

- **Personalize your content:** One of the main purposes of ad tracking is while knowing better your audience, to create tailor made content for each of them. That practically means to offer the most suitable content to the most relevant user.

- **Boost conversations:** Providing with the most relevant content to users, makes it easier for them to convert. Conversion might be a positive response to a desired action such as sales.

With Ad tracking marketers have a variety of metrics to decide which of them are valuable so to focus on for their campaign. Data gathering is incredibly sophisticated and now forms a cornerstone of modern digital marketing. Ad tracking technology allows marketing teams to track the following aspects (Figure 21):

1. **Clicks** the number of times users clicked on an ad and navigated to website

2. **Impressions** the frequency that a user view an ad

3. **Views or Visits** number of users navigating to the destination page

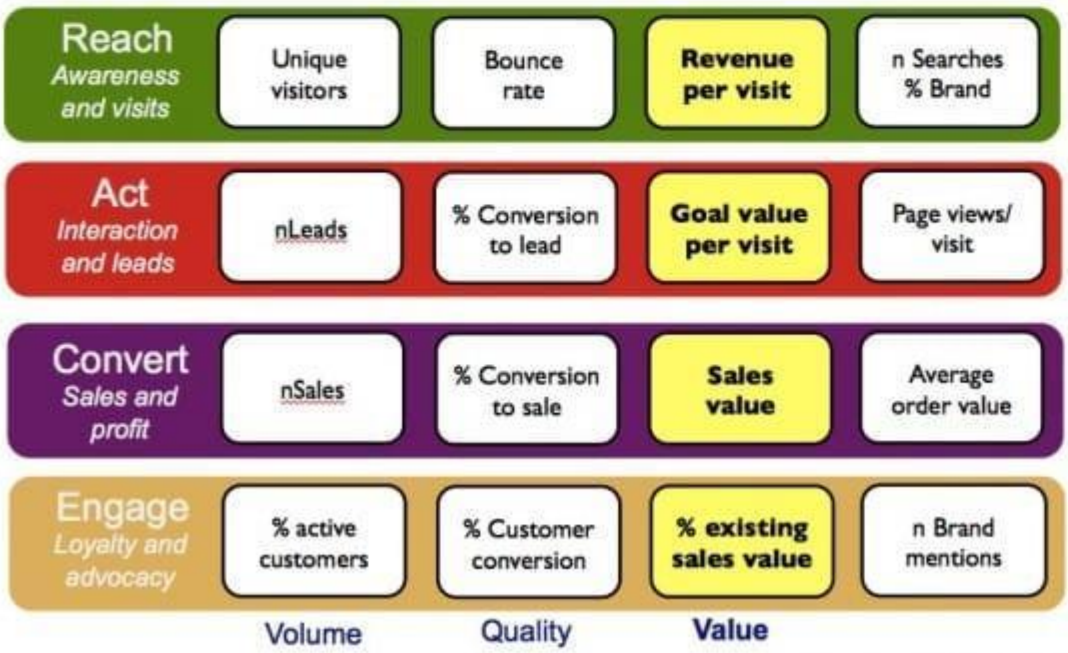
4. **Conversions** number of times that user performed a desired action (for example sale, video view, sign up to a form)

5. **Cost** amount of value an advertiser spent for a campaign

6. **CPC** Cost per click

7. **Value of Sales** Amount of revenue gained for a product sale originated from the campaign

8. **ROI** Return on advertising spent, the real value gain for the company excluding cost



	Volume	Quality	Value	
Reach *Awareness and visits*	Unique visitors	Bounce rate	**Revenue per visit**	n Searches % Brand
Act *Interaction and leads*	nLeads	% Conversion to lead	**Goal value per visit**	Page views/ visit
Convert *Sales and profit*	nSales	% Conversion to sale	**Sales value**	Average order value
Engage *Loyalty and advocacy*	% active customers	% Customer conversion	**% existing sales value**	n Brand mentions

Figure 21 – Online Advertising KPIs

2.5 Recommender Systems

According, to a famous survey [25] held in 2005, regarding the factors of successful purchases, a positive word of mouth about a product or a service was responsible for 31% of the purchases, while personal search was responsible for the 22%. Advertisements and the mass media were responsible for 14%. The rest 33% was ascribed on sales staff influence and situations where the choices were lacking or there was no choice at all. As we can easily understand, the word of mouth has the biggest influence in situations where there are many choices and it is the crucial factor to get convinced or convince the surroundings. If there was a way to combine the first three cases, we would have a situation where we could attract consumers and offer them most choices that lead to earnings. This combination can be established in a virtual centralized environment or network, such as a digital market or digital platforms.

Similarly, to the word of mouth, suggestions and information about products or services can be contagious. Especially in online social network platforms like Facebook and Twitter, there is a massive and parallel distributed spread of concepts and content, at times to such a degree that they become viral or even a new trend. This phenomenon can play a huge role in marketing, as customers spread awareness of interest in products through interactions. Such social interactions structure the main basis for network-based demand shifts for specific products. However, today's platforms can also enable demand to spread across different and potentially competing products. Summarizing, the necessity in systems that can guide the consumer into choosing the most relevant item, is now more prevalent than ever before.

Recommender systems were introduced in the mid-1990s to help users choose the most suitable product or service for them from the plethora of options available. The idea that made this concept prosperous, was that we people rely on the opinions of our peers before encountering a situation which is not familiar. Let's think of a situation before buying a new vehicle or equipment, before visiting an attraction, before going to a new coffee shop or even when searching for a doctor. Their emergence in the research field started about the 1970s in Duke University in North Carolina. Afterwards, many projects and a lot of effort had been done by scientists and analysts in order to conclude in a fortunate result. The very early, operational recommender system ever designed was Tapestry_1 which was developed at the Palo Alto Research Centre (PARC) founded by XEROX. The main factor that has driven scientists to its expansion, was the huge amount of emails that users received and especially those which were too disturbing or even difficult to sort or manage. The process to overcome this issue was that users grouped their mailing list, and it was impossible to receive emails from users that were not belonging in their contact lists. Furthermore, they could hear only accepted groups of people while the others were sent to the spam list, exactly how it works in present in our email accounts.

Recommender systems have the ability to revise the way of interaction between available options and users. At the same time, these systems grow rapidly with the purpose to enrich shopping potential and maximize the targeted advertising. They also try to predict the preference of users about a specific item, in order to suggest it to a specific target group of people. Today the most used section of these systems can be found on commercial applications and websites. They are also applied in a range of areas and can be easily identified as content organizers (e.g. playlist, most recent) mainly for music and video on demand services or product recommenders. We can also find these systems on content recommenders for social media platforms like Facebook and Twitter or in e-commerce corporations like eBay and Amazon. These systems can serve the users, using a single input like music, or multiple entries on each platform as a search query. They are proven crucial and profitable mechanisms both for bigger and smaller companies that are focused on exporting personalized customer experiences.

As we know, the digital and online area is becoming one of the most powerful advertising fields, simultaneously the available methods for tracking user's activity on the Web are proven unsuccessful in delivering an accurate and individual user profile (**Figure 22**). Although precise information about users' actions such as the visited pages, duration of visit, device used, financial status is captured and available on server-side, it is mainly processed with traditional techniques like machine learning and statistical analysis. Indeed, these techniques grant automated identification of customer groups with common interests. On the other hand, most of the traditional machine-learning approaches count on a datapoint, vector dependent input.

They also require the natural content of the web supplies to be converted in this type of format. Despite the existence of techniques that achieve these actions, the conversion usually takes place along with important losses of semantic information with immediate consequences on the quality of the final output. Consequently, personalization or user profiling is the most suitable and demanding solution which grants access to user’s significant information that can be used in classification and item ranking according to the user's unique interests. A User Profile is a collection of elements and patterns that can represent the user purely. This process is crucial for digital companies, in order to provide customized products, services and improvement of user satisfactions. It is also important for providing high-quality web services. Specifically, using a well-organized user profile, online advertising can be further accurate and mainly targeted, focusing both on users’ interests and on his current position and status. Finally, there is also an item profile construction procedure, where each item is represented by its own details and functions.

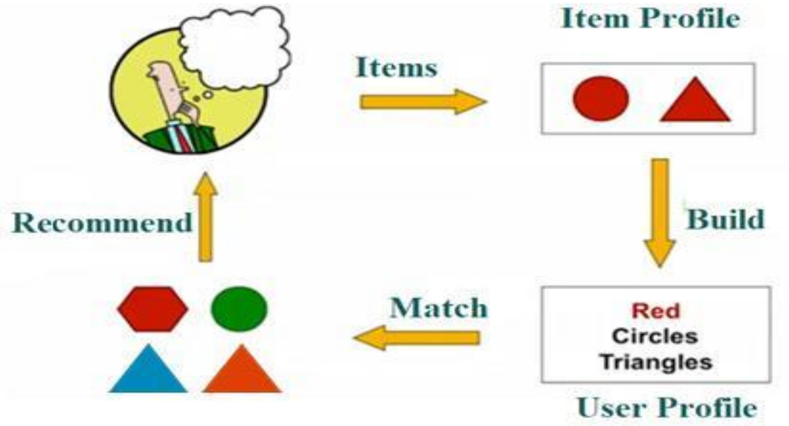


Figure 22 – User to profile to item-profile correlation

The first mechanism focuses on the products themselves and recommends other products with similar attributes with the ones that the user prefers. It is named Content-based filtering and it depends on the characteristics of the products, so it doesn’t need any information about other users or their interaction with other products before taking the decision. The main process is performed by a content-based recommender, which collects the attributes and stores the preferences of a user in a profile, in order to match them with the attributes of an item , so as to recommend new interesting items to the user as in **Figure 23**.It should be mentioned that this type of mechanism is not useful for new users or users with multiple interests.

The second case is often presented as people-to-people interrelationship and it is named collaborative filtering. The main concept of that case is that two or more individuals, sharing a part of similar activities in one field, tend to converge in similar items or products belonging in another area too. It is a process that evaluates items using the opinions of other people. This similarity upon users, can be evaluated by the click-through rate method (the ratio of users who clicked on a specific link compared to the number of total users who viewed a page), browsing pattern and ratings. In contrast with a content-based approach, it is very helpful for new users or users who have many interests.

The third very interesting case is Demographic systems and as we can understand by its name, they are focused on the demography of the users or on the region that each user belongs. For example, consider an online shop that the user must select its country before creating an account, so automatically, the online shop serves the user with products which are only available to ship in his country.

The last type of systems to be described, is called Community-based. This kind of system makes recommendations to the user, purely based on the preferences of the user's connections (friends, followers). A good example is Facebook's concept, where the friends' suggestion list, is mainly based on common friends between our accounts.

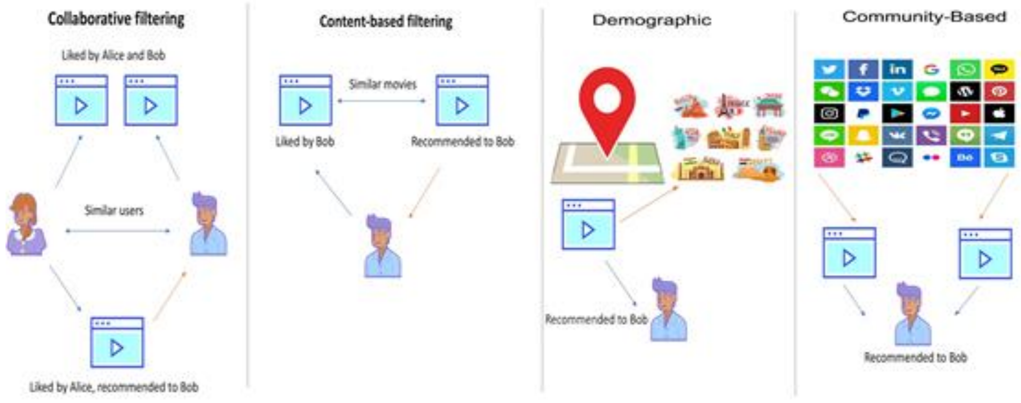


Figure 23 – Types of Recommender Systems

3. TECHNOLOGIES

As we mentioned, companies are gathering and analyzing demographic data from customers and adding them to information from previous purchases, product ratings, and user statistics in order to predict the next move as we already know. More specifically, have you ever felt that Amazon or eBay understands your desires better than your own friends? Have you ever searched a place for holidays in TripAdvisor or Google and then realize that you see a lot of ads trying to sell your local products or suggest your activities to do? This is thanks to recommender systems served to you combined with the science of Big Data. In our world today, businesses collect an extreme amount of data from users on their data farms and use them to anticipate what they can sell them next. Many people may support that data has always been big since the last 30 years that technology forcefully evolved. At this moment, we can think about how the customer data businesses collected personal data 20 years ago, with ways like transaction info, responses to direct mail campaigns, coupons exchange, offline subscriptions and etc. Then think about the customer data collected today by online purchase data, click-through rates, browsing behavior, social media interactions, mobile device usage, geolocation data, etc. Comparatively speaking, there’s no comparison at all. Think of big data as your secret ingredient, your raw material, your essential element. It’s not the data itself that’s so important but, it’s the insights derived from big data. Specifically, it’s the decisions you make and the actions you take that make all the difference. Big data can offer wisdom into not just who your customers are, but where exactly they are, what their preferences are and what influences their loyalty or attention.

The main question is: what is the exact way by which companies collect our data in order to combine them with recommender systems? After the user profiles are created, there are two ways that can convert our personal data into datasets, in order to supply the profiles with important information. The first way is explicit, by evaluating the user's actions like ratings, surveys , purchases ,thumbs up/down or any other distinct action. The actual meaning of explicit in a dictionary, is to state something clearly and in detail but on our occasion, things are different as the explicit collection of data is not always the most efficient way. If we think about ratings, most people use to rate a movie or an item only on extreme feelings like situations of a pleasure or a total disappointment with the latter being more possible. Also, some people are lenient with their ratings while others are strict about the ratings they provide. Moreover, the context or the emotional situation when a user purchases an item or uses a service (e.g. watching a movie), cannot be taken into consideration in the result. In addition, there is a need for a separate system like matrices where we can store this type of data in order to examine them asynchronously. The second way is called implicit, and the collection of data takes place passively. This type of data could be users' geographical region, the name of his Internet Service Provider, the type of the users' browser, or demographic (age, gender, profession, and education). Since it can be gathered silently without disturbing users and in great quantities, the potential impact of implicit feedback is ultimately a great method, but it doesn’t directly reflect the interest of the user as it does not measure the negative preferences. Also , there is another technique in existence , AJAX, an acronym for Asynchronous JavaScript And XML, where a server-side application can transfer data without reloading the page and it was first introduced as remote scripting in 2002.The collection of data takes place passively once more , but it is happening in real time . It became highly popular when the engineers of Google started to use it in Google Suggest, Google Business, Gmail and Google Maps. It is a combination of various existing technologies, using XMLHttpRequest object to exchange data asynchronously with a server, JavaScript in order to display them and DOM to interact with the information. As we have dealt with the techniques and algorithms used in recommender systems and data collection, privacy is always an absolute argument. Even though, in explicit data the user is informed about the collection and storage of data about him, there are some strict instructions that must be observed. The same level of importance is in force even in the implicit data collection method, where a user does not know and does not recognize the information acquiring process. There are many instructions and regulations both nationally and internationally that define standards for collecting, storing and managing data, but finding a middle ground solution between marketing section and governmental orders, remains a very challenging task. Even though recommender systems are very useful, more than 75% of web users refuse to exchange personal information with the web pages in specific cases.

Recommender system engineers and internet service providers (ISP) are obligated to find an approach of operation, which authorizes them to generate valid predictions and at the same time, maintain the privacy of the users. The main risk of privacy originates because of the collection and storage necessity. Although, If we consider some modelling schemes , where the users' data are handled by multiple intermediaries in order to provide us a more mature prediction , there is a high chance of exposing information that users are not aware of , or even worse ,that they never consented to be visible by third parties.

3.1 Real Time Bidding

A completely new concept appeared in 2009, this is the real-time bidding (RTB), or programmatic buying, which is a tool for advertisers to help them make decisions for every impression (auction). Before RTB (**Figure 24**) came into the online ad industry the display advertising market consisted of two categories, premium contacts which take about 40% of impressions and ad networks which take the rest of impressions.

The first category with premium contacts is a direct negotiation deal between publishers and advertisers for a certain amount of impressions. Advertisers, not considering the identities of users, how or when they have seen the ad, buy a specific amount of impressions from the given placements. From the publisher's part, must guarantee the display of the specific number of impressions so as to avoid a contingency penalty. For premium contacts used mostly cost-per-mile (CPM), also is a bit difficult for advertisers to develop goal-driven campaigns (e.g. booking a ticket) than branding one's (e.g. announcing a new product) for the reason that they have no control over the users.

Ad networks use a more complex procedure from contracts, publishers register placements and then offer impressions for sale. Impressions are sold through second price auction in ad networks. The main responsibility of the ad network is to understand the user and the webpage and find the ideal advertisers considering their predefined targeting rules. The Understanding of the users (a.k.a. behavior targeting) is a process in which ad networks use the browsing history of users to identify their most useful for target matching information such as location, local time and interests. On the other hand, understanding of webpage (a.k.a. contextual advertising) is a process where ad networks detect, analyses and extract keywords which summaries the target. In ad networks can be used cost-per-click (CPC) or cost-per-acquisition (CPA) pricing models. Adopting these price models, advertisers can pay only when a certain goal is achieved.

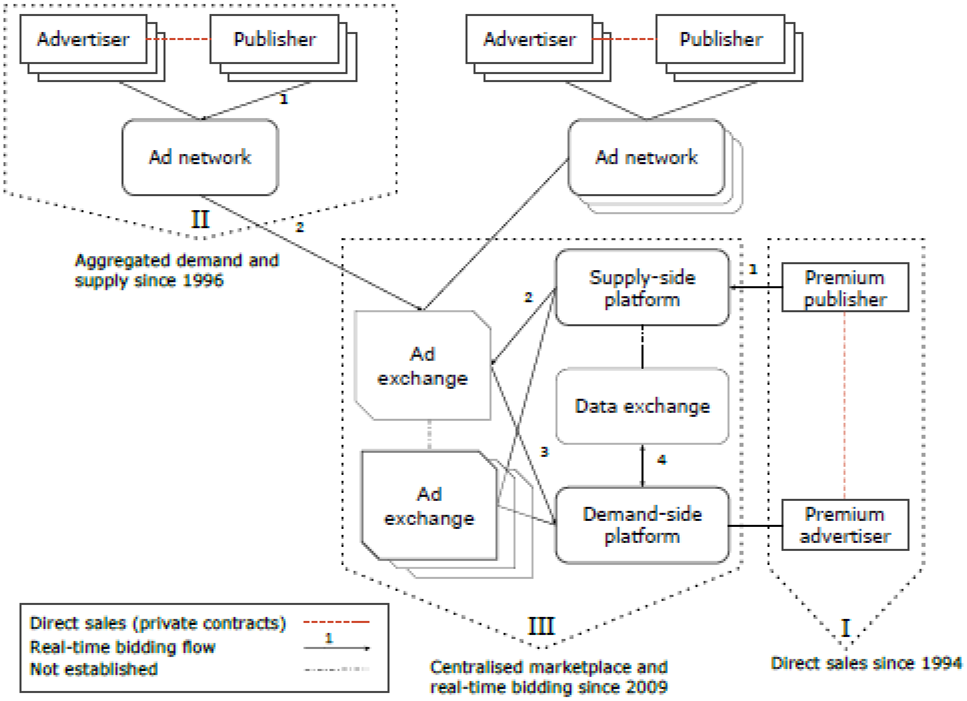


Figure 24 – Real time bidding market since 1994 [45]

The existence of more and more ad networks was a problem that led the market to the birth of ad exchanges. Having as a rule that it is better to have more demand rather than supply, since healthy competition will increase revenue for both sides, ad networks and publishers. Apart from that, ad networks try hard to find buyers and sell all the available impressions. From the other side advertisers had a common practice, to register to multiple ad networks in order to find enough impressions within their budget or to find cheap inventories. So, ad exchanges play a significant role in display advertising because now advertisers have the opportunity to locate more and more impressions with preferred targeting rules, and on the other hand publishers are able to increase their profit because more bidders potentially.

