Online Tracking — Analysis of the TwitchTV streaming platform

Tim Bierth, Oleksii Kulikov, Jurek Olden

July 2018

1 Introduction

Humanity has always been obsessed with observing, measuring and analyzing its environment. We have built sophisticated instruments, equipped with sensors, to gather data. There are metrics one can measure directly, such as distance or voltage, while other units must be inferred from a number of assessments. We tend to measure and deduct from very feasible heuristics. Usually, the more one observes, the greater and more accurate the conclusion becomes. In our time and age, utilizing the strong computational capabilities at our disposal, we even dare to predict human behavior. The starting point for such predictions is data, which nowadays can be easily collected online. The Internet is foremost a web of servers and clients, connected through coaxial cables, wireless transmitters and receivers. Apart from some autonomous systems, the web is a passive medium. This means that individuals decide all by themselves which website they want to visit. This decision making is exactly what data collection on the Internet is about — human behavior. Because we are dealing with a passive medium, behind every click sits a living person, which has the monetary means to acquire various products. Data scientists now ask themselves, how is this person or group going to spend their money? The answer can be given by thoroughly analyzing the group.

An exemplary procedure would be to start by classifying the Internet users according to some pattern. This entails agreeing on a unique identifier for a group of users, such as the their country of origin. The next step would be to observe this group's behavior. For example, one could be interested in the number of times a person accesses a certain website per day. One can regress this behavior in data structures attributed to artificial intelligence. Now, given a user of unknown origin and their frequency of accessing a certain website, we can draw conclusions about the user's country of origin.

The utility of this example may be questionable due to simplicity of the experiment. However, similar principles can be applied to human behavior on the web. The more one knows about a person or a group, the greater the ability to influence that entity.

This paper is divided into four major parts: relevant literature, technical background, introduction of TwitchTV (in the following simply 'Twitch'), and the critical analysis of the platform. Within the section on relevant literature we cover some of the resources, mostly available on the web, that we consider suited for further reading on the topic of data collection and data analysis. The technical background section delivers a basic understanding of how user data can be collected and explains the most common tracking practices. In addition to that, a brief overview of computational intelligence is given to gain insight into how conclusions can be drawn from data. The

introduction of the Twitch services covers the main aspects of the streaming platform, focusing on its business model and its implications to modern society. The analysis section concentrates on the data flow around the Twitch platform and the way Twitch interacts with its customers and partners. At this point, we conduct a thorough analysis of the Twitch's data and cookie policy and introduce the findings of our experimental proceedings. Finally, we sum up our results in a conclusion while giving a brief overview of possible future work.

2 Relevant literature

While there are not many resources focusing specifically on Twitch and the way it treats user data, the entire field of critical data studies is currently on the rising edge of the modern society. Numerous books, articles and other information sources have emerged studying the social impact of exponentially growing data accumulation. The process of collecting data from increasingly many areas of human life and of transforming this information into economic value is called *datafication*.

Some of the sources relevant for this research take an optimistic approach to the matter, like for instance M. Salganik (2016) [1]. In his book on the evolution of social research he states that, despite being a social scientist himself, he is willing "to adopt the optimistic tone of a data scientist" [1]. In his work, he tries to abstain from using field-specific terminology in order to target readers who work on the narrow verge between social science and data analytics. Thus, the book is advisable as a read for interdisciplinary researchers.

Another author taking more of a negative approach while writing specifically on the topic of corporate surveillance is Wolfie Christl (2017) [2]. "With the actors guided only by economic goals, a data environment has emerged in which individuals are constantly surveyed and evaluated, categorized and grouped, rated and ranked, numbered and quantified, included or excluded, and, as a result, treated differently" [2] - this is how the digital rights activist based in Vienna, sees the current development in the area of data collection. One of the leading questions he tries to answer in his article is "How do online platforms, tech companies, and data brokers collect, trade, and make use of personal data?" [2]. The train of thought that he follows is extensively visualized by various graphical representations of data flow between large data-driven companies, such as Google and Facebook, ever-present data brokers, and other entities taking part in the processes of data gathering and data analytics.

For the readers looking for a more scientific approach to the implications of datafication "Critical questions for big data" [3] by K. Crawford et al. (2012) can be a good starting point. The authors of this scientific article offer "six provocations to spark conversations about the issues of Big Data" [3]. These 'provocations' concentrate on the areas of technology, analysis and mythology, and are expected to create a lively discussion with the reader, while focusing on both utopian and dystopian aspects of datafication.