In 2009 ad exchange introduced one of the most important properties which is known as real-time bidding. Real-time-bidding is a subset of programmatic media buying and refers to an auction setting process, where advertisers are bidding and buying ad impressions instantly within a fraction of seconds. At the same moment, multiple advertisers can bid on a single impression of a publisher's inventory, and the one with the highest bid is shown to the users. The benefit of RTB for advertisers is that they can apply fine-targeting and focus on the inventory that is most relevant to them and have a better ROI. Another benefit is that because this process is happening real time, they can adjust their campaign budgets and optimize based on actual campaign performance.

3.1.1 KEY PLAYERS

There are four major types of players: advertisers, publishers, ad network or ad exchanges and users (**Figure 24**). Apart from them there is a whole ecosystem [44]:

- **Supply side platforms (SSP)** serve publishers by registering their inventories (impressions) from multiple ad networks and accepting bids and placing ads automatically.

- **Ad exchanges (ADX)** combine multiple ad networks together [Muthukrishnan, 2009]. When publishers request ads with a given context to serve users, the ADX contacts candidate Ad Networks (ADN) in real time for a wider selection of relevant ads.

- **Demand side platforms (DSP)** serve advertisers or ad agencies by bidding for their campaigns in multiple ad networks automatically

- **Data exchanges (DX)**, also called Data Management Platforms (DMP), serve DSP, SSP and ADX by providing user historical data (usually in real-time) for better matching.

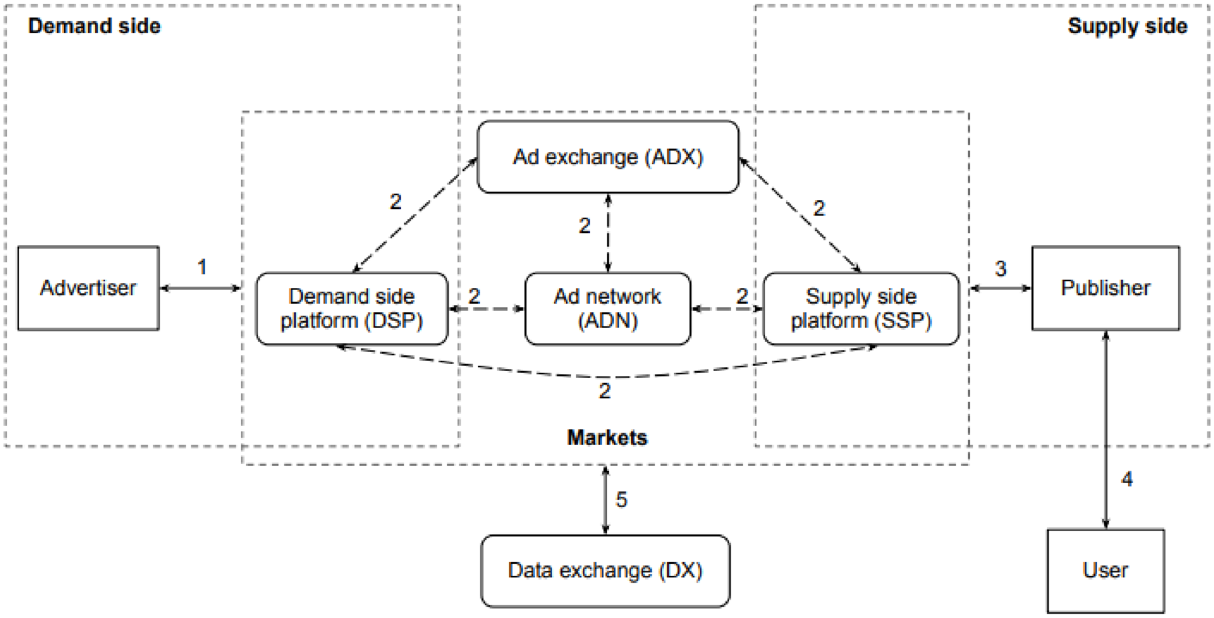


Figure 25 – The various players of online advertising and the ecosystem [44]

In the current RTB ecosystem, advertisers can target specific users based on their previously observed behavior. In Figure 25, it is described in detail the process of RTB between the various players of online advertising and the ecosystem [48]:

1. The advertiser creates campaigns in the market

2. An impression is created on publisher's website, when a user visits a web page

3. The bidding request is sent to ad exchanges through ad network or SSP

4. The ad exchanges query DSPs for advertisers' bids

5. The DSP could contact data exchanges for 3rd party user data

6. If the advertiser decides to bid, the bid is generated and submitted; the winner will be selected at ad exchanges, then at SSP. Following the reversed path, the winner's ad =will be displayed on publisher's website to the specific user.

8. The markets can query data exchanges also known as data management platforms, user profiles in real-time. Platforms are tracking whether a user clicked the ad or completed a conversion.

The above process marks a fundamental departure from contextual advertising as it puts more focus on the underlying audience data, rather than the contextual data from the web page itself [48]. Today most of the e-commerce platforms are using behavioral targeting to re-target users who have previously seen a product from the website but have not converted yet. The final goal of this retargeting method is to direct users back to the website and convert.

3.2 Cookies Matching/Synching

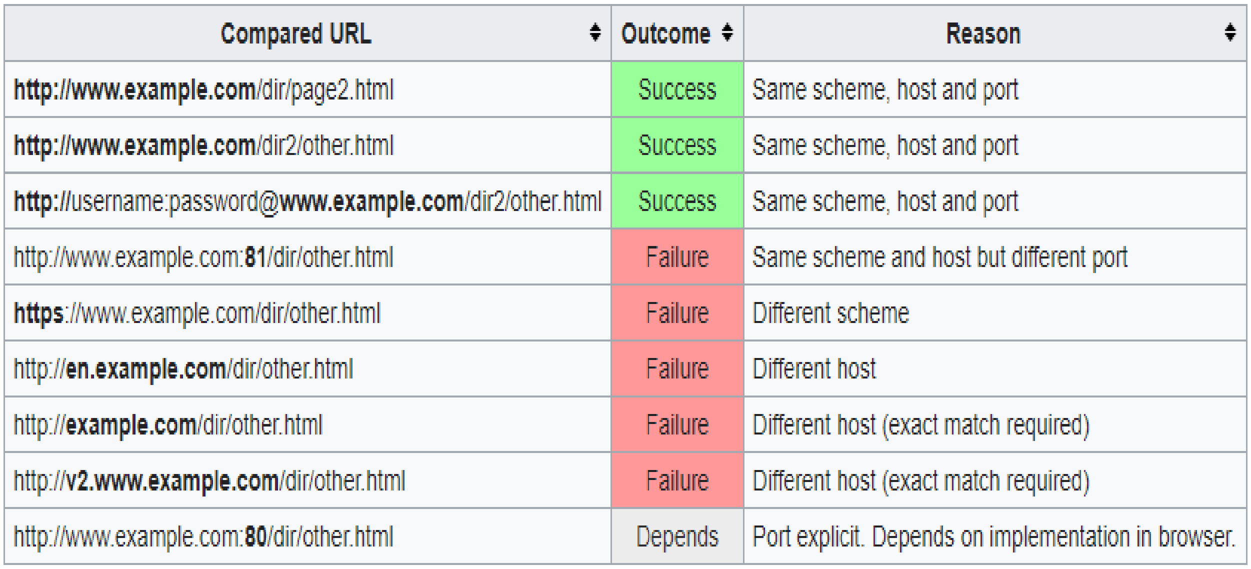
Nowadays tons of user data are generated by the increasing number of users accessing the Internet. Cookies remember user choices and information in order to bring an ordered and more efficient experience. They can also be stored in a name-value pair that is accessible only by the parent domain. However, these tiny amounts of data are also the heart and soul of online advertisers. Advertisers' main purpose is to retrieve as much information as possible, like anonymous information from a parent domain or user's data from external sources in order to overcome their competitors. This process has a huge impact on a company's overall market value. As a next step they use this information to create profiles containing important details about a user or they sell this information to third parties for advertising or other purposes.

Advertisers' main challenge is the expansion of the online display advertising system and the ability to target the right audience. As mentioned earlier ,the most efficient way to maintain some kind of state on the web and identify or authenticate users across different sessions and domains, is by using 1st-party cookies for users who repeatedly visited the same site, or 3rd-party cookies to track users when they move from one website to another with the cost of exposing their data. For that reason, the Same-Origin Policy was invented to restrict the amount of information that trackers can collect from users and exchange them with other 3rd-party platforms. The algorithm used to calculate the "origin" of a URI is specified in RFC 6454, Section 4.Mainly ,the origin is defined by the triplet {scheme, host, port} as seen in Figure 5. Origin checks are applied by the browser in every case of an interaction between elements from different origins. More specifically:

- Restrictions on JavaScript code and the Document Object Model (DOM), which represents the document as nodes and objects in such a way that programming languages can connect to the page. For example, a page cannot access the content of its iframe unless they are of the same origin.
- A session cookie for a particular site cannot be sent to a page with a different origin.
- AJAX calls, especially between *XmlHTTPRequest objects for http requests, are restricted between scripts of different origins*.

Although, Same-Origin Policy (**Figure 26**) does not completely abort the interaction between different origins or domains. The browser may estimate if the interaction should be considered as a threat or not, and allows some specific actions:

- There is an option of writing between origins. It can create cross-origin links and submit cross-origin forms.
- A part of embedded data between origins is also granted. For example, the usage of content from a different origin in an *iframe,* or embedding an *image*, a *css*,or a *script* from a different site.
- Reading between origins is usually blocked. It means that a user can send a cross-origin request but cannot read the reply.



Compared URL	Outcome	Reason
http://www.example.com/dir/page2.html	Success	Same scheme, host and port
http://www.example.com/dir2/other.html	Success	Same scheme, host and port
http://username:password@**www.example.com**/dir2/other.html	Success	Same scheme, host and port
http://www.example.com:**81**/dir/other.html	Failure	Same scheme and host but different port
https://www.example.com/dir/other.html	Failure	Different scheme
http://**en.example.com**/dir/other.html	Failure	Different host
http://**example.com**/dir/other.html	Failure	Different host (exact match required)
http://**v2.www.example.com**/dir/other.html	Failure	Different host (exact match required)
http://www.example.com:**80**/dir/other.html	Depends	Port explicit. Depends on implementation in browser.

Figure 26 – 5 Same origin policy examples

In order to overcome this restriction, and create unified identifiers for each user, the ad-industry invented the Cookie Synchronization process [24].It is a mechanism that can practically slide over the Same-Origin Policy, which allows web companies to exchange cookies/user data between them, while matching the different IDs assigned to the same user and detecting its profile. In detail, it is a process of mapping a user from demand-side platform (DSP) to data management platform (DMP) by assigning users with a unique ID.

As mentioned in 2.2.1 in Chapter 2, different platforms and components like SSP, DSP, DMP, publisher and advertiser, store a range of information characterizing a single user. In order to maintain a fully detailed user profile, these platforms share user data between them, with the purpose of identifying a common user. Let's consider an online environment where a user tends to buy a new helmet, but he can't afford it yet. Let us also assume that this user has placed the helmet on the online shopping cart and then canceled his preference. If that user visits another webpage with a totally different content, there is a high chance that the DSP of this website is assigned the task to find the user on the web and match him with a profile. As a next step SSP will be triggered and show him a relevant ad about the specific helmet. The main question is, how did these platforms work together and found out that it was about the same user. The simple answer is by using cookie syncing, a simple data sharing process. Foreign parties interested in cookie syncing, pairing the data they have collected about the users, using behavior matching algorithms. Using this information, advertisers are able to target users in a real-time auction.

As stated in [24], Google's method in order to populate users' match tables, a match tag is placed on every page of the interested domain in matching advertising data. The match tag is represented by a pixel which tracks visitor behaviors and records information about them. Anytime someone lands on the site or clicks on any link, administrators or advertisers will automatically receive that data. For example, the advertiser can either desire to match every user who visits his domain using pixels on nearly all pages, or he could match converting users by placing pixels on a conversion page. Generally, a more widespread pixel will lead to higher match rates. The match also contains user's cookie matching profile ID and an encoded cookie ID at the following format:

```
<img src="https://cm.g.doubleclick.net/pixel?google_nid=adh_customername&google_
hm=Q29va2llIG51bWJlciAxIQ" />
```

This match tag is what initiates communication between advertiser and the Google cookie matching services and a step by step process is explained below:

- A user visits a page with a match tag.
- The match tag initiates a series of redirects to DoubleClick, Google, and YouTube matching services. The requests contain that user's ID or cookie from your website, plus the Google cookie in each of the matching service's ID spaces.
- A transparent 1x1 pixel is returned to the browser to confirm that the request was fulfilled.

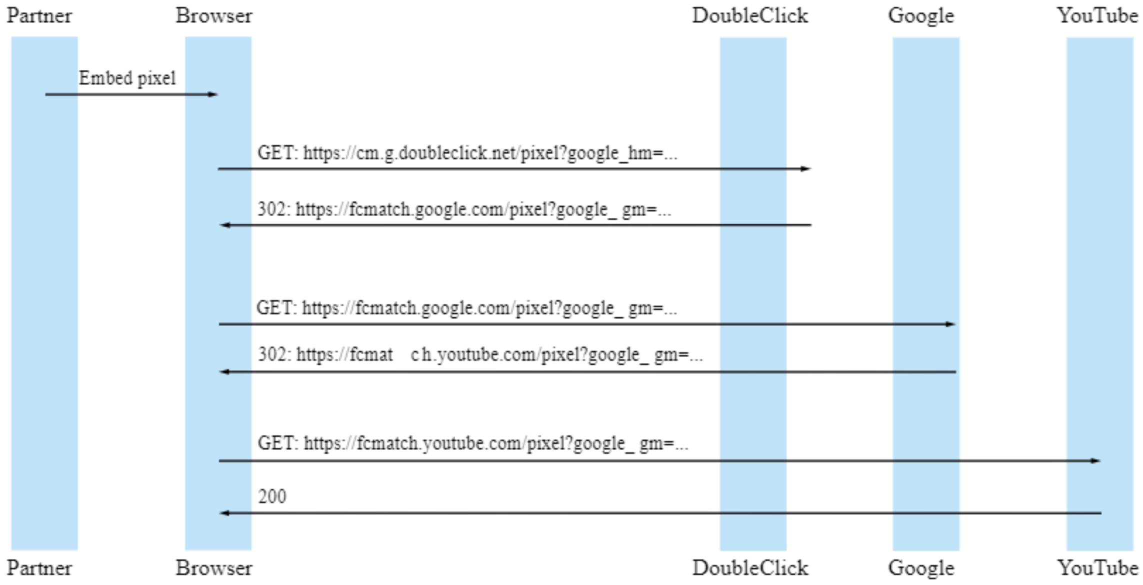

Partner
Browser
DoubleClick
Google
YouTube
Embed pixel
GET: https://cm.g.doubleclick.net/pixel?google_hm=...
302: https://fcmatch.google.com/pixel?google_ gm=...
GET: https://fcmatch.google.com/pixel?google_ gm=...
302: https://fcmat ch.youtube.com/pixel?google_ gm=...
GET: https://fcmatch.youtube.com/pixel?google_ gm=...
200
Partner
Browser
DoubleClick
Google
YouTube

Figure 27 – Google's Cookie Syncing mechanism

Despite Cookie syncing is one innovative technique, it also faces a lot of challenges. Locating and matching a user based on structured data from so many intermediaries like DMPs or publishers can be tough and very demanding. A lot of publishers or platforms do not want to share their user's data and they take advantage of them exclusively, placing a "wall" in cookies syncing procedure (**Figure 27**). Many social media platforms like Instagram and Facebook, target their own users without sending users' data out of these platforms but at the same time behave as publishers and advertisers.

Also, there are many deviations as only 60% of user's data are correctly matched. There are multiple users with similar profiles or cases with multiple users with the same profile because of some confusions between the involved components. That means that there is a big gap in group targeting methods since now and some auctions as described in 3.1 Real Time Bidding may not have the appropriate result.

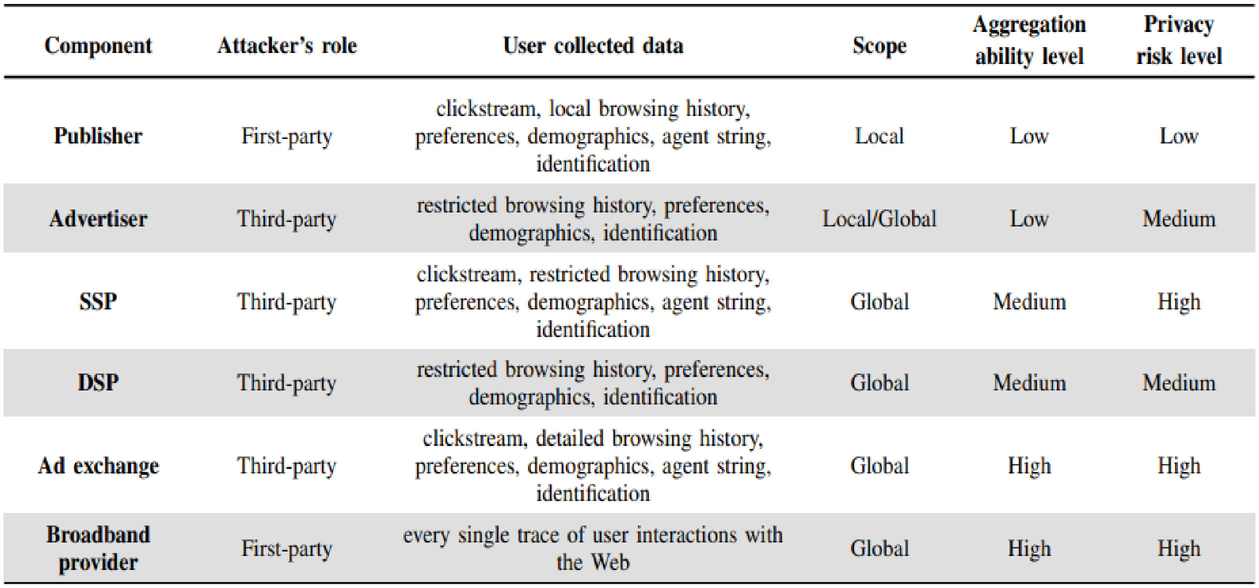
Unfortunately,[23] recent results showed that most of the third parties involved in Cookies Syncing, are also sharing user data with other parties than those agreed by the user. Specifically, 157 of top 200 websites, almost 80%, support 3rd-party data, which synchronizes cookies with at least a foreign party, who is able to rebuild more than half of a user's browsing history. Also, 95% of pages that support 3rd-party data ,make requests to potential trackers and 78% attempt to transfer unsafe data .Finally, a mechanism for respawning cookies has been created by ad networks, with consequences in the reconstruction of users' browsing history, even if they delete their cookies. The information discussed in the whole chapter but also in this paragraph, generates a lot of anxiety and questions about the protection of users' privacy and personal data and in which degree they are violated. We will focus and analyze this issue in the next Chapter.

4. PRIVACY THREATS IN ONLINE ADVERTISING

4.1 Online Advertising: Analysis of Privacy Threats

The health of the online advertising ecosystem has turned to a critical concern for both advertisers and users. Sharing and transferring valuable and sensitive information is a vital process in many conditions of our modern lives, including areas like energy, health, financial, research or e-commerce sectors. Some years ago, these processes were mostly hand-operated, but due to the rapid improvements in computation resources and the raise of software science like machine learning or artificial intelligence, they are now becoming increasingly automated. As mentioned in the previous chapters, except for the automation approach, the personalization factor has now become a standard and the majority of web components are adapted to that situation. As a result, users are facing many security and exposure risks, when personal or sensitive data about their daily life are exchanged in the name of personalized advertising by an infrastructure that operates in the shadows without any surveillance or penalties.

The majority of the prospective attackers are not directly involved in the raw web traffic spawned by a user or at a user's side, as there are many innovative security solutions both for the web and for local user's components, but it is happening at the level of ad-serving entities. A great amount of the concerns regarding privacy and privacy threats in online advertising are originated from the mistreatment of this huge amount of user data, which is taking part at a logically centralized spot, the advertising platforms. It can be easily considered as a single point of failure approach and a mistreatment of user information, might include common privacy issues such as data leakage or unauthorized collection and share of data after a malicious or adversary attack[46].Any possible attack can be defined in terms of the amount and quality of information that potential attackers might be able to collect about users. It could be just simple pieces of information like gender, location and preferences or at the worst case, a full detailed access such as knowledge about every single trace of user interaction through the Web. The type of access and value of information is dependent on the attacked and exposed component, and it could be the publisher itself or even a national service provider as can be seen in Figure **28.**



Component	Attacker's role	User collected data	Scope	Aggregation ability level	Privacy risk level
Publisher	First-party	clickstream, local browsing history, preferences, demographics, agent string, identification	Local	Low	Low
Advertiser	Third-party	restricted browsing history, preferences, demographics, identification	Local/Global	Low	Medium
SSP	Third-party	clickstream, restricted browsing history, preferences, demographics, agent string, identification	Global	Medium	High
DSP	Third-party	restricted browsing history, preferences, demographics, identification	Global	Medium	Medium
Ad exchange	Third-party	clickstream, detailed browsing history, preferences, demographics, agent string, identification	Global	High	High
Broadband provider	First-party	every single trace of user interactions with the Web	Global	High	High

Figure 28 – The type of component that leaked the information [46].

There can also be another categorization of attacks as described in [49], which are not focusing on the intermediate information of ad platforms but in the operation of them , and they are called ad fraud, ad injection, privacy theft, and advertising respectively:

- **Advertising fraud**, also called ad fraud, is the most widely deployed attack on online advertising. The attackers are trying to maximize the gain from the ad ecosystem, even without the existence of a valid audience. It is divided into subcategories determining more clearly the type of attack. The first subcategory is impression fraud and it is the simplest form of ad fraud. The attack focuses on generating multiple HTTP requests and flooding either the publisher's page, or the ad server directly, and by cheating, it increases the actual amount of traffic that must be managed. It clearly targets the Cost Per Impression pricing model, where the advertiser is charged based on how many times its product is viewed by the ad traffic. The second subcategory is called Click Fraud and it is the most popular technique because of the widely adopted Cost Per Click pricing model, in which advertisers are charged based on the number of clicks on their products. It can also be achieved by generating HTTP requests to advertisement click URLs, when an ad has been spotted. There are two types of click fraud [50], click inflation and competitor clicking. The first type is where publishers make more money than they should, by creating fake traffic using artificial clicking. Competitor clicking is a different approach where advertisers make false clicks in a competitor's ads to consume the competitor's advertising budget. The third subcategory is named Conversion Fraud. If a user follows an ad that leads to the advertiser's website and makes a purchase, this process is called conversion. The conversion can be imitated, if the visitor did not in fact make the purchase but the publisher only made it appear to be a conversion to receive the credit. This method can only work if the action does not require deposits, but payments made after a service has been provided.

- **Ad Injection**. Most of the personal data violations are happening in the form of ad injection, which has a direct impact on a publisher's revenue. The Internet follows the end-to-end approach, which implies that any intermediary node between two communicating processes must not alter the original information. The ad delivery process is through client-server communications, which is mostly via HTTP. So, the interference in the end-to-end process can easily alter ad contents and their delivery process. Even though secure Hypertext Transfer Protocol HTTPS can protect general Man-in-the-Middle attacks, proxies at HTTPS endpoints can still modify pages which are not familiar to the client. Even worse there are many websites that have not deployed Transport Layer Security (TLS) at the proper level , for example some services are supporting versions up to TLS1.1 or TLS1.0 and they are totally exposed to Lautenbacher attacks or chosen ciphertext discover ad injection and the value chain behind it, a client-side DOM scanner was deployed in many of Google's websites. The scanner is able to detect and report crooked ad elements. According to [49] reports', around 5% of unique daily IP addresses accessing Google, are impacted by ad injection. Many legitimate applications use advertisements to earn money while providing the application to users for free. However, malicious applications can take advertising a step further with unfair advertising practices. Ad injection can ruin a publisher's reputation in a few seconds. There are many examples such as politically sensitive advertisements or inappropriate content injection into the publisher's website.

- **Privacy Theft**. Privacy can be defined as secrecy that covers personal information. Any user would want to maintain the control over the collection, retention, and distribution of his personal data while visiting any publisher's domain [51]. A victim of privacy theft can be someone who accessed publishers' domain or any ad network that knowingly mistreats the user's personal information, such as providing the user's physical address and phone number to other parties for a fee without the user's consent. It can be easily done by platforms which use behavioral advertising. On the other hand, about the mobile section, sensitive user's information such as demographic or user's preferences are delivered through trackers to third party entities. With the excuse that many users are satisfied to receive relevant ads, they are being tracked and their personal data are sold to advertiser networks. The loss of privacy is the user's loss of control over the collection, retention, and distribution of his personal information. Many regulations are in existence, focusing and restricting both targeting data and tracking techniques that the online advertising industry can use. Although, the concern about the loss of privacy is increasing dramatically. As an example, there is a serious vulnerability on the Facebook behavioral advertising platform as reported in [52].In that case, users' personal identifiable information (PII) -data that if used alone or with other relevant data, can identify an individual- are collected and advertisers are authorized to choose which users see their ads, without any user's consent prior. There is also the option for advertisers, to upload personal information of the targeted audience to Facebook's behavioral advertising platform in order to examine it and have more specific targeting.

- **Malvertising**. Online advertising has been exploited and manipulated by cybercriminals to spread malware to web users. This phenomenon is named malvertising, a technique that takes advantage of the browser and distributes malware to vulnerable devices through online advertising [53].More simply, a user visits a website that contains an external advertising link in the form of an injected iframe and then he is getting redirected to an invisible exploit kit landing page. Exploit kits are web based services designed by hackers to exploit vulnerabilities of browsers or users by downloading malicious executables files. After that, information about the victim's system is transferred to the attacker's server, which is then used to select a malicious exploit file that is being automatically downloaded to the user's device. The downloaded file exploits a vulnerability on the system that allows the attacker to install malicious files or even control the victim's device. This is also called Drive By Downloads. Two newborn popular malvertising operations are AdGholas [55] and Ramnit [54], which infected general computer browsers and Android smartphones. About AdGholas, security researchers managed to shut down a large-scale malvertising operation that used sophisticated techniques and was undetected for months and exploited millions of computers. According to security vendor Proofpoint, the gang behind it managed to distribute malicious advertisements through more than 100 ad networks, attracting between 1 million and 5 million page hits per day. Although it is hard of ad networks and ad exchanges to filter malicious ads, attackers can use many techniques to bypass detection, including fingerprinting, redirection, just-in-time assembling and timing-based evasion. At a lower level ,users' computers with updated anti-virus protection can prevent damage to a high degree. As for the smart devices, the majority are not equipped with anti-virus protection.

In order to overcome this serious issue, researchers are concentrating on advanced mechanisms in order to anonymize or block the information exposed to third parties, while trying to keep a compatibility with the current Internet's system. By now, many commercial solutions are focused on providing services in a way that is blocking tracking mechanisms at the cost of seriously damaging the Internet business model.

4.2 Online Advertising: Privacy Protection Approaches

With the purpose to face the threats that were analyzed in the previous section, researchers have focused on developing two individual categories of countermeasures, offline and online. Offline countermeasures are aiming to limit and identify any security concern before or after the advertising delivery process and they are actually local mechanisms on the user side. On the other hand, the online approach focuses on encountering security issues during the ad delivery process and it could be in any location remotely from the user side. Both of them, are explained further below:

- **Offline Countermeasures**: The first approach is Data Analytics, where Artificial Intelligence’s methods like machine learning, are used for detecting ad fraud or malvertising by using patterns. In case an advertiser is getting thousands of clicks from a single IP address, none of which is related to any paying customers, the ad network may start filtering and restricting all clicks from that specific IP address or the whole subnet where the IP belongs. Secondly, a static analysis of mobile applications, and their executables files, can detect security issues before they happen. In mobile advertising, to enable ad services or third party interactions, developers need to include third-party libraries in their applications, and these libraries are mostly binary files. The ad library grants the same permissions as its hosting applications. This means that when a user sets the permissions in libraries required by the application during installation, the ad libraries automatically gain the same permissions as the application’s domain. By this way ,the application’s domain owner can set strict restrictions from the beginning. The last approach is game theory, that can be considered as theoretical analysis which provides wisdom and at the same time possible correlations. There can be an analysis of interactions between different players in the advertising ecosystem, but it will be always based on possibilities.

The three principles of information and network security are defined by the CIA triad **(Figure 29).** Confidentiality, Integrity & Availability. Confidentiality measures protect information from unauthorized access and misuse. Integrity measures protect information from unauthorized alteration and Availability measures protect timely and uninterrupted access to the system. Publishers of online ads face two serious challenges. Mainly, they must ensure ads will neither violate the integrity of publishers web pages, nor breach confidentiality of user data present at them. Ads are integrated into publisher web pages, and therefore must coexist with high integrity content and sensitive information. A very popular offline framework is AdJail [62] a tool that aims at authorizing publishers to separate the content elements to which ads will have access to - like not having access to user’s sensitive information-.Specially, this approach is operating by creating a sandbox where ads are executed. From this sandbox, ads may have access to user or publisher specific content through a configurable set of enforcing policies. Even though the real goal of AdJail is to protect the confidentiality and integrity of user and publisher data, user privacy can also be provided by applying strict policies based on privacy agreements between publishers and their users. The trouble that

AdJail faces, however, is that its scope is limited to the publisher’s domain. In other words, users can have access to sandboxing approach only if this mechanism is deployed in the owner’s website.

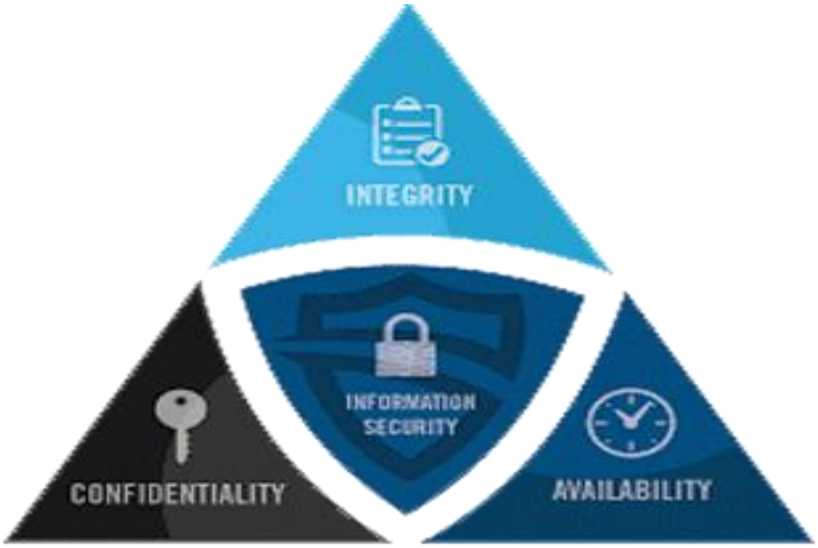


Figure 29 – The Three Goals of Information & Network Security - CIA Triad

- **Online Countermeasures**: The following approaches are aiming to protect advertisers, publishers, or consumers during the ad delivery process. The first approach is by using bluff ads [58]. In this method, tricky or baiting ads are designed in such a way to be detected and triggered by bots or scripted click farm workers in order to expose them. If clicked, the server-side component considers the click as fraudulent and bans the IP or restricts its actions. Secondly, for Man In The Middle scenarios, reference [59],[60] presents a method for fingerprinting transactions between client and server sides.There is an authentication process for valid clicks, where some of them are identified as legitimate or others are rejected. The whole idea is based on authenticating requests using cryptography certifications -like a digital signature- by clients and then a token is used for authentication. By this way, there are unique identifications for every session. Another approach is suggested in [61] where data about user mouse movements are collected and then inserted into machine learning engines in order to detect fraud. This is a direct and effective method to differentiate bots from humans. However, the user profile data obtained by mouse movements, are too large and they demand a lot of space in order to be stored. Also,they are non-structural, which require more effort in applying machine learning techniques .Additionally, creative bots from genius hackers have been reported, which can fool the machine learning algorithms by injecting noise which may be handled as mouse movements by the machine learning engines. The last and very popular countermeasure is ad-blocking, a procedure where users can block ads based on blocklists they made, letting browsers bypass some content, then having to run very time-consuming fetching and rendering of ad. However, ad-blocking is now considered as a threat for publishers that rely on advertising to run their services in order to maximize their income. In order to face this phenomenon, publishers in their majority, have defined their own rule, and their content or services are only available for those who are willing to see their ads, forcing users to disable ad-block scripts.

A popular offline framework is ObliviAd and it was proposed by Backes [57]. It relies on secure coprocessor brokers that use the online advertising ecosystem. Their goal is to securely transfer user's information to the users and at the same time financial reports to advertisers, during the delivery of ads. To achieve that, it provides an individual secure storage of sensitive users or advertisers' data. As for the profiling, the data are being encrypted when transferred to the broker side. Then, the encrypted data are provided and displayed on the user side. The same thing happens for billing data intended for advertisers.

- Another very effective but not valuable approach is using ad hoc networks to filter out ad fraud attacks. In ad hoc approach, the nodes rely on one another in forwarding a packet to its destination, and as there is a limitation in the range of each mobile host's wireless transmissions, the HTTP flooding of ad fraud will be a difficult scenario[56].It can be categorized as a hybrid approach , containing both offline and online elements.

4.3 Privacy Regulations in Online Advertising (ePrivacy Regulation)

Large-scale collection of customer's data via Online advertising has increased customers' concerns about their privacy. Consequently, governments worldwide have started considering new privacy regulations to delimit the collection and use of consumers data by Advertisers. Undoubtedly, this has an impact on campaigns' performance but also restricts many of the most important benefits of user tracking.

In 2017, European Commission published the first draft of ePrivacy Regulation in order to replace the ePrivacy Directive (the 'cookie law'). The ePrivacy Regulation (ePR) is a proposition for the regulation of multiple privacy-related issues, in relation mostly to electronic communications within the European Union. It is important to emphasize that ePrivacy Regulation is not just about cookies, but it concerns electronic communications and the right of confidentiality, data/privacy potential and more [65]. Electronic communications do not only mean the Web but also through the Internet, apps, email, instant messaging and telephone. As a result, privacy regulations are also concerned about spam, direct marketing, Online advertising, mobile Apps and other forms of Electronic communication such as Internet of things (IOT). The ePrivacy Regulation will supplement the GDPR’s general rules on personal data processing by providing specific rules governing electronic communications. Both ePR and GDPR constitute the revision of the EU data protection framework, that includes a new set of rules related to Non-personal data flow in the EU.

ePrivacy Regulation aims to simplify rules on cookie collection. Users will be able to control any sensitive information on their devices, without even clicking on a banner to give their consent for every privacy pop-up in a new website. Instead they will be able to set up browser settings, and users will be able to accept or refuse cookies. In addition to that, for non-privacy intrusive cookies that are used for improving Internet experience (Remember email, shopping cart etc.), no consent will be needed.

According to [66], ePR privacy rules for all electronic communications include:

- **New Players:** privacy rules will be applied also in the future to new players using popular services such as WhatsApp, Skype and Facebook in order to make sure that they have the same level of confidentiality

- **Stronger rules:** all European users and business will be influenced by the same level of protection of their electronic communication

- **Communications content and metadata:** privacy is guaranteed for the content of communications and their metadata, for example location or time. Metadata has a high privacy component and without users giving their consent, they will be deleted unless it is needed for the payment process.

- **New business opportunities:** once users give their consent for use of communication data; businesses have more opportunities to provide extra services and to develop their business. An example could be the creation of a heat map indicating where each individual is located, and it can possibly be used for infrastructure projects to a transport company.

- **Protection against spam:** this means that electronic communications via email, phone or SMS will be banned. Advertisers, depending on national laws, are obligated to display their phone number or use a specific prefix that indicates that this is a marketing call.

- **More effective enforcement:** the imposition of privacy rules in the Regulation will be controlled by Data protection authorities, that are already managing the rules under the GDPR.

In Europe, where privacy laws have started being implemented, most of the online advertising types have experienced a reduction on average in effectiveness of 65 percent in terms of changing stated purchase intent [64]. Regulation is a trade-off between the benefits of consumer privacy and the benefits to consumers of a potentially broader, less intrusive advertising-supported internet [64]. As being reviewed by [64], ePrivacy regulation will impact the effectiveness of general content websites such as news and web service rather than product specific websites. In addition, the biggest impact has been found in advertisements with no audiovisuals.

4.4 GDPR

The right to privacy is part of the 1950 European Convention on Human Rights, which states, “Everyone has the right to respect for his private and family life, his home and his correspondence” [29]. For that reason, the European Union has managed to guarantee the protection of this right by applying strict laws.

While the technology was evolved and the Internet was invented, the EU recognized the need for modern protections. For that reason, European Data Protection Directive (EDPD) was founded in 1995, establishing minimum data privacy and security standards, but each country could adapt it to its own implementing laws. Nonetheless, the Internet was transformed to the anarchic approach with millions of information and devices. In 1994, the first banner ad appeared on the web and six years later, many financial institutions offered online banking. In 2006, Facebook opened to the public and then became a great platform with millions of incomes per year. In 2011, a Google user accused the company for having access to his personal emails and that news went viral. After two months, Europe’s data protection authority stated that the EU needs “a broader approach on personal data protection” and then focused to update the old EDPD which was created in 1995.For that reason ,the General Data Protection Regulation(GDPR) was founded in May 2016 ,formed in May 2018 and is still in progress till today, by the European Union. It is the toughest privacy and security law in the world provided by the European Union with only purpose to put in order the chaotic personal data handling or collection and at the same time protect e-privacy.

The General Data Protection Regulation (GDPR) (**Figure 30**) protects the collection, processing, and use-of personal information of EU residents as well as all firms/companies based in Europe or firms/companies with European offices. It also expands the definition of personal information more than private personal data, including individual high-level data like IP addresses. Further, its’ fines can reach the larger of 20 million euros or 4% of global turnover. More simply, companies are now obliged to ask for users’ consent before collecting and sharing their data. Additionally, users now have the legal right to inspect the personal information collected about them. Before GDPR regulations, users were not informed about collection or usage of their data and were unaware of the consequences. They also had very few options if they wished to regain control of their data or even try to evaluate what type of information were collected or any conclusions about them. According to recital 68 of the GDPR , the right to data portability is meant to support an individual in gaining control over one’s personal data by allowing access to the data stored about him or her “in a structured, commonly used, machine-readable and interoperable format”. [32],[34]

Figure 30 – General Data Protection Regulation (GDPR)

The consent requirement forms a wall of privacy protection for consumers and sometimes threatens the business model of firms that are based on users' data. This option provides a simple but effective way of privacy protection. If a user does not consent, he immediately stops the process of a website from collecting personal data and sharing it with third-party affiliates. At the same time, these denials are blocking firms from tracking consumers across time and across websites as a result, not allowing them to create user profiles. There is also the option of stopping the opt-in and users can block the collection of their data at any time. The GDPR requires marketers to make it as easy to opt-out as it was to opt-in. In addition, consumers using their consent right can significantly restrict the firms' ability to predict consumers' behavior and target their services and advertising, accordingly, creating a serious operational issue for firms.[35]

Under the GDPR, firms face rights obligations and risk obligations [31]. The rights-related obligations require that firms allow individuals to exercise their rights in an easy and timely manner. As for risk-related obligations, firms must appoint a Data Protection Officer to manage compliance activities and must control internal data processes. Also, firms must encrypt and anonymize personal data as well as minimize data collection. In case of a data breach, firms must promptly notify the regulator and affected individuals and inform them about possible consequences. These obligations enforces potentially large compliance costs on companies. Many companies are spending over 10 million dollars each year to make sure they comply with the law and many of them are still coming into compliance after May 25, 2018 (PWC 2018). The GDPR defines the legal bases for processing personal data. Firms can process data in order to fulfill a contract or legal obligation, to protect the public interest and to protect the vital interest of the individual. Otherwise, firms may obtain an individual’s consent. Consent must be affirmative and clear to the user, freely given, fully detailed to the purpose of processing (e.g. website analytics, behavioral advertising, newsletter), and must list all third parties who are going to process the data. According to Article 7 of GDPR, the conditions of a consent should strictly follow the requirements below:

- Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.