Furthermore, one of the few Twitch-centered resources available on the web is the official Data Science Blog [4] based on the company's own platform. It provides numerous entries and articles on the various aspects of what and how Twitch does with the massive data sets it collects. Some of the entries are of profoundly technical nature and focus on the practical aspects of what can be done with the acquired data. Others, however, include sociological elements and sometimes even spark a discussion between blog followers.

3 Technical Background

Websites are delivered to *user agents*, such as web browsers, via the *hypertext transfer protocol* (HTTP). HTTP is an inherently stateless protocol [5]. This means that when some client requests a website, the site is unknowing of any previous visits or actions of this client. This poses a challenge for web services which do not only display information and media, but also rely on state. The typical example of a web feature requiring state is the shopping cart of an online store: By visiting the store, a client initiates a *session*. During this session, the client may put an item into the shopping cart, but since HTTP is agnostic to state, the website would not be able to know about the item the client has put into the shopping cart once the user loads another page, e.g. by clicking a link or performing a search.

This problem is solved by using *cookies*. A cookie is a piece of state data, which is stored **on the client**. A website may request the storage of a cookie by providing a name and content together with the actual web content it is serving to the client (Figure 1, Page 4) [6]. On every subsequent visit to the same domain, the client then sends the cookies which were earlier set by the site along with the HTTP request [6, Section 3]. It is mandatory that this piece of state data ist stored on the client, since the client is able to recognize a homepage when visiting it subsequent times, but the same does not hold true the other way around: HTTP uses the *transmission control protocol* (TCP) for the underlying connections [5, Page 13]. Using TCP, a client is known to a web server only by the pair of its IP-Address and source port, which are obtained during the TCP Handshake [7, Section 2.7]. This initially unique

```
GET /index.html HTTP/1.1
Host: www.shop.example

[..]
```

(a) HTTP request

```
HTTP/1.0 200 OK
Content-type: text/html
Set-Cookie: sessionId=a39d8f9;
  Expires=Sat, 09 Jun 2018 10:18:14 GMT

[..]
```

(b) Server response with cookie header

Figure 1: Example of a website requesting cookie storage along with delivering content

identifier may change dynamically before any subsequent request to the web server, which makes it impossible for web servers to reliably recognize users. It is therefore mandatory for cookies to work as intended that the client is governing them; this also means that the client is able to deny a cookie by ignoring the web servers request to set the cookie [6, Section 5.2]. The client may also delete cookies at a later point or even modify them. In practice, this is achieved using the privacy options of a users browser.

So, using cookies, the online store is now able to generate a *session id* for some visiting user, which the user's web browser then saves as a cookie. The user's web browser sends this session id (in form of a cookie) to the website every time a request to this domain is sent. The website is then able to correlate the session id with the internally saved shopping cart of this user and will deliver the appropriate site content.

As we can see not only by looking at the online store use case, but also by considering any site where e.g. users need to log in, cookies are tremendously useful and essential for how we use the Internet today.

3.1 Third Party Cookies

So far, we have considered cookies as a storage for the state of a client session at a single domain, governed by the client itself and consisting mostly of identifiers. More interesting scenarios arise when visiting more complex websites, where the content is not delivered from a single web domain, but some elements on the page refer to external links (Figure 2, Page 5). When a user agent, such as a browser, tries to render the page, it has to follow these external links to retrieve the referenced content: The user agent has to send an additional request and the third-party server to which the request is sent may add set-cookie headers to the response. Though again, since cookies are governed by the user agent itself, they may be rejected or dealt

```
[..]
<div>
  <object type="text/html" data="http://third-party.abc.com/">
  </object>
</div>
[..]
```

Figure 2: Page element which has to be resolved by sending a request to abc.com

with any other way [6, Section 5.2, Section 7.1]. In real-world applications, third-party cookies are used for legitimate features: An example are third-party login providers (such as 'log in via Google/Facebook' etc. (Figure 3, Page 5)) or any kind of external poll or form etc. Since these use cases require state, the third-party content providers have to be able to set their own cookies. When used for tracking, third-party cookies are typically set

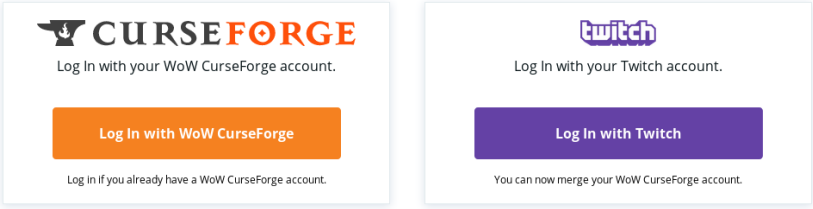

Figure 3: Feature to login via a twitch account instead of using an account specific to the site [8]

when the user agent requests third-party content like advertisement banners or *tracking beacons*. Beacons are usually invisible 1×1 pixel pictures hidden in the page content, which may also be an HTML email (Figure 4, Page 5). In the email case, they usually simply check if they are being retrieved to confirm that some specific recipient read the email, on websites they often act as an external page hit counter. Otherwise they can also work just like other

```
[..]
<img src="img.abc.com/set-a-cookie.png" height="1" width="1">
[..]
```

Figure 4: Example of a one by one sized pictured to be requested from abc.com

content designed for tracking: After the tracking cookie with an id has been set for the first time, this cookie will be sent back to the tracking server every time the client resolves third-party content from the tracking servers domain while rendering any other site. Thus the tracking server immediately knows that those two sites have been visited by this specific user [6, Section 7.1]. For advertisement purposes, this may for example be leveraged to identify user interests and tailor the next ad banners the user will see: should the user visit a third web page containing an ad banner of the same provider which has already identified the user as described above, it may show some advertisement referring to his or her interests.

3.2 Real world example

By combining two important conclusions, namely that a browser without a restrictive cookie policy will resolve third party content embedded in a website by sending an HTTP GET request to the link provided in the HTML source of the original website, and the fact that a browser will attach all cookies set by some domain to any HTTP request to that same domain, we can account for how a site like Facebook may be able to track our interests. The online site of the german newspaper "Frankfurter Allgemeine Zeitung" embeds social media buttons on each article, including a 'Share on Facebook' button (Figure 5, Page 6). By inspecting the actual HTML source code,

Figure 5: Social media elements on the article page

which was sent to us by the webserver (Figure 6, Page 7), we can note two things:

- The button is actually an external link (third-party content). When

```
<div class="atc-ContainerSocialMedia_Buttons">
<ul>
[...]
"facebook,twitter,xing"
[...]
data-customsharelink="http://www.faz.net/aktuell/politik/ausland/ \
        schlagabtausch-mit-israel- \
        in-syrien-und-irans-ambitionen-15583553.html"
[...]
<a href="https://www.facebook.com/sharer/sharer.php?u=http%3A%2F%2F \
        www.faz.net%2Faktuell%2Fpolitik%2Fausland%2F \
        schlagabtausch-mit-israel- \
        in-syrien-und-irans-ambitionen-15583553.html \
        %3FGEPC%3Ds2" title="Auf Facebook teilen"
[...]
</ul>
```

Figure 6: Original HTML code from Figure 5, Page 6

a browser renders this page, it needs to request the button from the `facebook.com` domain.

- The external link contains a variable which embeds the article link into the request.

If a user has logged in to Facebook prior to visiting this news site, the user's browser will have accepted some session-id cookie by Facebook, which will now be attached to the HTTP request it sends to `facebook.com` in order to resolve and render the 'Share on Facebook' button. So the Facebook server will receive a request for a specific resource along with a session-id which it can connect to the user; it also knows which website this user is currently loading, since this information is embedded in the link. Facebook has now successfully gathered that this specific user has been visiting this specific article.

3.3 Flash cookies

Before HTML5, the Adobe Flash Player has been widely adopted by websites to display interactive content such as videos and games [9]. For a client to use Flash elements in websites, one had to install the Adobe Flash Player, which was using cookies too, intended to store Flash specific settings. Flash

cookies work analogous to HTML cookies, but in this case the user agent is not the browser, but the Adobe Flash Player, which is problematic because this makes Flash cookies independent of the browser and also non-trivial to administer and delete [10]. In 2009, more than half of websites from a sample using flash cookies deployed them as a 'backup' for HTML cookies [11]; this means that HTML cookies are being re-set by websites after deletion by the user. The cookie content is read from the Flash cookie and then set again as a HTML cookie. The nature of flash cookies and their usage for tracking has led to controversies in the public discourse, the pinnacle being a 2010 lawsuit in the United States of America which has been settled with a 2.5M $ donation to resarch [12] [13]. With the adoption of HTML5 and other open standards, Flash is seeing diminished usage — in the Google Chrome browser, the number of users who had loaded at least one Adobe Flash element has dropped to 8% in 2018 from 80% in 2014 [14] — and the Adobe Flash Player end-of-life has been announced by Adobe for 2020 [15].