- If the data subject’s consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

- The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.

- When assessing whether consent is freely given, it is considered whether the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

As for web analytics, the European Union states that processing data for web analytics even for improving a service by the site cannot rely on any official agreement as web analytics are not obliged to fulfilling a contract. Instead, EU regulators have indicated that firms should seek consent or may use legitimate interest as a legal basis (EDPB 2019). Using legitimate interest is potentially legally risky as it is not clear if web analytics would pass the EU’s proscribed balance of interest test (Article 29 Working Party 2014). Prior to the GDPR, online firms collected this data and there a chance that they have provided an e-Privacy Directive cookie notice to users and offered them that choice. By protecting individual privacy, the GDPR can hurt firms that rely on customer analytics to make decisions or personalized marketing. Online firms collect detailed web analytics data on how users navigate through websites using platforms like Adobe Analytics. Online firms use web analytics data to better draw users to their sites and to improve site content and usability. Under the GDPR, firms may choose to collect less web analytics data or may find that fewer users consent to data collection. The GDPR also increases the cost of personalized marketing channels like email and display ads that draw users to websites as there must be a detailed description of the personal data that are being and will be collected. As a simple example, the main goal for every manufacturer is to sell the product to as many people as possible and get a lot of earnings. In order to succeed, manufacturer needs to inform his clients about his products and at the same time convince them that it will be essential for them. The only way to achieve that, is to explore his clients' preferences and analyze them.

GDPR also changes user preferences for browsing online by making privacy more transparent and actionable. Thus, the GDPR could hurt online firms, by restricting online advertising and changing user browsing preferences, or by reducing the web analytics data with a huge impact on their profits.

At the end of 2010, the total income of the online display advertising market was more than $11.2 billion [39]. The evolution of this section was innovative, and the total income is supposed to be five or more times greater now. Also, Google has developed many highly profitable advertising display platforms with the most popular to be AdSense that generates an estimated $6 billion in revenues by displaying plain content-targeted text ads. The GDPR though, has increased the legal risk for targeting emails and online display advertising as they are purely based on personal data. As the cost of personal information has become great, the personalized marketing brands that use too much online traffic are trapped. As a result, both the quality and the quantity of advertising through these channels will be reduced.[40]Also, as privacy concerns have grown the last few years and all the users are noticed about their data, these notices may have changed user preferences and affect the total time users spend online and which sites they will trust. The GDPR demands that firms reduce the users' data collection. The majority of firms, are enforced to minimize the number of web analytics vendors that they provide their data to, or even do it with a strict users' consents. If the site collects and respects user consent, the share ratio of data will be a function of the total consenting users. Quantcast, the dominant GDPR consent management platform, reports that now, any average website consent rate exceeds 90%.

Nowadays, firms should focus on persuading users about the safe handling of their data, in order to make them feel safe and as a result to get their consent. After achieving that, if they keep complying in the EU laws, they can have access to a large amount of personal information as they used to, at the prior GDPR era. Many firms though, are trying to trick users and get their consent by using immoral tactics like not allowing access to the main part of their web page without consent , or auto select the consent option ,or even not making clear to the user about how his/her data will be processed and by who. The European parliament should consider the unethical behaviors of many firms in order to restore the trust of users to the personal data collected by them and especially for third party data and at the same time apply the transparency as it’s one of the main goals of GDPR.

Some of the main recitals of the GDPR, are described below:

- The protection of natural persons in relation to the processing of personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union (the ‘Charter’) and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her.

- The principles of, and rules on the protection of natural persons with regard to the processing of their personal data should, whatever their nationality or residence, respect their fundamental rights and freedoms, in particular their right to the protection of personal data. This Regulation is intended to contribute to the accomplishment of an area of freedom, security and justice and of an economic union, to economic and social progress, to the strengthening and the convergence of the economies within the internal market, and to the well-being of natural persons.

- The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality. This Regulation respects all fundamental rights and observes the freedoms and principles recognized in the Charter as enshrined in the Treaties, in particular the respect for private and family life, home and communications, the protection of personal data, freedom of thought, conscience and religion, freedom of expression and information, freedom to conduct a business, the right to an effective remedy and to a fair trial, and cultural, religious and linguistic diversity.

- The economic and social integration resulting from the functioning of the internal market has led to a substantial increase in cross-border flows of personal data. The exchange of personal data between public and private actors, including natural persons, associations and undertakings across the Union has increased. National authorities in the Member States are being called upon by Union law to cooperate and exchange personal data to be able to perform their duties or carry out tasks on behalf of an authority in another Member State.

- Rapid technological developments and globalization have brought new challenges for the protection of personal data. The scale of the collection and sharing of personal data has increased significantly. Technology allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities. Natural persons increasingly make personal information available publicly and globally. Technology has transformed both the economy and social life and should further facilitate the free flow of personal data within the Union and the transfer to third countries and international organizations, while ensuring a high level of the protection of personal data.

➢ Those developments require a strong and more coherent data protection framework in the Union, backed by strong enforcement, given the importance of creating the trust that will allow the digital economy to develop across the internal market. Natural persons should have control of their own personal data. Legal and practical certainty for natural persons, economic operators and public authorities should be enhanced.

➢ Where this Regulation provides for specifications or restrictions of its rules by Member State law, Member States may, as far as necessary for coherence and for making the national provisions comprehensible to the persons to whom they apply, incorporate elements of this Regulation into their national law.

➢ The objectives and principles of Directive 95/46/EC (which is the European Directive of personal data protection that was valid since 2018) remain sound, but it has not prevented fragmentation in the implementation of data protection across the Union, legal uncertainty or a widespread public perception that there are significant risks to the protection of natural persons, in particular with regard to online activity. Differences in the level of protection of the rights and freedoms of natural persons, in particular the right to the protection of personal data, with regard to the processing of personal data in the Member States may prevent the free flow of personal data throughout the Union. Those differences may therefore constitute an obstacle to the pursuit of economic activities at the level of the Union, distort competition and impede authorities in the discharge of their responsibilities under Union law. Such a difference in levels of protection was due to the existence of differences in the implementation and application of Directive 95/46/EC. The GDPR, as a Regulation, addresses this harmonization problem since it is directly applicable to all Member States.

➢ In order to ensure a consistent and high level of protection of natural persons and to remove the obstacles to flows of personal data within the Union, the level of protection of the rights and freedoms of natural persons with regard to the processing of such data should be equivalent in all Member States. Consistent and homogenous application of the rules for the protection of the fundamental rights and freedoms of natural persons with regard to the processing of personal data should be ensured throughout the Union. Regarding the processing of personal data for compliance with a legal obligation, for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, Member States should be allowed to maintain or

> introduce national provisions to further specify the application of the rules of this Regulation. In conjunction with the general and horizontal law on data protection implementing Directive 95/46/EC, Member States have several sector-specific laws in areas that need more specific provisions. This Regulation also provides a margin of maneuver for Member States to specify its rules, including for the processing of special categories of personal data ('sensitive data'). To that extent, this Regulation does not exclude Member State law that sets out the circumstances for specific processing situations, including determining more precisely the conditions under which the processing of personal data is lawful.

The data exchange between third parties, is supposed to be shrinked by about 40% after the GDPR took effect [32]. However, the general structure of how the third parties are connected stays the same.

This indicates that the GDPR did not revolutionize the ad ecosystem, but rather has a direct effect on the amount of information sharing in the ecosystem and protect the users from the uncontrollable collection of their data without any consent.

5. HIGH LEVEL PRIVACY

5.1 How to get protected from 3rd party data

Technology has made human lives easier, but it also means that the personal data are no longer personal. Companies and websites track every action of users as described in the previous chapters. Many ads, social network buttons or shares, information about the location, browsing habits, and more. The collected data with a user consent or without, can reveal a lot of important and sensitive information through user profiling. Third-party applications are an effective way for businesses to decrease overhead and achieve goals with a few resources. Despite being cost-effective and boosting companies to remain competitive and profitable, they tend to be the most vulnerable point for hackers aiming to take advantage of security weaknesses.

Protecting user's data is a necessity for any organization. Information security should be a part of every business process to prevent data leaks and breaches especially when there are exchanges of data with third-party vendors. Thankfully, there are a few strategies in order to maximize the data protection and a very effective one ,is by decentralizing the process of authenticating and accessing the personal data storage by using block chaining as described in [63].This framework, mainly focuses on ensuring that users own and control their personal data and decide with which third parties they share their data at any given time. As such, the system recognizes the users as the owners of the data and the services as guests with delegated permissions. There is also full transparency by users over what data is being collected about them and how or from who they are accessed. Moreover, the most important thing offered by this approach is the flexibility and access control. A very frustrating concern with web and especially mobile applications is that users are required to set some permanent permissions upon sign-up. These permissions are granted instantly and there is only one way to modify the agreement, by opting out. Although, in this framework, at any given time, the user may alter the set of permissions and restrict access to previously collected data by first or third parties.

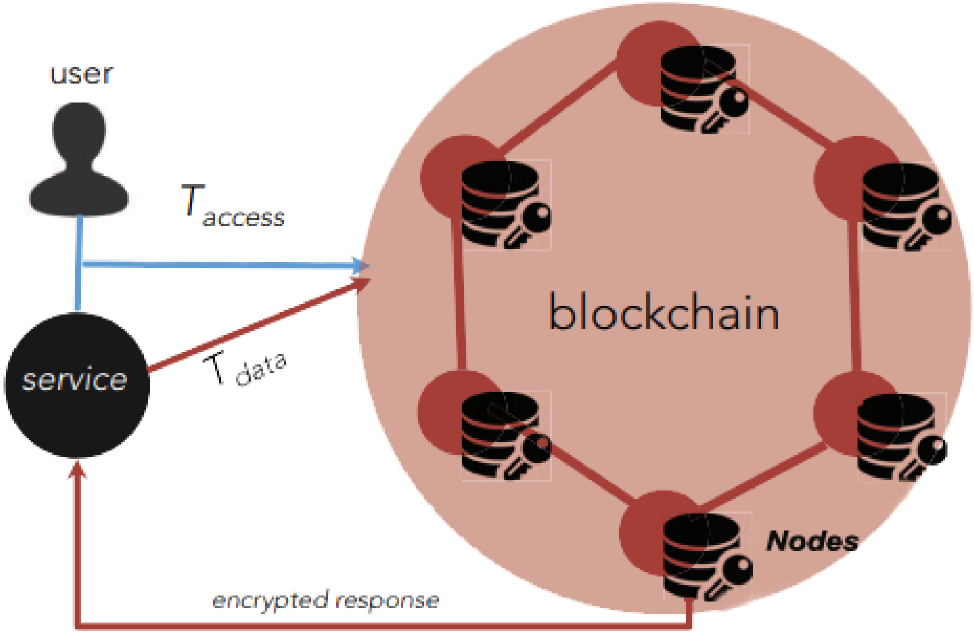


Figure 31 – Overview of the decentralized platform using the block chaining approach [63].

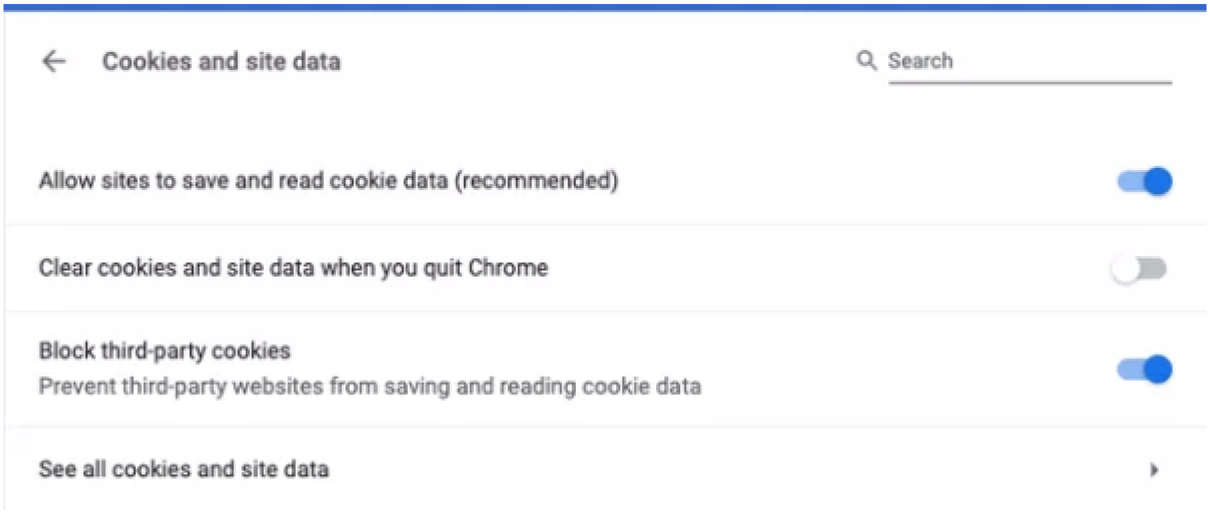
As shown in **Figure 31**, the system is consisted by three entities which are regular users, interested in using applications. Secondly, the services which are the providers of such applications who are responsible for processing personal data before reaching the data storages or businesses related reasons (e.g., targeted ads, personalized service).Finally, the nodes are maintaining the blockchain and grant access to verified users only, using a distributed private key in order to authorize and recognize them.

The framework supports two different types of operation, Taccess used for access control management, and Tdata, for data storage (read or write).

Every time a new user signs up, a new shared user-id variable is generated and sent, including the associated permissions, to the blockchain in a Taccess transaction, for later identification if he wishes to sign in. Personal user data such as location, username, gender are encrypted using a shared encryption key and sent to the blockchain in a Tdata transaction, which afterwards are routed to a hashtable using a key-value pair, while nodes retaining only a pointer to the data (the pointer is the SHA-256 hash of the data). Both the service and the user can now access the data using a Tdata transaction with the pointer associated to it. The blockchain then verifies that the digital signature belongs to either the user or the service. Finally, the user can change the permissions granted to a service at any time by issuing a Taccess transaction with a new set of permissions, including the restriction on accessing data that he previously stored.

Personal user data may not be trusted in the hands of third parties, as there can be vulnerabilities to attacks and misuse. Users should own and control their data without negotiating security issues, and at the same time not limiting companies to provide personalized services. Using the framework above, users have the option to not trust any third-party and are always aware of the data that is being collected about them and how it is used. In addition, the blockchain recognizes the users as the owners of their personal data. Using this approach, users are ensuring that their personal data are accessible by first parties only, or third parties that users have authorized before and clearly consented with.

As the provisioning of privacy against third party data trackers is now an absolute strategy, even for simple browsing, there are anti-measures taken by platforms like Google. Google has ensured that on their own browser, Google Chrome (**Figure 32**), there are tools for blocking first-party cookies ,which are used to store user preferences for a particular site, as well as third-party tracking cookies which record users online activity across sites.



← Cookies and site data

Search

Allow sites to save and read cookie data (recommended) ●

Clear cookies and site data when you quit Chrome ○

Block third-party cookies
Prevent third-party websites from saving and reading cookie data ●

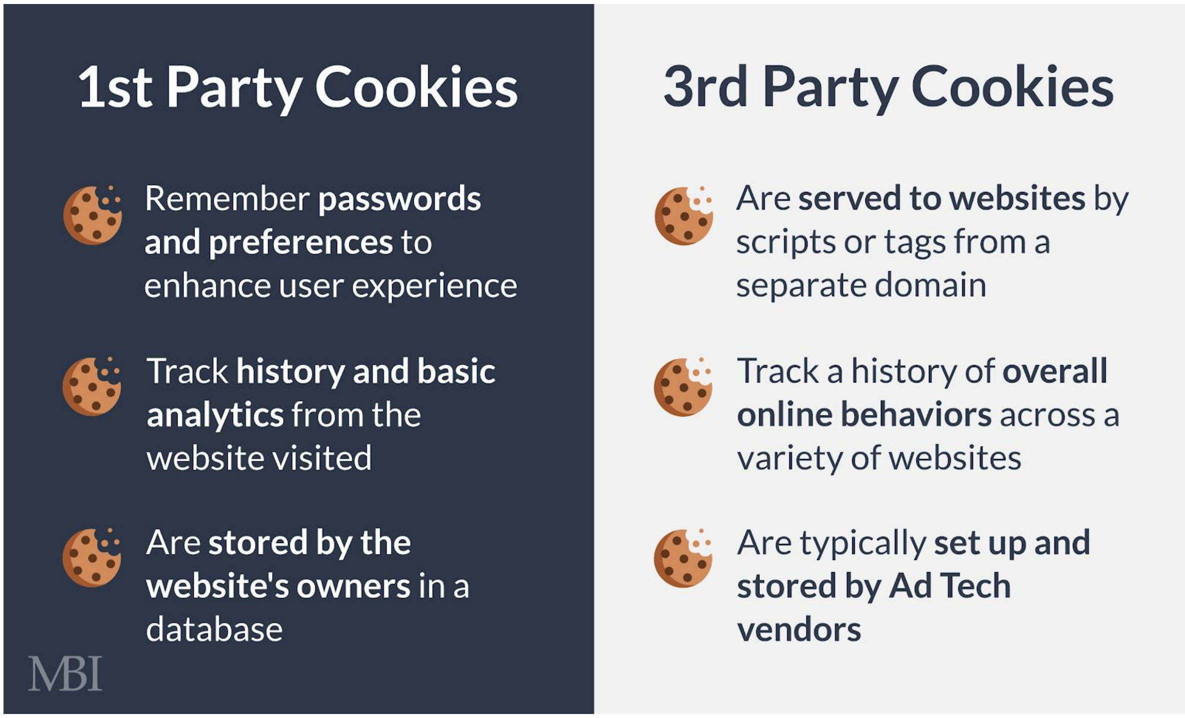
See all cookies and site data ▸

Figure 32 – Google Chrome tools for blocking third-party cookies.

Google also stated that Chrome will phase out third-party cookies until 2022 in favor of an alternative system to cookies using innovative technology that Google is developing.

5.2 Online Advertising without 3rd Party cookies

In January 2020, Google Chrome announced their intention to phase out all third-party cookies within two years. Google Chrome is utilized by 69% of the global internet users, and with an advertising revenue stream, they are walking a tight line to satisfy consumer privacy needs and meet their revenue goals. This was following previous announcements by browsers such as Safari and Firefox who had already outlined plans for cookie restrictions and new privacy regulations. These moves are an attempt to respect consumer privacy in an increasingly regulated environment.



1st Party Cookies

- Remember **passwords and preferences** to enhance user experience
- Track **history and basic analytics** from the website visited
- Are **stored by the website's owners** in a database

3rd Party Cookies

- Are **served to websites** by scripts or tags from a separate domain
- Track a history of **overall online behaviors** across a variety of websites
- Are typically **set up and stored by Ad Tech vendors**

Figure 33 – Google Chrome tools for blocking third-party cookies.