3.4 Data transfer

Based on these findings, it is clear that a company which is able to place its content on the maximum number of web domains, preferably websites with high traffic volume, will be able to gather the most user data. But the knowledge of a users interests and habits is most likely incomplete at best — it is not feasible for a company to be present on each and every website. While it is conceivable to assume some sort of sharing or trading of user data between large tracking networks, the claim is hard to prove and it may only be possible to heuristically find indicators of any such procedure.

3.5 Do-Not-Track Request

In 2009, an additional header field for HTTP requests has been devised, the Do-Not-Track (DNT) field [16]. It may be set to 0, which means the user consents to third-party tracking, or 1, which signals an active opt-out. NULL means that no preference was set [17]. The DNT header field is supported by all major browsers, but is widely disregarded by online services due to the lack of standardization and legal obligation.

3.6 Brief Overview Computational Intelligence

Given collected data, how can this data be processed and what knowledge can be gained? Computational intelligence is more of a set of universal algorithms, which are applicable to different data scenarios, than an actual

intelligence. The main goal of such a data based algorithm is to learn some correlation between metrics contained in the data set. There are two major tools of computational intelligence: clustering and neural networks. Both will be briefly presented to further the understanding on what conclusions can be drawn, given certain sets of data. The presented methods of data analysis are meant to visualize that there are simple algorithms behind key words, such as computational intelligence, and that there is by no means some sort of intelligence in the sense of human intelligence sifting through your data.

3.6.1 Clustering

The goal of clustering is to organize a data set based on criteria of similarity. The process of clustering can be easily understood on the example of the common clustering algorithm called *k-means*. When applying the k-means algorithm, each data point is presented by a vector $a_1, .., a_n$. Each element of the vector is a numerical value, which however can represent a degree of class. At the beginning, k random data points are generated. These are called cluster points. Then each data point in the data set is attributed to one of these random points, resulting in each cluster point having its own set of data points. It is called a *cluster set*. The cluster set is used to update the position of the cluster point. If c is a vector representing a cluster point, then the new i-th element of the vector is the mean of the i-th element of each data point in the cluster set:

$$c_i' = \frac{1}{|C|} \sum_{c \in C} c_i \quad where\ C\ is\ the\ set\ of\ clusters$$

Updating the cluster points is repeated until they do not change anymore or rather do not change in regard to some specified threshold. An example for the resulting clusters is given in Fig. 7. In the context of Twitch, each user could represent a data point. Each cluster, representing a user group with certain interests, could be dealt differently with to maximize profit.

3.6.2 Neural Networks

Neural networks are at their core a purely mathematical model, however inspired by nature. Computational intelligence tries to mimic learning processes that can be found in nature. Learning is nothing more than optimizing your future decisions based on past observations. It can be found in all kinds of mammals; and humans are the ones who excel at it. The structure used to learn is the brain, which consists of many tightly knitted neurons, a special

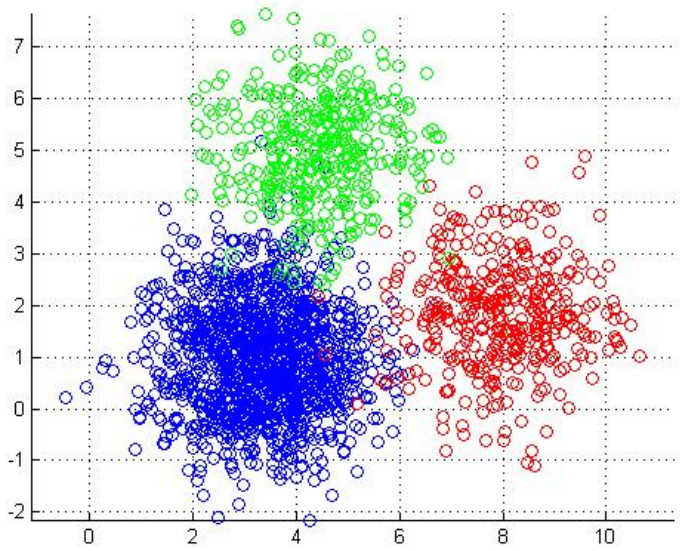

Figure 7: Example clustering performed by the k-means-algorithm.

type of cell. The input and the output of each single neuron can be measured in voltages. Frank Rosenblatt (1957) developed a mathematical model to emulate the behavior of a single neuron. This model is called the perceptron. A single modeled neuron consists of a set of weighted inputs, the propagation function, the activation function and an output. Input, propagation function and activation function are set from the beginning by the programmer or are predetermined by the data. The output is calculated as:

$$f_{act.}(f_{prop.}(inputs))$$

$$f_{prop.}(inputs) = \sum_i o_i \cdot w_{i,j}$$

The activation function is usually picked as a gaussian or sigmoid function, because a weighted sum of these functions is able to approximate next to every correlation. For a single perceptron, there is only one input and one output. The result of the propagation function comes down to a single weighted value, which is then fed into the activation function. Comparing the output of the neural network to the value one would like to regress to, the weight is being adjusted. However, this procedure reaches its limits, if there are so called hidden layers. A hidden layer is a set of neurons which lie between the input and the output layer. Their weights cannot be trained by simply comparing them to the desired values of the dataset. This problem

can be solved by the back propagation algorithm. Neural networks are not limited to feed forward topologies. There is much more to explore (e.g. [18]). In the context of Twitch, one could derive interests from a certain set of parameters. As we have just seen, neural nets have a certain accuracy, which is dependent on the structure and the training set fed to it. This means that profiling users with this technology can be more revealing, the greater the user base.

4 Twitch

Twitch is an online streaming platform specializing in video games. Twitch is solely the name of the streaming service and should not be confused with the company it is owned by — namely Twitch Interactive. Another fact we consider significant to keep in mind throughout the course of this work is that Twitch Interactive is a direct subsidiary of Amazon.com, Inc.

4.1 Platform

In a study conducted by the company Lifecourse Associates[1], Twitch is said to be responsible for the fourth-highest U.S. peak Internet traffic, right after Netflix, Google and Apple with their respective streaming services [19]. Besides, while there are no hard facts provided on the matter, it seems sensible to assume that Twitch is about as popular in European countries as it is in the U.S. In total, the platform accounted for 355bn minutes of watched streams from all over the world in 2017 [20].

One can differentiate between three different groups using the service:

- Broadcasters, who produce video content. Twitch provides broadcasters with the needed technology to stream games, professionally produced e-sports events or even outdoor activities.
- Users, who consume the content by watching streams, videos, setting preferences and participating in integrated online live chats.
- Developers, who embed the streaming platform into their own applications as well as produce extensions to make the video content more interesting and interactive.

Naturally, the platform is an ideal location for brand marketing and product placement due to the high visitor numbers.

[1] Lifecourse Associates accessible under `https://www.lifecourse.com`.

Besides actively watching streams and participating in chat, Twitch offers some convenience functionality to its users. Viewers can follow a specific channel they like in order to receive information and updates on the selected channel. These updates are usually issued in form of an e-mail to the followers, hence requiring their respective e-mail-address. In addition to that, users can subscribe to a channel, which results in them getting access to customized emotes that can be used in the accompanying live chat of the stream and allows them to participate in a 'subscriber only' chat whenever the mode is switched on. Subscriptions are acquired on a monthly basis. Twitch also offers many other features, including various user settings and preferences. However, we restrain from listing all of them in this paper, as we are mainly interested in the presented core functionality for our research purposes.

4.2 Business Model

Twitch generally has two sources of income: the revenue they share with the streamers and the money they make through contracts with third parties. The shared revenue consists of subscriptions and donations provided by the users. The previously explained subscription functionality is available after a \$4.99 fee[2] is paid. Additionally, in 2017 the platform added two additional subscription tiers for \$9.99 and \$24.99 respectively [21]. Users who simultaneously are *Amazon Prime* customers, receive the option to subscribe to one channel for free once a month. This service was created to establish a stronger connection between Twitch and Amazon users. Twitch shares the money made by subscriptions with the respective broadcaster on a 50-50 ratio. Donations are made by viewers who specifically like a certain broadcaster and therefore want to support them by paying an arbitrary amount of money to them. Donations are also shared on a 50-50 ratio between Twitch and the broadcaster. The money Twitch makes through contracts with third parties mostly consists of advertising fees, which the streaming service receives for providing a convenient advertising platform to its partners, such as *Amazon* or *MSI* for instance.