5.2.1 The main implications of this change

There are many concrete advantages of the use of third-party cookies. Consequently, a cookie less world will have many impacts to various stakeholders. It is important to reflect on the website experience in the past before the use of third-party cookies that companies were struggling to distinguish between the new and recurrent customers. For customers, there is a clear benefit of third-party cookies phase out, because of the ability to control the personal data to share with the advertisers. In addition to that, it will be visible to them which companies specifically collect their behavior data, because they need to give their consent before the start of any form of data collection. As a result, unexpected data sharing cases by companies will eventually come to an end. On the contrary, there are also some negative impacts, since with the fade off third-party cookies, customers will not be able to get any personalized experiences. This will have therefore, to receive fewer interactive ads with no particular interest to their customers' individual needs. The only personalization strategy that will remain available, is the remarketing capabilities if they are owned by the brand itself.

Besides, for global internet users there will be an increased thoroughness for the intrusion of privacy while browsing online. As a result, internet users will be able to develop a general trust and confidence about the use and the recipients of their personal data.Additionally, the implication of blocking 3rd party cookies will have an immediate effect on all the parts of programmatic advertising since for more than 2 decades advertisers were relying on them for various reasons. Scalability issues will also be an outcome for advertisements, since the capabilities will be limited to only first party data. This indicates that advertisement companies need to explore new strategy formulation for relying on different marketing tactics. Finally, audience segmentation for better campaign targeting will be a challenge before companies will learn to adjust and find alternatives.

Moreover, the ban of third-party data collection will directly result in short term issues for publishers and demand side platforms (DSPs), such as loss of revenue and downturn freeze of digital marketing budget. However, in the long-term publishers will be able to adapt if they manage to remodel their approach and create new strategies for revenue generation, taking advantage of the audience data they produce. DSP vendors may need to change their digital product positioning, which currently relies on the various third-party data available for retargeting and personalization. Therefore, the past year strong media publishers such as Google and Facebook offer first-party audience without a cost in order to develop a new way of working within their media buying platform.

5.2.2 How to manage the change

As the countdown of 3rd party cookies has started, with Google Chrome being the last one from those who will end support them, a new question is burned. What will be the future in online advertising without 3rd party cookies and which techniques and methods will replace them? 3rd party cookies are one of the possible targeting techniques, some others are already here while others are still emerging.

Therefore, companies need to develop new strategies and evolve in order to meet the new custom needs:

- First and foremost, amplify company’s owned cookies by improving the use of first-party data collection. A precise, careful and thorough implementation of data collections mechanisms will ensure that reliable data will be available as granular input for targeting and personalization of users who have already engaged with the brand.

- Enable best-in-class cookies permission solutions. It is important to make sure that there is a guarantee that first-party data is fully compliant with legislation. This can be achieved by creating communicating strategies on how they manage and secure data to boost customer trust.

- Discover second-party data from technology leaders and publishers. Facebook and Google offer aggregated but granular audiences created by the platform in use. In addition to that,

- Expand the application of personalization strategies. Deep-personalization campaigns with smart geo-targeting and time parting, also known as moment marketing.

- Explore non-cookie-based third-party audiences created for targeting purposes using a transparent way. This will result in an overall shift from highly scalable third-party data that will be based on relationships.

- Another solution could be context mapping that maps various audience segments on the open marketplace which would be created with publishers first party data. This solution is similar to Google's federated learning of cohort's proposal which would use machine learning to l better understand the browsing habits of similar users and then will group them together.

- Leverage Google Privacy Sandbox solution proposal that transfers the storing and processing of customers data into the browser. But also create a better open dialogue between Google and those who function within the advertising space to discuss and test the third-party cookie alternatives and data privacy regulations that benefit users' advertisers and publishers

- Use of emerging technology solutions to help advertisers be more independent on the use of third-party cookies. An example will be the investment of extensively in privacy-compliance, demographic-based IDs that provide a consistent and accurate view of a customer.

- At browser or device level fingerprinting solutions will incorporate all public, IP address and general information into one single ID. Unlike third-party cookies collection, it is not necessary for customers to receive any alert. This solution will be impossible for customers to avoid manually since with any browser search, the information stored in the fingerprint is publicly distributed.

6. DETECTING AND BLOCKING THIRD PARTY TRACKERS

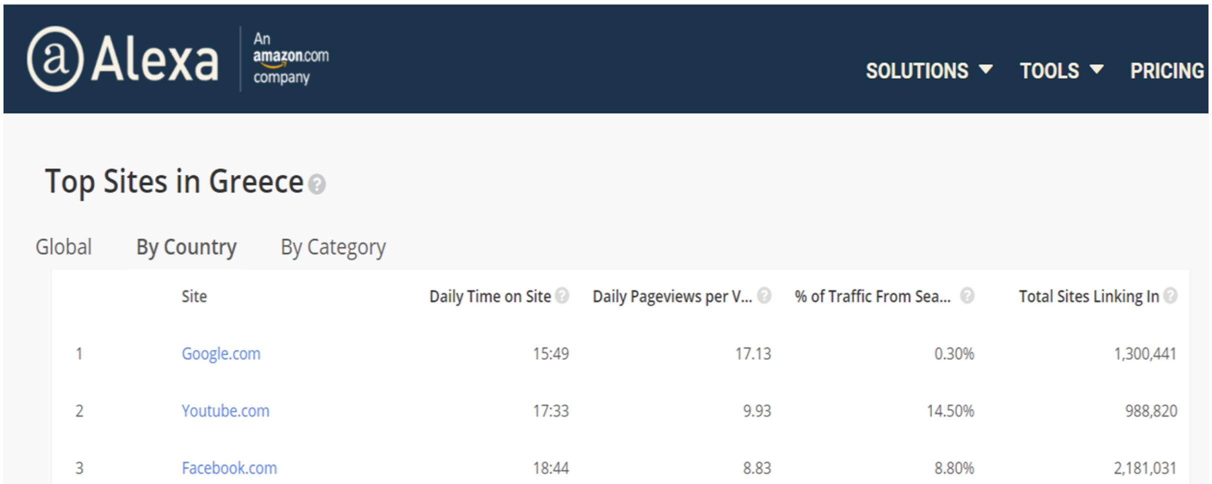
6.1 Software & Tools to Identify Third Party Trackers

6.1.1 WHO IS TRACKING ME EXTENSION

In the previous chapters, an extensive analysis has been made about the online advertising, the benefits for the users but also a detailed description about all the privacy issues and the upcoming threats for the e-society. Once there is one third-party entity on a page, that entity can turn around and invite any number of other third-parties to the first-party webpage.

In order to imprint and prove all the details that were discussed in the whole thesis, we ran some detection systems while browsing the web and observing a rich ecosystem, exploring some popular websites in Greece and all over the world, according to Amazon Alexa software, and at the same time used our tools to explore the third party trackers that collect our data on mobile applications.

Initially, we focused on the top 3 Greek websites from Alexa and top 3 global websites, and an extension of Google Chrome, named" Who **is Tracking Me"** that detects the third-party entities at the time a user visits the site. **In Figure 34,** Top 3 Global websites across by Alexa, are being displayed**.**

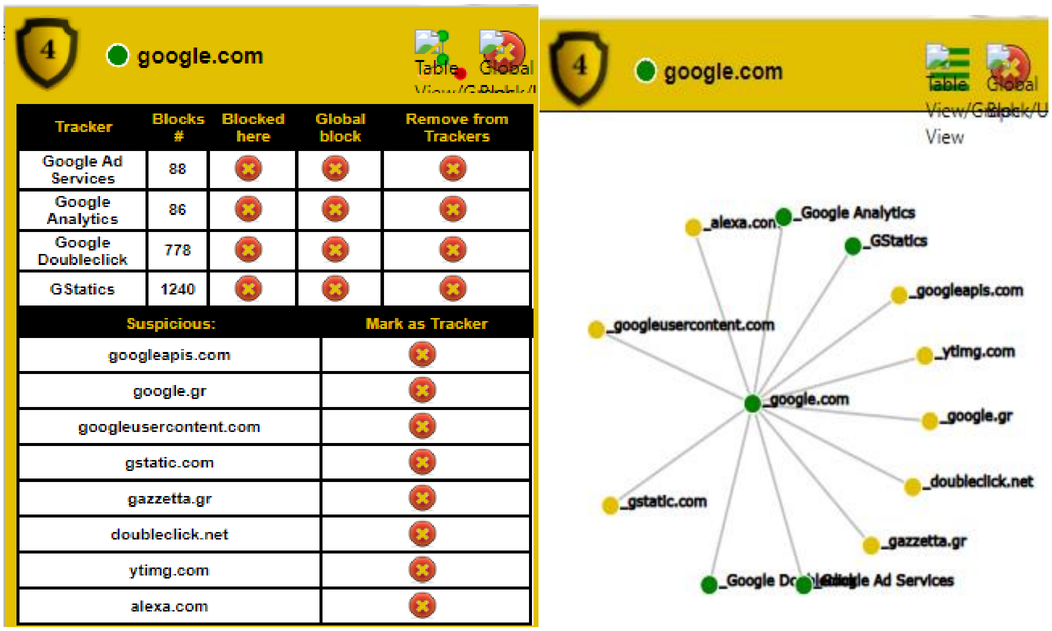


Top Sites in Greece

	Site	Daily Time on Site	Daily Pageviews per V...	% of Traffic From Sea...	Total Sites Linking In
1	Google.com	15:49	17.13	0.30%	1,300,441
2	Youtube.com	17:33	9.93	14.50%	988,820
3	Facebook.com	18:44	8.83	8.80%	2,181,031

Figure 34 – Top 3 Greek websites by Amazon Alexa.

Using the Google Chrome extension "Who is Tracking me" on Google Chrome, after visiting google.com, the first listed website, the results are provided on the scheme in **Figure 35**.



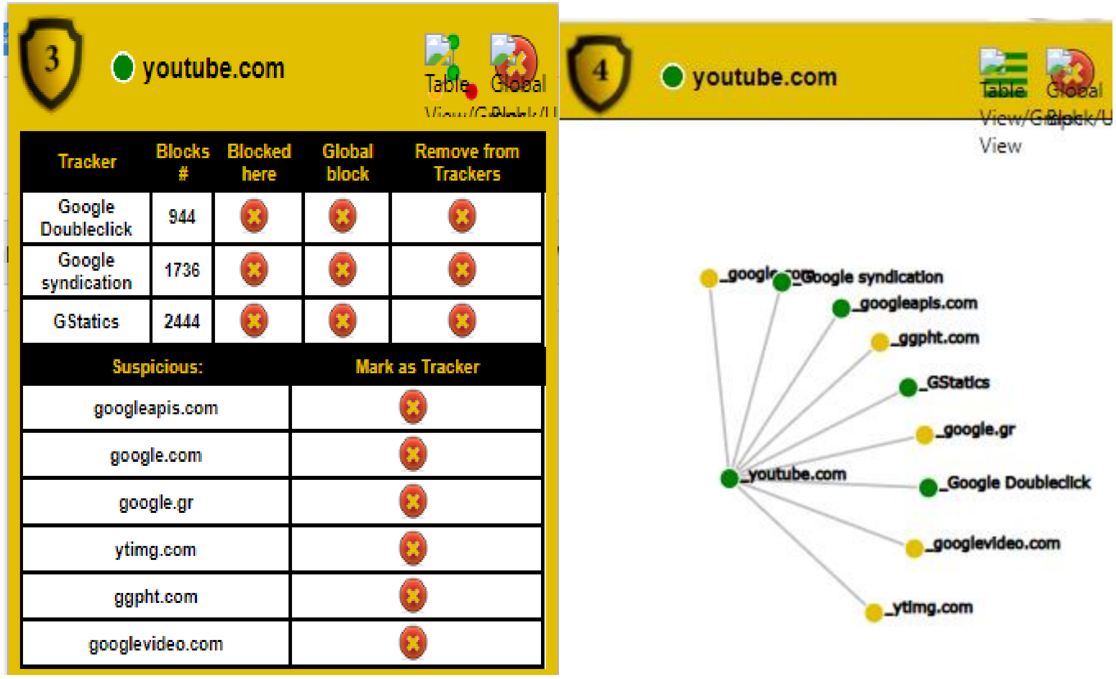
Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Google Ad Services	88	⊗	⊗	⊗
Google Analytics	86	⊗	⊗	⊗
Google Doubleclick	778	⊗	⊗	⊗
GStatics	1240	⊗	⊗	⊗

Suspicious:	Mark as Tracker
googleapis.com	⊗
google.gr	⊗
googleusercontent.com	⊗
gstatic.com	⊗
gazzetta.gr	⊗
doubleclick.net	⊗
ytimg.com	⊗
alexa.com	⊗

Figure 35 –Google Who is Tracking Me Extensions results 1.0.

Google services (Google Ad Services, Google Analytics, Google DoubleClick Statics) by majority are the ones that track users while searching for information in Google (Figure 35). On the other hand, there are some suspicious or candidate websites that may take a closer look on user data. One of them is Alexa itself. On the second column of the scheme, there is a graph with the confirmed trackers that are being featured by green color, and the yellow color is used for suspicious activity that perhaps tracked our footprints at a time.

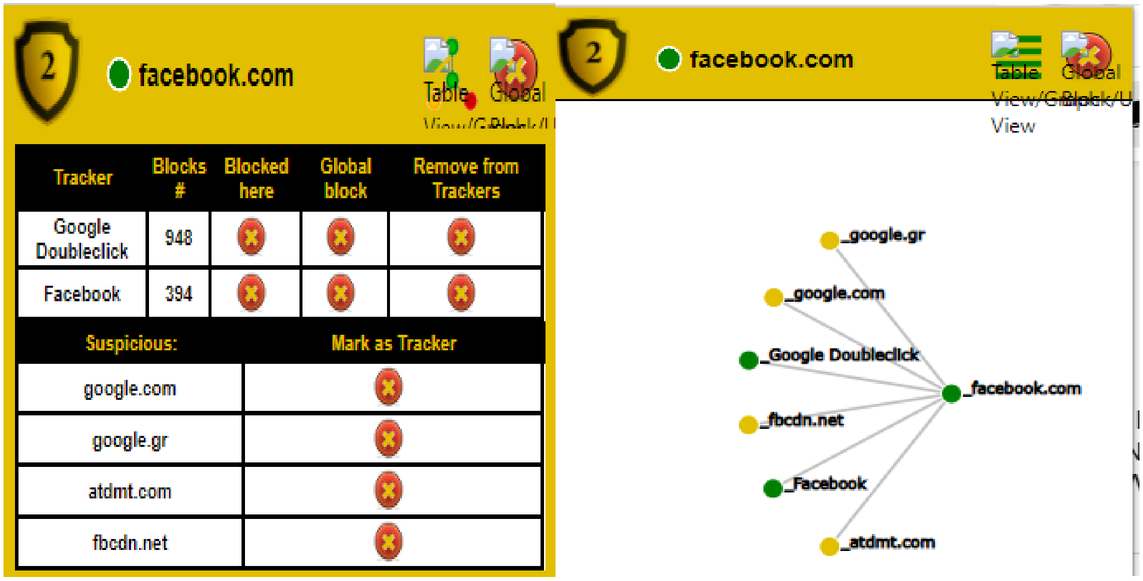
The results of the two other websites (YouTube & Facebook) of Alexa's list, are also being displayed **in Figure 36**, and **Figure 37.** It is very clear that Google's trackers are involved on every commercial and top visited Greek section and the 2 out of 3 Global websites.



Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Google Doubleclick	944	☒	☒	☒
Google syndication	1736	☒	☒	☒
GStatics	2444	☒	☒	☒

Suspicious:	Mark as Tracker
googleapis.com	☒
google.com	☒
google.gr	☒
ytimg.com	☒
ggpht.com	☒
googlevideo.com	☒

Figure 36 – YouTube Who is Tracking Me Extensions results 1.0.

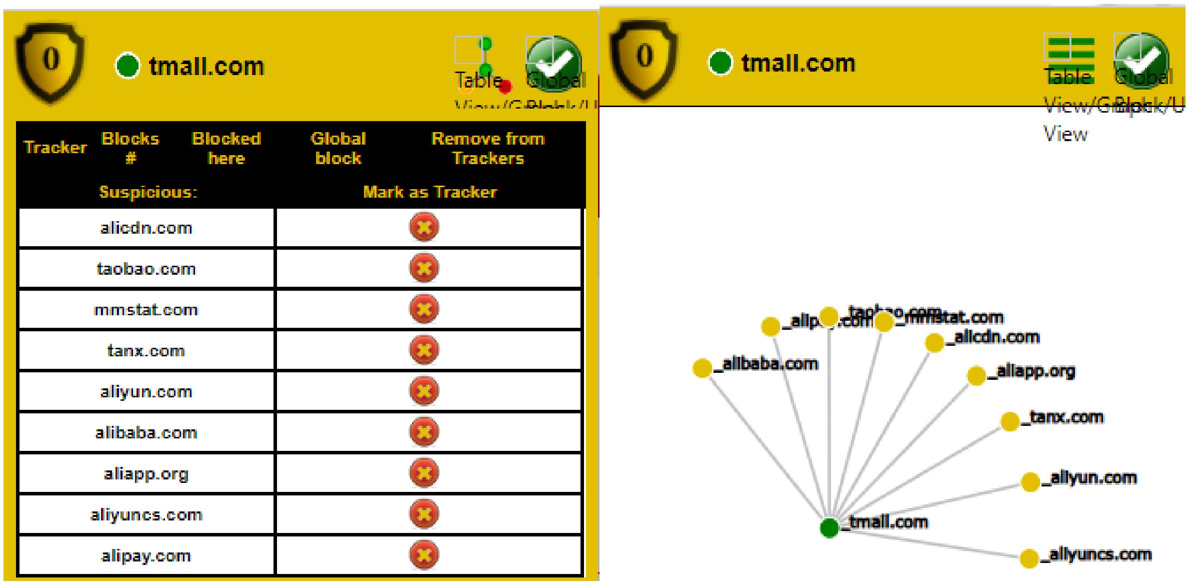


Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Google Doubleclick	948	☒	☒	☒
Facebook	394	☒	☒	☒

Suspicious:	Mark as Tracker
google.com	☒
google.gr	☒
atdmt.com	☒
fbcdn.net	☒

Figure 36 – Facebook Who is Tracking Me Extensions results 1.0.

Finally, on the top third global website named Tmall.com, a Chinese online retail shop, Google has not achieved to involve their own trackers yet, but there are only some suspicious not belonging to Google's services. It seems that Chinese market has achieved to limit the user tracking by Google.