4.3 Implications to society

Another metric that many of the modern companies consider extremely valuable is the number of users their platform can reach. In this regard, Twitch has been growing very rapidly over the last several years, thus attracting ever growing numbers of users [19]. However, unlike most social media outlets,

[2]\$4.99 at the time of research, last access on July 2, 2018

Twitch does not provide any significant political grasp. It is in the first place motivated by making money, as most businesses are, and thus mainly interested in exposing its audience to the ads or services of its partners. Twitch's viewership is mostly comprised of young people — teens to young adults[3]. These users provide an audience, which has a common interest in computers, games and technology, to Twitch's advertisement partners. Because the viewer base is so young, they are still in the process of forming shopping habits and trusting certain companies. This makes Twitch even more attractive to businesses cooperating with it. Therefore, an argument could be made, that Twitch forms a connection between playing video games and buying advertised products, thus bringing together the two 'worlds' of spending money and leisure. Because the platform accompanies teens from a young age until they are grown adults, behavioral patterns are likely to be formed in their subconsciousness. This practice can be compared to *FIFA* providing companies like *Adidas* or *Nike* with a platform to advertise their products. While social and behavioral implications of this sort should be kept in mind when discussing how Twitch treats its users' data, they are not directly subject of this paper. As the data is supposedly collected anonymously, there is no direct harm in adjusting advertisements to the field of interests of the customer. In how far this anonymity holds, we will discuss in our analysis of the platform (Section 5, Page 13).

5 Analysis

Being a relatively large platform, Twitch.tv provides an extensive overview of its data usage policies. In the following, we are going to analyze both the flow of data between Twitch and its partners as well as the way user data is managed within the platform.

5.1 Cookie Policy

When first entering the main web page of Twitch, one is welcomed by one of the currently trending streams and a list of suggested games to watch. We instantly notice that the website's content is in German, meaning that Twitch has access to our IP-address and hence the location of our computer.

On the lower part of the screen there is a discrete info box stating that the platform uses cookies and that further usage of the website implies that the user agrees with their cookie policy (Fig.8). There is a direct link to the

[3]According to [19], "Nearly half of Twitch's visitors are 18- to 34-year-olds."

aforementioned policy, in case the user wants to receive more information on the matter before proceeding.

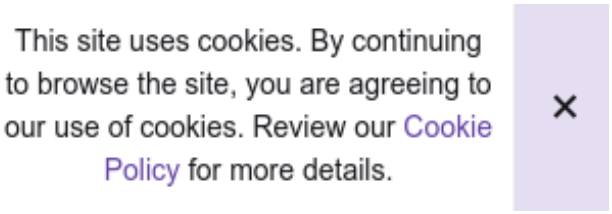

Figure 8: Cookie Usage Notification (English Version)

In order to understand what information Twitch gathers on its users and how it is processed, we follow the given link to their Cookie Policy and analyze it. The entire Cookie Policy can be found in [22].

The main reason why Twitch uses trackers is to improve the users' online experience, as it states. Hereby, not only cookies, but also pixels and web beacons are listed as possible trackers (we explained these in (Section 3.1, Page 4)). Later in the policy though, all of these trackers are simply referred to as *cookies*. For this reason, we assume that cookies are the type of tracker that enjoys the most prevalent usage by Twitch.

The cookies are furthermore classified into *Essential Cookies*, *Performance Cookies*, *Functionality Cookies*, *Targeting or Advertising Cookies* as well as *Flash Cookies*. Each type of cookie is responsible for a specific task.

5.1.1 Essential Cookies

These cookies are, as stated, essential for the proper functioning of the website and the features it provides. One of these features is access to paid content. At this point, Twitch needs to know which privileges the user has already paid for and which not. This information is seemingly stored in form of a cookie on the user's PC. Another feature is restricted access to secured areas of the website. Without further information given, we assume that secured areas are the ones protected by authorization measures, like e.g. login necessity. The cookie used for this purpose is the one called *twitch_session_id* (cf. (Section 3, Page 3)). This cookie is said to enable efficient navigation through the website.