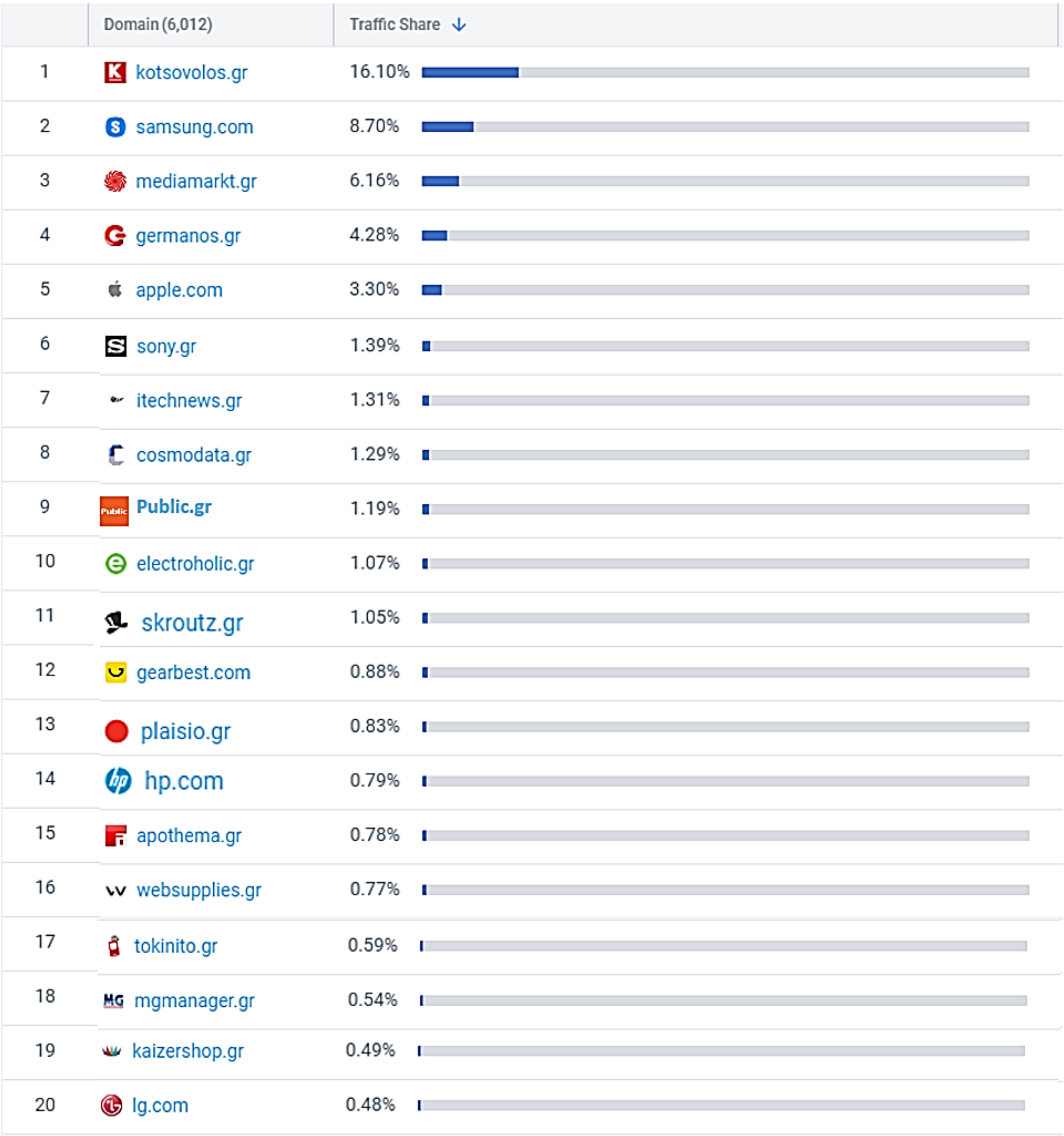


Suspicious:	Mark as Tracker
alicdn.com	☒
taobao.com	☒
mmstat.com	☒
tanx.com	☒
aliyun.com	☒
alibaba.com	☒
aliapp.org	☒
aliyuncs.com	☒
alipay.com	☒

Figure 36 – Tmall.com Who is Tracking Me Extensions results 1.0.

6.1.2 COOKIESERVE – COOKIES SCANNING TOOL

Furthermore, we also have done a detailed analysis, using another online tool named Cookieserve Online scanner tool, which is a website scanning tool that scans a website for cookies and generates a cookie scan report containing detailed information. Using the Similar web tool top 20 Greek websites have been extracted for Consumer Electronics Technology industry (Figure 37). The scope of this analysis was to analyze the number of cookies being tracked for each website but also the coverage of first party versus third party cookies.



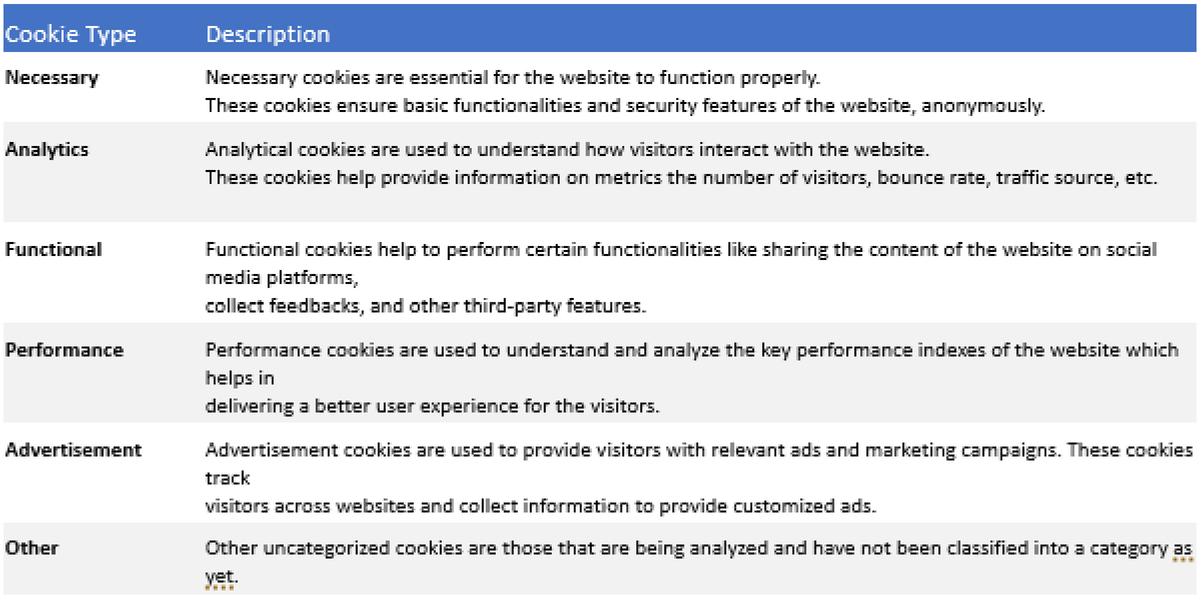
	Domain (6,012)	Traffic Share ↓
1	kotsovolos.gr	16.10%
2	samsung.com	8.70%
3	mediamarkt.gr	6.16%
4	germanos.gr	4.28%
5	apple.com	3.30%
6	sony.gr	1.39%
7	itechnews.gr	1.31%
8	cosmodata.gr	1.29%
9	Public.gr	1.19%
10	electroholic.gr	1.07%
11	skroutz.gr	1.05%
12	gearbest.com	0.88%
13	plaisio.gr	0.83%
14	hp.com	0.79%
15	apothema.gr	0.78%
16	websupplies.gr	0.77%
17	tokinito.gr	0.59%
18	mgmanager.gr	0.54%
19	kaizershop.gr	0.49%
20	lg.com	0.48%

Figure 37 – Top 20 Greek Websites in Consumer Electronics Industry

For each website, a list of the following datapoints have been collected and analyzed into a database using Cookieserve online tool:

- **Website URL**
- **Cookie name**
- **Cookie Domain**
- **Description** of use, why this cookie is being used

- **Duration** amount of time that cookie is running on each website
- **Cookie Type** as displayed in Figure 38
- **1st/3rd Party,** manual calculated column that check whether website equals domain name, or whether it is a third party because website does not equal domain name

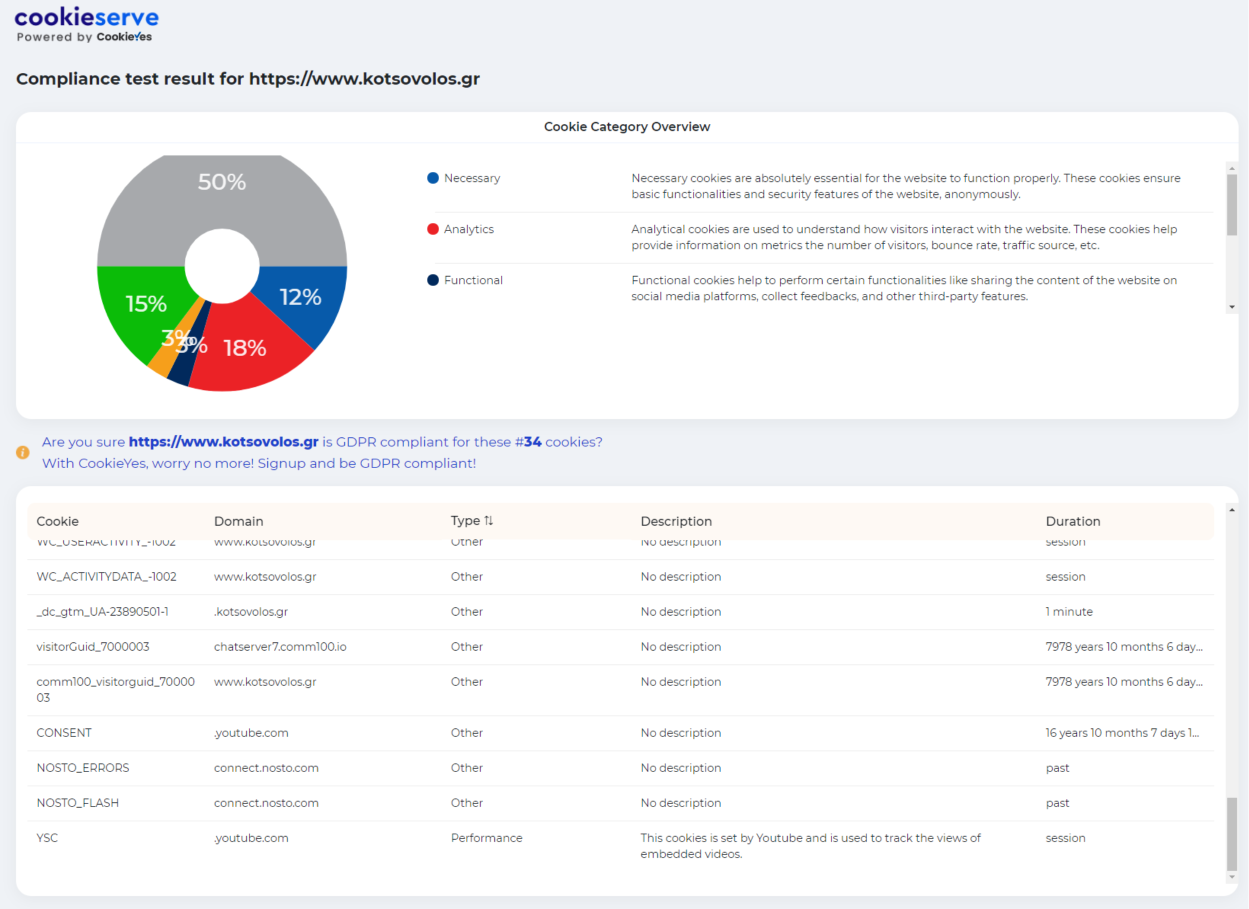


Cookie Type	Description
Necessary	Necessary cookies are essential for the website to function properly. These cookies ensure basic functionalities and security features of the website, anonymously.
Analytics	Analytical cookies are used to understand how visitors interact with the website. These cookies help provide information on metrics the number of visitors, bounce rate, traffic source, etc.
Functional	Functional cookies help to perform certain functionalities like sharing the content of the website on social media platforms, collect feedbacks, and other third-party features.
Performance	Performance cookies are used to understand and analyze the key performance indexes of the website which helps in delivering a better user experience for the visitors.
Advertisement	Advertisement cookies are used to provide visitors with relevant ads and marketing campaigns. These cookies track visitors across websites and collect information to provide customized ads.
Other	Other uncategorized cookies are those that are being analyzed and have not been classified into a category as yet.

Figure 38 – Cookieserve Cookies type definitions

Furthermore, an example of the data being collected for each website is displayed in **Figure 39**. All these data have been inserted into an excel file and using Microsoft Power Bi platform have been analyzed. More specifically, what we analyzed were:

- Number of cookies of each website (**Figure 40**)
- Count of Cookies for all website by Cookie Type (**Figure 41**)
- Number of 1st party vs 3rd party cookie for each website (**Figure 42**)
- Overall split of 1st party vs 3rd party cookie for all websites (**Figure 43**)
- Count of most common 3rd party domains tracked for all websites (**Figure 44**)



cookieserve
Powered by CookieYes

Compliance test result for https://www.kotsovolos.gr

Cookie Category Overview

● Necessary — Necessary cookies are absolutely essential for the website to function properly. These cookies ensure basic functionalities and security features of the website, anonymously.

● Analytics — Analytical cookies are used to understand how visitors interact with the website. These cookies help provide information on metrics the number of visitors, bounce rate, traffic source, etc.

● Functional — Functional cookies help to perform certain functionalities like sharing the content of the website on social media platforms, collect feedbacks, and other third-party features.

Are you sure **https://www.kotsovolos.gr** is GDPR compliant for these #**34** cookies?
With CookieYes, worry no more! Signup and be GDPR compliant!

Cookie	Domain	Type ⇅	Description	Duration
WC_USERACTIVITY_-1002	www.kotsovolos.gr	Other	No description	session
WC_ACTIVITYDATA_-1002	www.kotsovolos.gr	Other	No description	session
_dc_gtm_UA-23890501-1	.kotsovolos.gr	Other	No description	1 minute
visitorGuid_7000003	chatserver7.comm100.io	Other	No description	7978 years 10 months 6 day...
comm100_visitorguid_7000003	www.kotsovolos.gr	Other	No description	7978 years 10 months 6 day...
CONSENT	.youtube.com	Other	No description	16 years 10 months 7 days 1...
NOSTO_ERRORS	connect.nosto.com	Other	No description	past
NOSTO_FLASH	connect.nosto.com	Other	No description	past
YSC	.youtube.com	Performance	This cookies is set by Youtube and is used to track the views of embedded videos.	session