5.1.2 Performance Cookies

Technically, these are analytics cookies, such as *Google Analytics* and *Mixpanel*. The purpose of these cookies is to improve the way the site works.

This is done by collecting information on which pages of the website are frequently used, whether any difficulties are being encountered, and whether the ads that are shown are clicked on by the users.

5.1.3 Functionality Cookies

Functionality cookies can be used to customize website appearance to fit the preferences/settings of the current user. The Cookie Policy does not provide any examples for this type. From our experience though, the preferences mentioned could be the likes of chat color and preferred stream resolution.

5.1.4 Targeting or Advertising Cookies

Targeting or Advertising Cookies provide ads that are more relevant to the user's interests. These cookies contain information on what the user has watched and which services provided by Twitch they have used. The platform allows itself to share this information with its partnered advertisement providers. A subsidiary of Google named *DoubleClick* is mentioned. We assume that there might be a correlation between *Google Analytics* and *DoubleClick*, as both services are technically powered by Google. Hence, internal information exchange between the two services seems likely.

5.1.5 Flash Cookies

Flash Cookies are slightly different from the rest of cookies used, as we mentioned in the Technical Background (Section 3.3, Page 7). They cannot be deleted out of the user's browser, but instead have to be managed on the website of *Adobe Flash Player*. Twitch uses this sort of cookies to deliver visual content to users, such as video clips and animations. Without Flash Cookies some of the streaming content provided by the platform is not available.

5.1.6 Third-Party Cookies

The "darkest" part of Twitch's cookie world are Third-Party Cookies. These can be used by either extension developers for Twitch broadcasters, third-party advertisers, or "other organizations" [22, Section 3].

- *Extension Developers* use cookies to allow user settings for these extensions. This set of cookies is limited to the extensions the user decided to opt in for.
- *Third-Party Advertisers* use cookies to collect information about the user's behavior on the websites of advertisements they might have

clicked on. This way, the ads can be tailored with the user's interests in mind. It is up to the advertisers to decide which way they do that.

- We are not sure what Twitch means by the vague entity of *"Other Organizations"*, because no further information on what these organizations are or what their cookies do is given. We only can guess that these organizations are not allowed to gather or use information beyond the limits of both the U.S. and local country's jurisdiction. For further inquiries, however, the email address `privacy@twitch.tv` is given.

5.1.7 Conclusion on Cookie Policy

We conclude that Essential Cookies are the only ones that are necessarily needed for the website to function correctly. The rest of the cookies appear to be of only insignificant value to the user in terms of user experience. While the way Twitch treats data gathered by these cookies will be analyzed in the next section, we believe that the user may deactivate all but the Essential ones, without fearing significant service shortages.

5.2 Privacy Policy

Having analyzed the Cookie Policy, we switch our attention to the information given in the Privacy Policy of Twitch to determine the ways it treats its users' data.

5.2.1 What data Twitch gathers

In their Privacy Policy, Twitch classifies the data being gathered into three different types:

1. Information you can **provide to Twitch yourself** when using their services, such as:
    - user name
    - email address
    - postal mailing address
    - telephone number
    - credit card number
    - billing information

2. Information Twitch **automatically collects** when you visit their website, such as:

    - IP address
    - device type
    - browser type
    - software and system type

3. Information from **"other sources"**: In particular, if you have previously connected Twitch to one of their partners, which are mostly online social networks, Twitch can collect your data from these partners. However, it is not stated what kind of information this is.

As we have previously seen, some of these data types are saved in cookies. Some other data might be stored directly on Twitch servers though, like e.g. your billing information.

5.2.2 What Twitch does with the data

We already illuminated some of the things Twitch does with the data in the Cookie Policy section (Section 5.1, Page 13). The Privacy Policy extends the list of these data procedures in a significant way.

For one, there is the feature to automatically install updates to the Twitch application on the user's computer. This procedure logically requires type 2 data (automatically collected data), such as device and operating system information. Without, it would not be possible to determine the correct update package.

Another way Twitch uses our data is to communicate with the user, e.g. via email. Such reasons might be policy updates or promotional information. For the purpose of communication Twitch requires type 1 data (provided by users themselves), such as user name, email address, or postal address.