Figure 39 – Cookieserve Example Result for Kotsovolos.gr

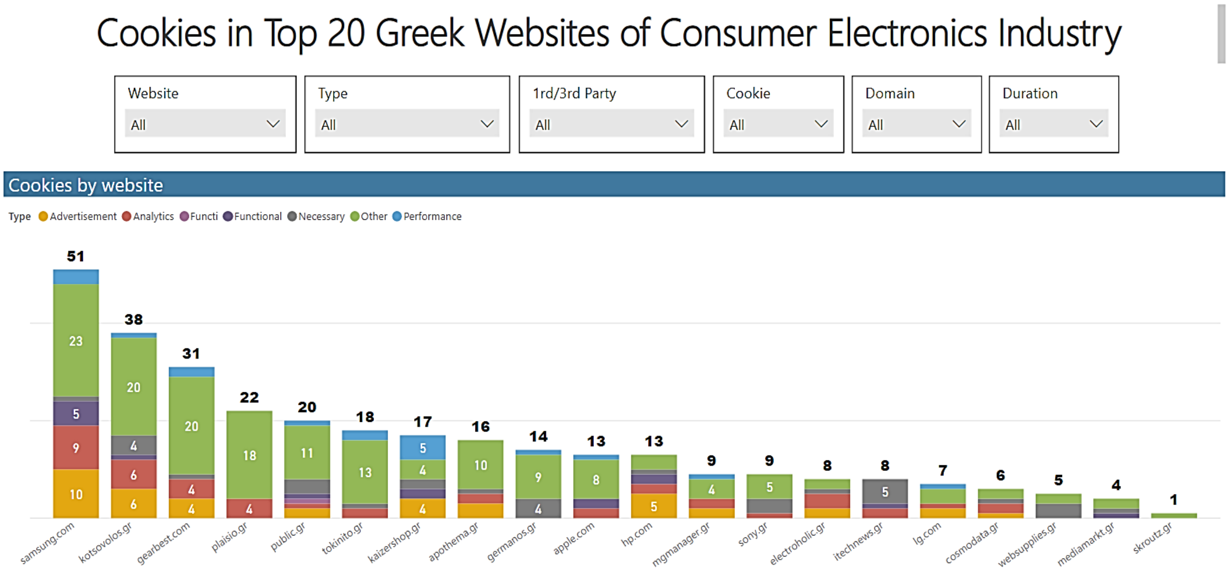


Figure 40 – Number of cookies by website

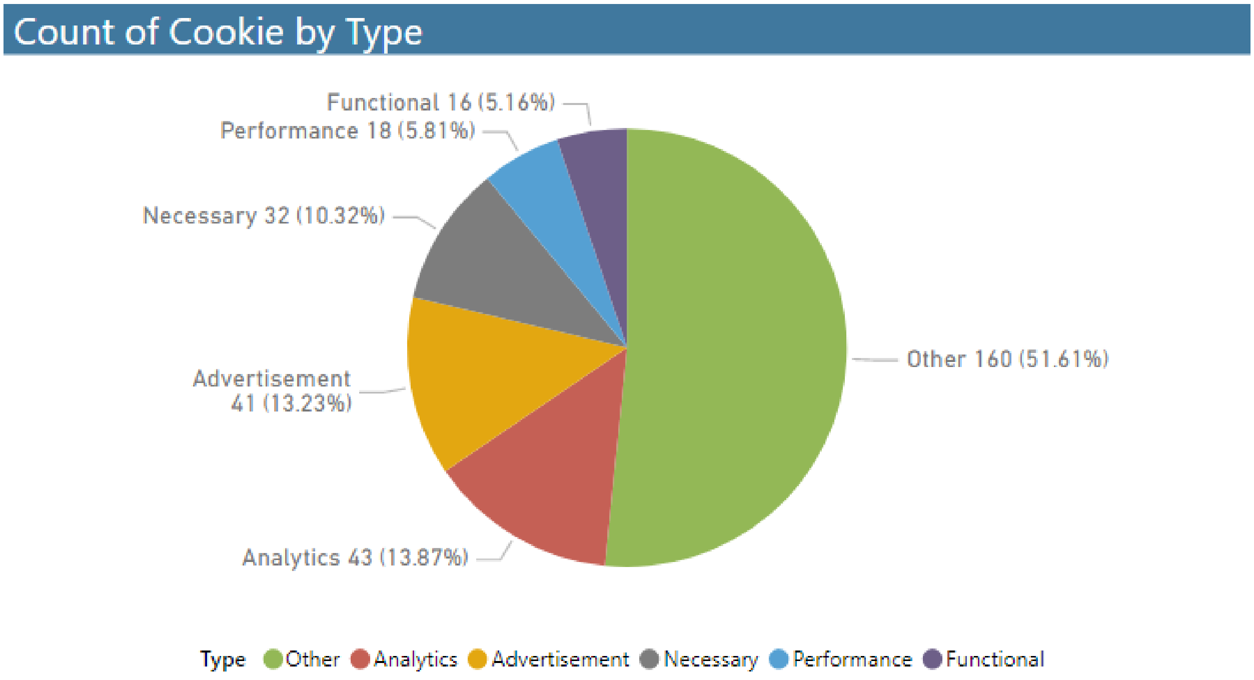


Figure 41 – Count of Cookies by Type for all websites

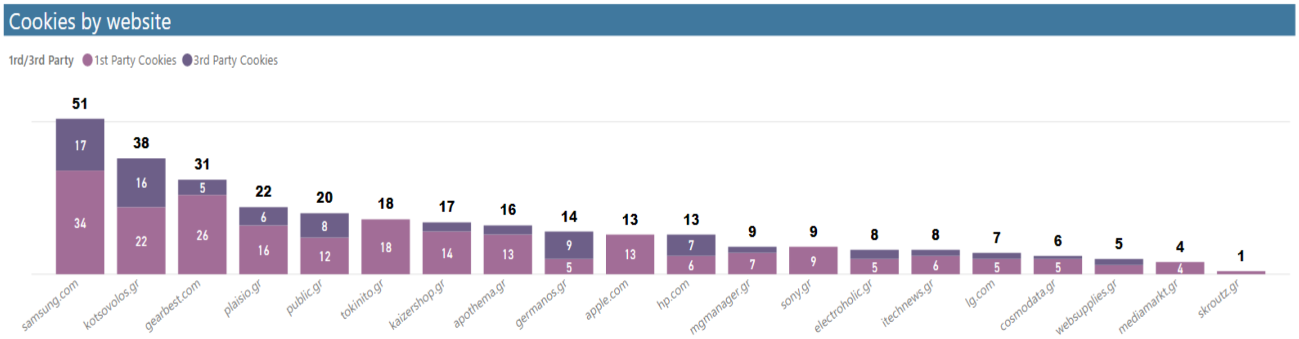


Figure 42 – Count of 1st party vs 3rd party data for each website

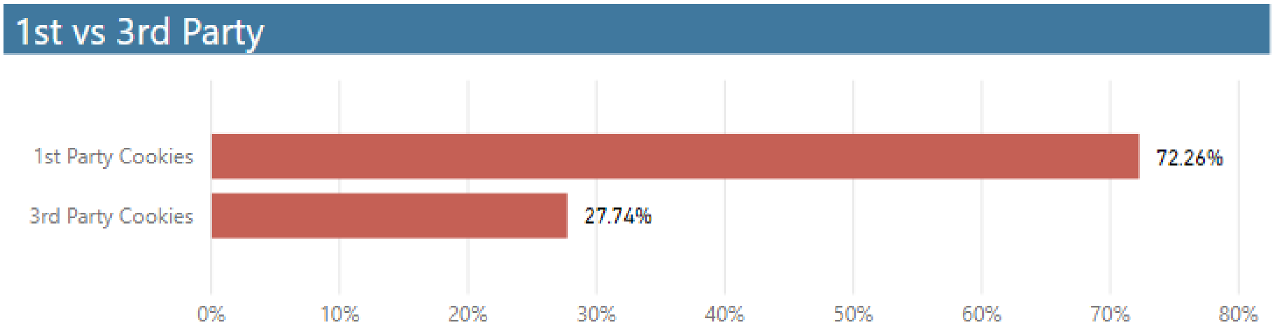


Figure 43 - 1st party vs 3rd party data for all websites

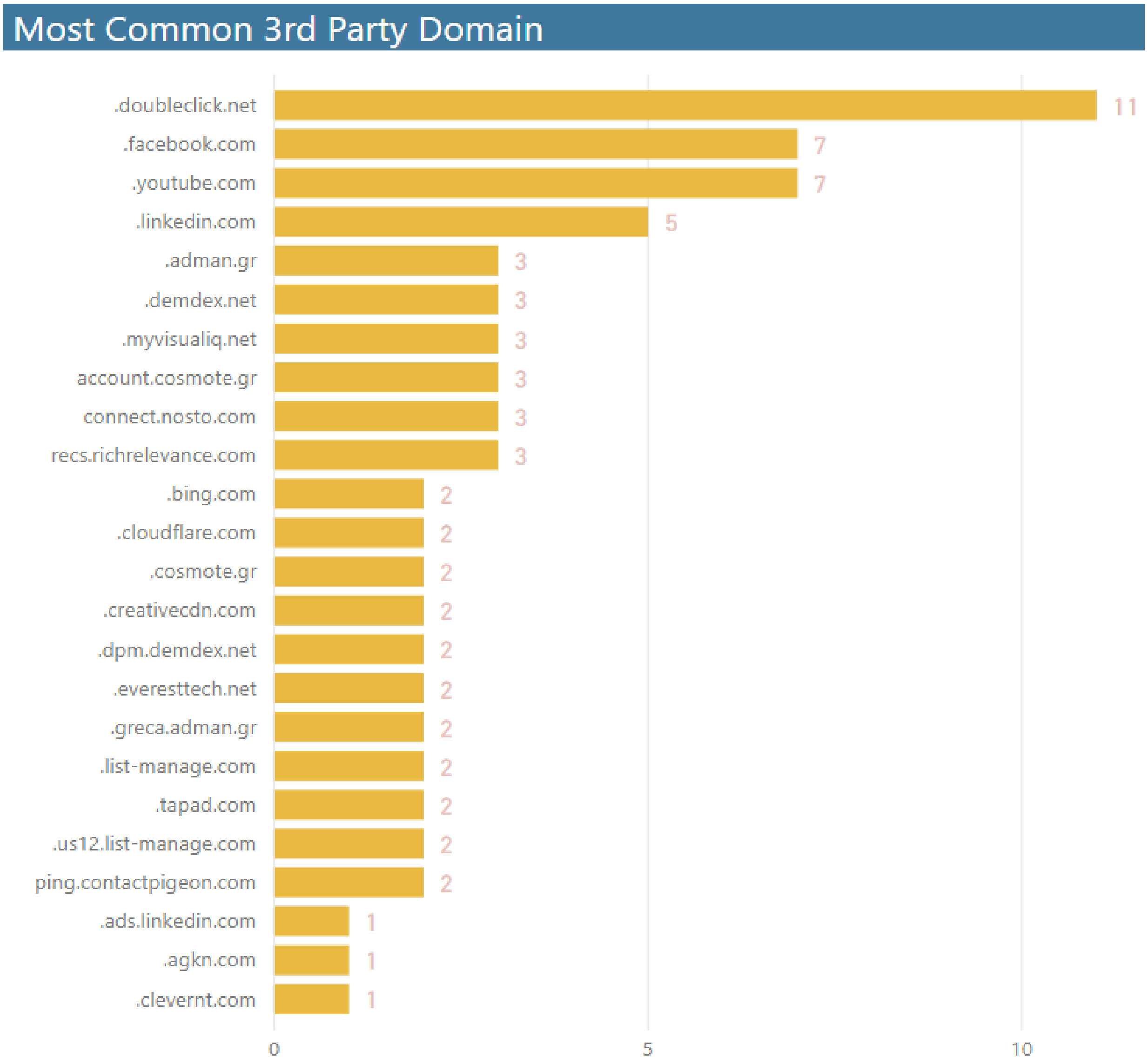
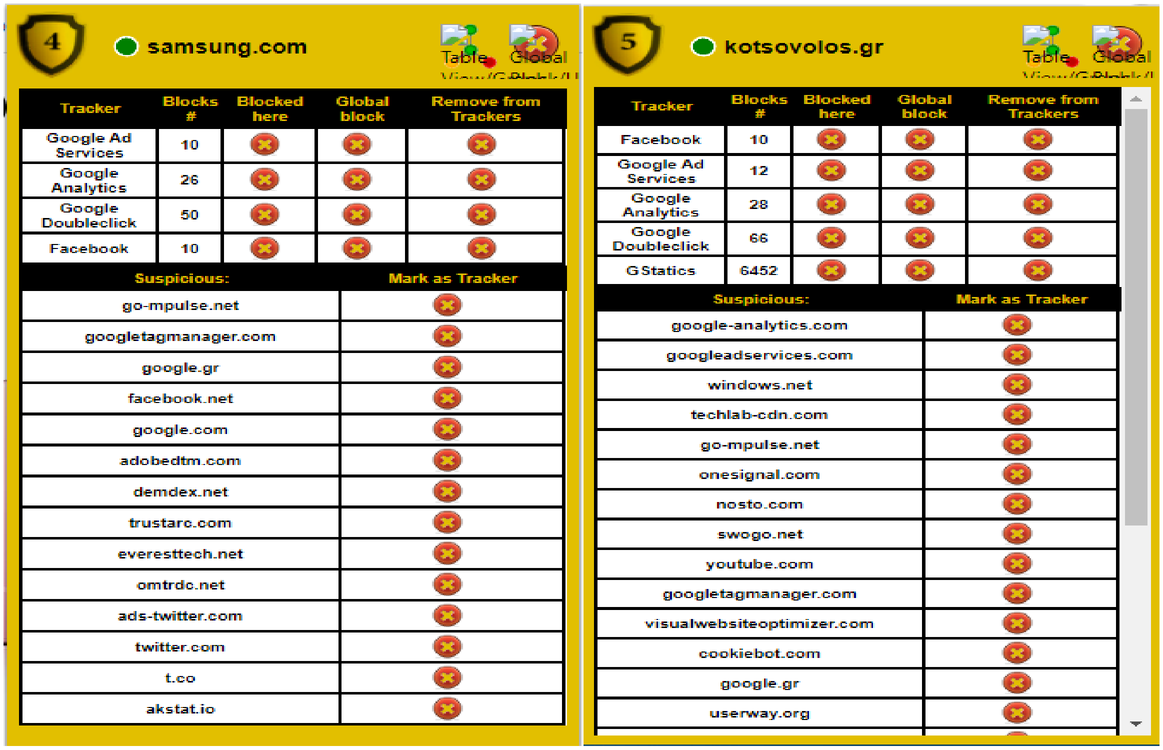


Figure 44 – Most Common 3rd party Domains for all websites

6.2 User tracking without consent

Using the results of previous analysis, we focused on the 10 top websites, shown in **Figure 37.** The scope of this analysis was to collect user traces on these websites, and whether there was a consent about user data collection or not. All the data below are collected without pressing “I consent” or “I accept cookies”. Active trackers are marked with X and blocked trackers are marked with ✔.

On the top 2 Greek Consumer Electronic websites, Samsung and Kotsovolos.gr.gr, as presented in **Figure 45** we found four and five active trackers respectively and a lot of suspicious inactive domains. It is clear from **Figure 45** that without any user consent, Google’s and Facebook’s trackers had been activated, bypassing user’s rights.



4 samsung.com

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Google Ad Services	10	✖	✖	✖
Google Analytics	26	✖	✖	✖
Google Doubleclick	50	✖	✖	✖
Facebook	10	✖	✖	✖

Suspicious:	Mark as Tracker
go-mpulse.net	✖
googletagmanager.com	✖
google.gr	✖
facebook.net	✖
google.com	✖
adobedtm.com	✖
demdex.net	✖
trustarc.com	✖
everesttech.net	✖
omtrdc.net	✖
ads-twitter.com	✖
twitter.com	✖
t.co	✖
akstat.io	✖

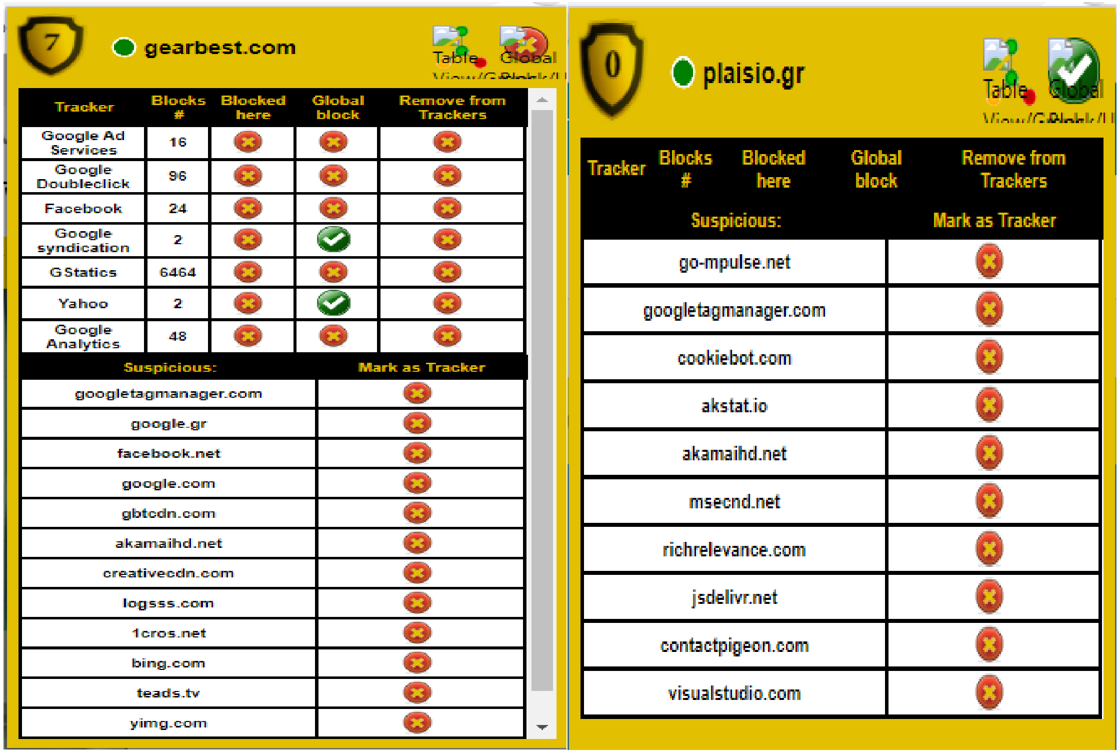
5 kotsovolos.gr

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Facebook	10	✖	✖	✖
Google Ad Services	12	✖	✖	✖
Google Analytics	28	✖	✖	✖
Google Doubleclick	66	✖	✖	✖
GStatics	6452	✖	✖	✖

Suspicious:	Mark as Tracker
google-analytics.com	✖
googleadservices.com	✖
windows.net	✖
techlab-cdn.com	✖
go-mpulse.net	✖
onesignal.com	✖
nosto.com	✖
swogo.net	✖
youtube.com	✖
googletagmanager.com	✖
visualwebsiteoptimizer.com	✖
cookiebot.com	✖
google.gr	✖
userway.org	✖

Figure 45 – Cookies tracked without user consent for Samsung and Kotsovolos.gr

On the other hand, Gearbest.com and Plaisio.gr.gr have limited the active trackers without users' consent. Gearbest.com has blocked 2 out of 7 trackers and waits for user’s approval. Plaisio.gr.gr have not allowed any tracker to be involved in the process before user consents as shown in Figure 46.



gearbest.com

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Google Ad Services	16	✗	✗	✗
Google Doubleclick	96	✗	✗	✗
Facebook	24	✗	✗	✗
Google syndication	2	✗	✓	✗
GStatics	6464	✗	✗	✗
Yahoo	2	✗	✓	✗
Google Analytics	48	✗	✗	✗

Suspicious:	Mark as Tracker
googletagmanager.com	✗
google.gr	✗
facebook.net	✗
google.com	✗
gbtcdn.com	✗
akamaihd.net	✗
creativecdn.com	✗
logsss.com	✗
1cros.net	✗
bing.com	✗
teads.tv	✗
yimg.com	✗

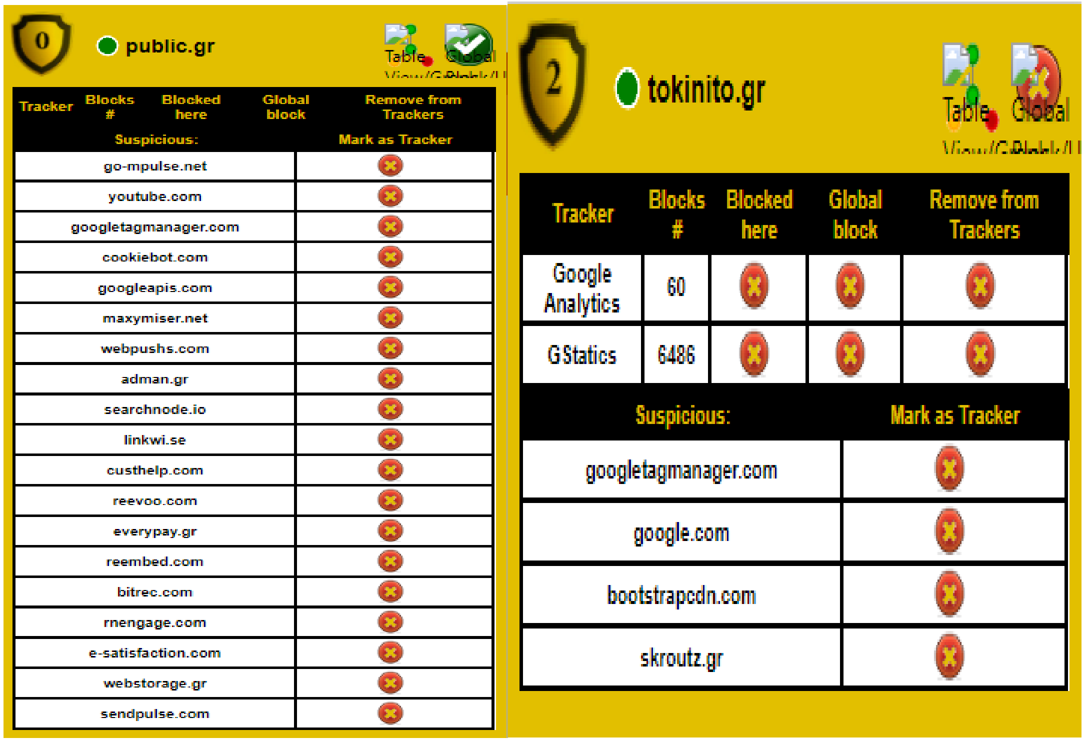
plaisio.gr

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers

Suspicious:	Mark as Tracker
go-mpulse.net	✗
googletagmanager.com	✗
cookiebot.com	✗
akstat.io	✗
akamaihd.net	✗
msecnd.net	✗
richrelevance.com	✗
jsdelivr.net	✗
contactpigeon.com	✗
visualstudio.com	✗

Figure 46 – Cookies tracked without user consent for Gearbest.com and Plaisoi.gr

In addition, Public.gr did not allow any tracker to be involved in the process prior to user consent. On the other side, tokinito.gr have activated Google Analytics and G statistics without letting the user know, as presented in **Figure 47.**



public.gr

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers

Suspicious:	Mark as Tracker
go-mpulse.net	✗
youtube.com	✗
googletagmanager.com	✗
cookiebot.com	✗
googleapis.com	✗
maxymiser.net	✗
webpushs.com	✗
adman.gr	✗
searchnode.io	✗
linkwi.se	✗
custhelp.com	✗
reevoo.com	✗
everypay.gr	✗
reembed.com	✗
bitrec.com	✗
rnengage.com	✗
e-satisfaction.com	✗
webstorage.gr	✗
sendpulse.com	✗

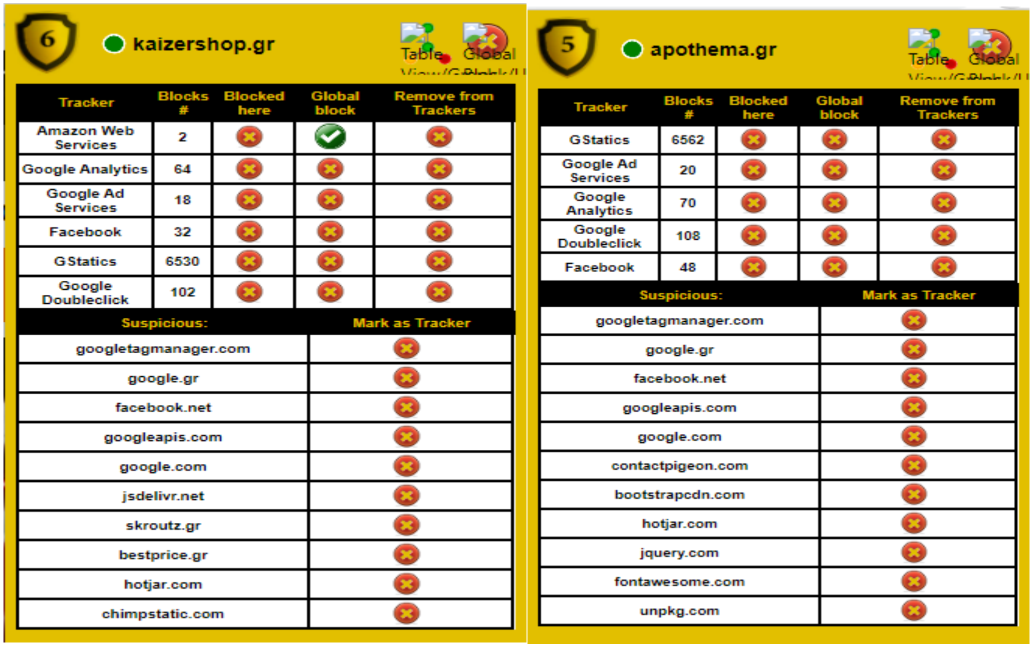
tokinito.gr

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Google Analytics	60	✗	✗	✗
GStatics	6486	✗	✗	✗

Suspicious:	Mark as Tracker
googletagmanager.com	✗
google.com	✗
bootstrapcdn.com	✗
skroutz.gr	✗

Figure 46 – Cookies tracked without user consent for Public and tokinito.gr

Kaizershop.gr and Apothema.gr, as shown **in Figure 47** both have the allowed trackers without any user consent. Kaizershop.gr has allowed Amazon web services, Google Advertising and Analytics, G statistics and Facebook. Apothema.gr on the other hand has also allowed similar trackers without having users' consent.



Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
Amazon Web Services	2	✗	✓	✗
Google Analytics	64	✗	✗	✗
Google Ad Services	18	✗	✗	✗
Facebook	32	✗	✗	✗
GStatics	6530	✗	✗	✗
Google Doubleclick	102	✗	✗	✗

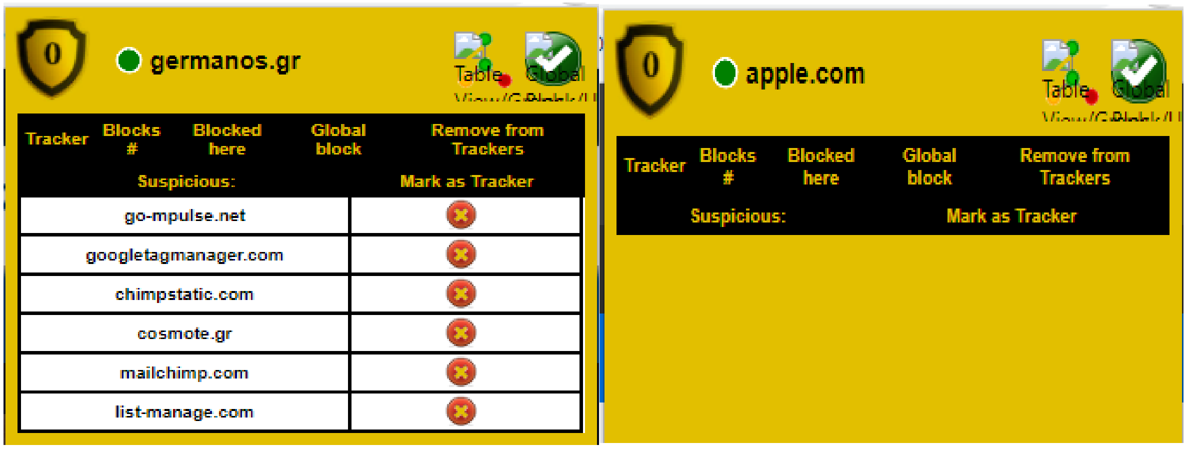
Suspicious:	Mark as Tracker
googletagmanager.com	✗
google.gr	✗
facebook.net	✗
googleapis.com	✗
google.com	✗
jsdelivr.net	✗
skroutz.gr	✗
bestprice.gr	✗
hotjar.com	✗
chimpstatic.com	✗

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers
GStatics	6562	✗	✗	✗
Google Ad Services	20	✗	✗	✗
Google Analytics	70	✗	✗	✗
Google Doubleclick	108	✗	✗	✗
Facebook	48	✗	✗	✗

Suspicious:	Mark as Tracker
googletagmanager.com	✗
google.gr	✗
facebook.net	✗
googleapis.com	✗
google.com	✗
contactpigeon.com	✗
bootstrapcdn.com	✗
hotjar.com	✗
jquery.com	✗
fontawesome.com	✗
unpkg.com	✗

Figure 47 – Cookies tracked without user consent for Kaizershop.gr and Apothema.gr

Lastly, Germanos.gr have not activated any trackers, but apple have also blocked any suspicious third-party entity and can be defined as the most trustworthy website. Apple is broadly known to be strict with privacy issues and user sensitive data handling. In conclusion, the majority of the Greek commercial websites are collecting user data with active trackers without user's consent, ignoring the explicit GDPR European law and without any consequences. Although, there are some exceptions that do not activate any tracker without user's assent and even not allowing any suspicious entities, following literally the General Data Protection Regulation and respecting user's freedom.



Tracker	Blocks #	Blocked here	Global block	Remove from Trackers

Suspicious:	Mark as Tracker
go-mpulse.net	✗
googletagmanager.com	✗
chimpstatic.com	✗
cosmote.gr	✗
mailchimp.com	✗
list-manage.com	✗

Tracker	Blocks #	Blocked here	Global block	Remove from Trackers

Suspicious:	Mark as Tracker

Figure 48 – Cookies tracked without user consent for Germanos.gr and Apple.com

In Table 1, we can see summarized the result of this analysis:

Table 1: Summary of results

Websites	Number of Trackers without user's consent	Suspicious Domains
Samsung.com	4	14
Kotsovolos.gr	5	14
Gearbest.com	6	12
Plaisio.gr	0	10
Public.gr	0	19
Tokinito.gr	2	4
Kaizershop.gr	6	10
Apothema.gr	5	11
Germanos.gr	6	0
Apple.com	0	0

7. CONCLUSION

The goal of this thesis was to figure out how online users react to privacy issues in relation to personalize online advertising. While online advertising has been connected with many privacy related issues, it has emerged as one of the most effective and profitable way of advertisement, specifically the past 5 years. Presently, more and more websites and applications collect and utilize users' data, in order to provide flexible personalized services but many potential privacy concerns could arise. With the authorization of privacy legislation and regulations worldwide, the conflict between advertising and privacy is more critical, because personalized websites are obliged to comply with strict provisions, which often have remarkable impacts on how personalization may be performed. Nevertheless, the sector of online advertising, have manipulated the current status of GDPR using methods such as third-party cookies or even sometimes claiming user's consent with deceptive processes. As the privacy and network security have become a major topic of the internet today, global and European regulations are already being composed in order to make privacy on the web, even a stricter factor. The first step is the abolishment of third-party cookies by 2022 and it is worth of discussion on how web advertising can even find tricky ways to overcome this situation in order to collect users' data from third party sources.

A detailed analysis of Top 20 Consumer electronic websites has been conducted in order to test the level of privacy compliance. From this analysis, it has been concluded that most of the websites have already taken decisive actions to address online privacy concerns to stay ahead of the competition, minimize risk and build online trust. These actions either include the consent of the user for any form of cookie or the use of first party cookies only. There were some limitations related to the available material for the fade out of third-party cookies and the focus on first party cookies as an alternative. However, with the use of trustworthy researches, we have listed all possible ways that we will see first party cookies being leveraged in the upcoming future. Additionally, for the scope of the privacy analysis of the websites, free apps and tools have been used with several limitations. For future research it would be interesting to use more advanced tools to test whether the same websites use cookies priorly to user consent, as well as to expand this research into the mobile ecosystem. A main outcome of the overall analysis is that all relevant stakeholders (organizations, wen site operators/developers but also legislators) should put much effort on forcing transparency to the relevant data processing, whilst at the same time effort should be put on further raising the users awareness. Since the ePrivacy Regulation has not finalized yet, the conclusions of research in this area provide useful feedback to the legislator, as well as to all relevant stakeholders.

REFERENCES

[1] Barbara K. Kaye and Norman J. Medoff. 2001. World Wide Web: A Mass Communication Perspective (2nd. ed.). McGraw-Hill Higher Education.

[2] Evans, David S. 2009. "The Online Advertising Industry: Economics, Evolution, and Privacy." Journal of Economic Perspectives, 23 (3): 37-60.

[3] Statista (2020). Digital Advertising - Worldwide | Statista Market Forecast. [online] Available at: <https://www.statista.com/outlook/216/100/digital-advertising/worldwide> [Accessed 19 October 2020].

[4] Yuan, Shuai & Wang, Jun & Zhao, Xiaoxue. (2013). Real-time Bidding for Online Advertising: Measurement and Analysis. 10.1145/2501040.2501980

[5] Goldfarb, Avi, and Catherine E. Tucker. "Privacy regulation and online advertising." Management Science,vol. 57, no. 1, 2011, p. 57+. Gale Academic OneFile,

[6] David Silverman, Partner, PwC(2016).IAB internet advertising revenue report: 2016 full year results. PricewaterhouseCoopers, Internet Advertising Bureau.

[7] Marketing Land. 2020. Display Advertising News & Trends | Marketing Land. [online] Available at: <https://marketingland.com/library/channel/display-advertising> [Accessed 1 December 2020].

[8] Econsultancy. 2020. *Paid Search Best Practice Guide*. [online] Available at: <https://econsultancy.com/reports/paid-search-marketing-ppc-best-practice-guide/> [Accessed 10 October 2020].

[9]J. R. Mayer and J. C. Mitchell, "Third-Party Web Tracking: Policy and Technology," *2012 IEEE Symposium on Security and Privacy*, San Francisco, CA, 2012, pp. 413-427, doi: 10.1109/SP.2012.47.

[10]Lotame Team "1st Party Data, 2nd Party Data, and 3rd Party Data." (2020, October 09). Available at:<https://www.lotame.com/1st-party-2nd-party-3rd-party-data-what-does-it-all-mean/>[Accessed 16 November 2020]

[11] Jacqueline Zote, (7 January 2020). *55 Critical Social Media Statistics To Fuel Your 2020 Strategy*. Sprout Social. Available at: <https://sproutsocial.com/insights/social-media-statistics/> [Accessed 11 October 2020].

[12] Statista(2020). Social Media Advertising - Worldwide | Statista Market Forecast.Available at: <https://www.statista.com/outlook/220/100/social-media-advertising/worldwide> [Accessed 12 November 2020].

[13] Global Web Index. (2019) Ad-Blocking: Consumer Trends, Behaviors And Motivations Available at: <https://www.globalwebindex.com/reports/global-ad-blocking-behavior> [Accessed 01 November 2020].

[14] Marjan Falahrastegar, Hamed Haddadi, Steve Uhlig, Richard Mortier Queen Mary University of London, Qatar Computing Research Institute, University of Nottingham "Anatomy of the Third-Party Web Tracking Ecosystem" ,arXiv:1409.1066v1 [cs.SI] , 3 September 2014.

[15] On Audience Ltd. (2020). What Is Raw Data? Definition, Examples | Blog Onaudience.Com. [online] Available at: <https://www.onaudience.com/resources/raw-data-definition-examples/> [Accessed 21 November 2020].

[16] Google Support (2020). Personalized Advertising - Advertising Policies Help. [online] Available at: <https://support.google.com/adspolicy/answer/143465?hl=en> [Accessed 9 November 2020].

[17] Google.com (2020). EU User Consent Policy – Company – Google. Available at: <https://www.google.com/about/company/user-consent-policy.html> [Accessed 10 October 2020].

[18] Ashish Kumar & Jari Salo (2016): Effects of link placements in email newsletters on their click-through rate, Journal of Marketing Communications, DOI: 10.1080/13527266.2016.1147485

[19] SendPulse Team. (2020, March 30). What is a Personalized Email: Definition and Guide. Available at: <https://sendpulse.com/support/glossary/personalized-email> [Accessed 15 November 2020]

[20] RSA Data Privacy & Security Survey 2019: The Growing Data Disconnect between Consumers and Businesses

[21] David M. Kristol.HTTP Cookies: Standards, Privacy, and Politics. ACM Trans. Internet Technol. Bell Labs, Lucent Technologies (9 May 2001).

[22] Krishnan Kamani, 2018. How Does... DSP, SSP And DMP Work?.Mumbrella. Available at: <https://mumbrella.com.au/dsp-ssp-dmp-work-505095> [Accessed 4 October 2020].

[23] Panagiotis Papadopoulos, Nicolas Kourtellis, and Evangelos P. Markatos. 2019. Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask. In Proceedings of the 2019 World Wide Web Conference (WWW ’19), May 13–17, 2019, San Francisco, CA, USA. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3308558.3313542

[24] Google Developers. 2020. Cookie Matching | Real-Time Bidding . Available at: <https://developers.google.com/authorized-buyers/rtb/cookie-guide> [Accessed 15 November 2020].

[25] East, Robert & Hammond, Kathy & Lomax, Wendy & Robinson, Helen. (2005). What is the Effect of a Recommendation?. The Marketing Review. 5. 10.1362/1469347054426186.

[26] David Berkowitz -Web Archive -Typepad. (11 August 2019). Marketers Studio - David Berkowitz's Marketing Blog: The Social Graph Ad Targeting Buyer's Guide. Available at: <https://web.archive.org/web/20140502003033/http://www.marketersstudio.com/2009/08/the-social-graph-ad-targeting-buyers-guide.html#axzz151ZHyNK0> [Accessed 18 October 2020].

[27] Bannersnack. 2020. What Is Display Advertising? (A Complete Beginner's Guide). [online] Available at: <https://blog.bannersnack.com/display-advertising-guide/> [Accessed 6 November 2020].

[28] THE MEDIA BUSINESS;October 24, 1994 Time Inc. Raises Its Multimedia Profile With an Internet Test, The New York Times.

[29] GDPR.eu.Official Website of the European Union. 2018. What Is GDPR, The EU’S New Data Protection Law? - Available at: <https://gdpr.eu/what-is-gdpr/> [Accessed 24 November 2020].

[30] Omniconvert. 2020 What is a Pop Up Banner?

Available at: <https://www.omniconvert.com/what-is/popup-banner/> [Accessed 4 November 202**0].**

[31] Goldberg, Samuel and Johnson, Garrett and Shriver, Scott July 17, 2019. Regulating Privacy Online: An Economic Evaluation of the GDPR. Available at SSRN< https://ssrn.com/abstract=3421731 or http://dx.doi.org/10.2139/ssrn.3421731>

[32] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann , 2018. The Unwanted Sharing Economy: An Analysis Of Cookie Syncing And User Transparency Under GDPR. Available at: <https://arxiv.org/pdf/1811.08660.pdf> [Accessed 1 December 2020].

[33] Andy Pitre, July 17 2019. Ad Tracking: What It Is & How To Do It. [online] Blog.hubspot.com. Available at: <https://blog.hubspot.com/blog/tabid/6307/bid/7249/a-marketer-s-guide-to-tracking-online-campaigns.aspx> [Accessed 2 December 2020].

[34] Jannick Sørensen and Sokol Kosta. 2019. Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites.In Proceedings of the 2019 World Wide Web Conference(WWW'19),May 13–17, 2019, San Francisco,Available at:<https://doi.org/10.1145/3308558.3313524>

[35] Guy Aridor, Yeon-Koo Che, Tobias Salz. 2020. The Economic Consequences Of Data Privacy Regulation: Empirical Evidence From GDPR.National Bureau of Economic Research

[36] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation

[37] Know Online Advertising. 2020. *Tracking - Know Online Advertising*. [online] Available at: <https://www.knowonlineadvertising.com/tracking/> [Accessed 2 December 2020].

[38] Ravi Pathak & Siegfried Stepke (2014). "Webinar: Understanding the Impact of TV Advertisement on Website Traffic"

[39] Goldfarb,Avi & Tucker,Catherine.(2011).Online Display Advertising: Targeting and Obtrusiveness. Marketing Science. 30. 389-404. 10.2307/23012474

[40] Goldberg, Samuel and Johnson, Garrett and Shriver, Scott, Regulating Privacy Online: An Economic Evaluation of the GDPR (July 17, 2019). Available at:< https://ssrn.com/abstract=3421731 or http://dx.doi.org/10.2139/ssrn.3421731 >

[41] Warren And Brandeis,December 15 1980. "The Right To Privacy". Available at: <https://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html> [Accessed 7 December 2020].

[42] Belle Woodward,Nancy Martin,Thomas Imboden. August 9 2011. Expansion And Validation Of The PAPA Framework.School of Information Systems and Applied Technologies Southern Illinois University Carbondale, IL 62901, USA Available at: <https://files.eric.ed.gov/fulltext/EJ1145453.pdf> [Accessed 10 December 2020].

[43] School of Information Systems. 2018. PAPA: PRIVACY, ACCURACY, PROPERTY, AND ACCESSIBILITY. Available at: <https://sis.binus.ac.id/2018/02/22/papa-privacy-accuracy-property-and-accessibility/#:~:text=Mason%20identified%20four%20such%20areas,Accuracy%2C%20Property%2C%20and%20Accessibility.&text=What%20information%20must%20people%20reveal%20about%20themselves%20to%20others%3F&text=What%20information%20does%20a%20person,protection%2C%20and%20under%20what%20conditions%3F> [Accessed 4 December 2020].

[44] Wang, J. et al. "Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting." ArXiv abs/1610.03013 (2017): n. Pag.

[45] Yuan, Shuai & Wang, Jun & Zhao, Xiaoxue. (2013). Real-time Bidding for Online Advertising: Measurement and Analysis. 10.1145/2501040.2501980.

[46]Jose Estrada-Jiménez,Javier Parra-Arnau, Ana Rodriguez-Hoyos and Jordi Forne(2017)"Online advertising: Analysis of privacy threats and protection approaches."

[47] Yuan, S., Wang, J., Chen, B., Mason, P., and Seljan, S.(2014). An empirical study of reserve price optimisation in real-time bidding. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1897–1906. ACM.

[48] Anagnostopoulos, A., Broder, A. Z., Gabrilovich, E., Josifovski, V., and Riedel, L. (2007). Just-in-time contextual advertising. In Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, pages 331–340. ACM.

[49]Yegui Cai, George O. M. Yee, Yuan Xiang Gu, and Chung-Horng Lung. (May 2020). Threats to Online Advertising and Countermeasures: A Technical Survey.Available at:<https://doi.org/10.1145/3374136>

[50] Brett Stone-Gross, Ryan Stevens, Apostolis Zarras, Richard Kemmerer, Chris Kruegel, and Giovanni Vigna. 2011. Understanding fraudulent activities in online ad exchanges. In Proceedings of the ACM SIGCOMM Conference on Internet Measurement. ACM, 279– 294

[51]George O. M. Yee. (2017). Visualization and prioritization of privacy risks in software systems.

[52]Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, and Oana Goga. (2018). Privacy risks with Facebook's PII-based targeting: Auditing a data broker's advertising interface. In Proceedings of the IEEE Symposium on Security and Privacy (SP'18). IEEE, 89–107

[53]Adrienne Porter Felt, Matthew Finifter, Erika Chin, Steve Hanna, and David Wagner. (2011). A survey of mobile malware in the wild. In Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices.

[54] Jerome Segura. (2017). Canada and the U.K. hit by Ramnit Trojan in new malvertising campaign. Available at:< https://blog.malwarebytes.com/threat-analysis/exploits-threat-analysis/20 17/03/canada-u-k-hit-ramnit-trojan-new-malvertising-campaign/.>

[55]Lucian Constantin. (2016). The AdGholas malvertising campaign infected thousands of computers per day.Available at : <http://www.computerworld.com/article/3101823/security/the-adgholas-malvertising-campaign-infected-thousands-of-computersper-day.html.>

[56]Molva R., Michiardi P. (2003) Security in Ad Hoc Networks. In: Conti M., Giordano S., Gregori E., Olariu S. (eds) Personal Wireless Communications. PWC 2003.. Available at <https://doi.org/10.1007/978-3-540-39867-7_69>

[57]M. Backes, A. Kate, M. Maffei, and K. Pecina(2012) "Obliviad: Provably secure and practical online behavioral advertising," IEEE Symposium on Security and Privacy.

[58] Hamed Haddadi. (2010). Fighting online click-fraud using bluff ads. ACM SIGCOMM Comput. Commun. Rev. 40

[59]William E. Sobel and Sourabh Satish.(2012). Methods and systems for detecting man-in-the-browser attacks. US Patent 8,225,401.

[60]Ari Juels, Sid Stamm, and Markus Jakobsson. (2007). Combating click fraud via premium clicks. In Proceedings of the USENIX Security Symposium.

[61]J. Chen, D. Lin, A. Kaufman, and Y. Villa. 2014. Click stream analysis for fraud detection. Available at: < https://www.google.com/ patents/US8880441>. US Patent 8,880,441

[62]M. Ter Louw, K. T. Ganesh, and V. Venkatakrishnan(2010) “Adjail: Practical enforcement of confidentiality and integrity policies on web advertisements.” in USENIX Security Symposium, 2010, pp. 371–388

[63]Zyskind G.,Oz Nathan,Pentland A. ’Sandy’(2015) Decentralizing Privacy: Using Blockchain to Protect Personal Data.IEEE CS Security and Privacy Workshops

[64] Goldfarb, Avi, and Catherine E. Tucker. “Privacy Regulation and Online Advertising.” Management Science 57.1 (2011) : 57-71.

[65] i-SCOOP. (2018). The New EU Eprivacy Regulation: What You Need To Know. [online] Available at: <https://www.i-scoop.eu/gdpr/eu-eprivacy-regulation/>

[66] Shaping Europe’s digital future - European Commission. (2017). Proposal For An Eprivacy Regulation - Shaping Europe’S Digital Future - European Commission. [online] Available at: <https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation> [Accessed 2020]