Additionally, Twitch has to provide user information:

- in compliance with U.S. laws or user residence country laws
- in response to court order, judicial request or any other government request from the respective country [4]

[4]The exact text of the policy might be better for accuracy purposes at this point: "Twitch may disclose user information if we believe in good faith that such disclosure is necessary to comply with U.S. state and federal laws or other applicable laws around the world (for example, in the country of your residence), or respond to a court order, judicial or other government request, subpoena, or warrant in the manner legally required. " [23, Section 3].

Without proper legislative knowledge of the particular country, it is difficult to estimate the extent to which the platform is obligated to share user data with the governmental structures. In the light of recent government espionage scandals, it seems advisable to keep this in mind.

Last but not least, Twitch retains the right to use data to protect itself from potential liability, third-party allegations and technically everything else that might damage Twitch as a company in any way. This point is vaguely formulated and leaves Twitch the necessary "freedom of movement" for the case of a force majeure situation. This part of the Privacy Policy makes us wary of the consequences for the user. It seems that in case of trouble, Twitch would not hesitate to put its own interests above the ones of the user.

5.2.3 Requirements to third-parties reusing the data

There are some requirements that third-parties have to comply with if they want to reuse data that Twitch shared with them. The first one is to only use data for the purpose for which it was originally shared. Hence, if Twitch, for instance, shared a user's location with their partner Amazon for analytics purposes, the latter would not be allowed to reuse the data for marketing purposes.

Another requirement to third-parties are potent confidentiality measures regarding the shared data.

Twitch does not guarantee though that the partners' usage of data will comply with the Privacy Policy issued by Twitch. So technically, except for the two aforementioned points and the legal boundaries, the user does not know exactly what third parties do with their data. The only way to find this out is by working yourself through the Privacy Policies of these third-parties. However, Twitch states that it does not share information with third-parties that would allow them to personally identify its users. This means that the data Twitch shares with its partners is anonymous. Nevertheless, third-party advertisers can indirectly conclude on one's identity. If a user clicked on a certain ad that was meant to target a specific group of people, the ad provider could deduce the user's personality from it.

5.2.4 Data security

Twitch does not by any means guarantee that the information users provide cannot be "accessed, disclosed, altered, or destroyed by breach of any of [their] physical, technical, or managerial safeguards" [23, Section 10]. This way, Twitch makes sure that it cannot be held responsible for any type of technical failure. On the one hand, this sounds alarming, as it might seem

that Twitch does not take data security seriously. On the other hand, Twitch is most likely aware of the effect that a massive data breach would have on its reputation. Hence, we assume that they do take the necessary measures to provide decent security of data.

5.2.5 Conclusion on Privacy Policy

The conducted analysis of the Privacy Policy shows us that it is unlikely for Twitch itself to misuse its users' data. However, we could also see that the platform can both voluntarily (partners and ad providers) and involuntarily (governmental requests) disclose and share the users' data with other parties. While these third-parties underlie their own Privacy Policies and legislative regulations of their respective countries, it still cannot be concluded without doubt that the data they receive from Twitch will not be misused.

5.3 Data Flow

Twitch's Cookie Policy and Privacy Policy gave us an insight into what types of data the company collects and how the data is being processed. To evaluate user information, Twitch essentially shares it with a number of entities (Fig. 9). While working through the Privacy Policy issued by Twitch, we were not able to identify a paragraph which would mention that the platform's partners are not allowed to reshare the data they received from Twitch. This means that the travel of data originating from the user is in no way limited to the entities shown. Nevertheless, we hope that our excerpt from the data flow around Twitch is detailed enough to demonstrate the main dependencies of the big picture.

On the one hand, there is the (simple) user of the services offered by Twitch. The user usually does not receive any personal data on other users or entities, and only serves as an origin to the information involved. In addition to that, the user sometimes fills the role of a 'trigger'. As soon as Twitch receives data from a newly registered user, it is able to request additional information on the person from both its *Social* and *Analytics Partners*. This process is significantly facilitated as soon as the user connected their Twitch account to one or more of the partners. This functionality is illustrated in Fig.10.

Keeping in mind that most of Twitch's relationships are bidirectional, the platform usually shares the newly acquired data on the registered user with its partners. The Privacy Policy does not directly state that Twitch is obligated to do so in exchange for information it receives from other entities. In the times of capitalism though, it seems likely that Twitch has to share

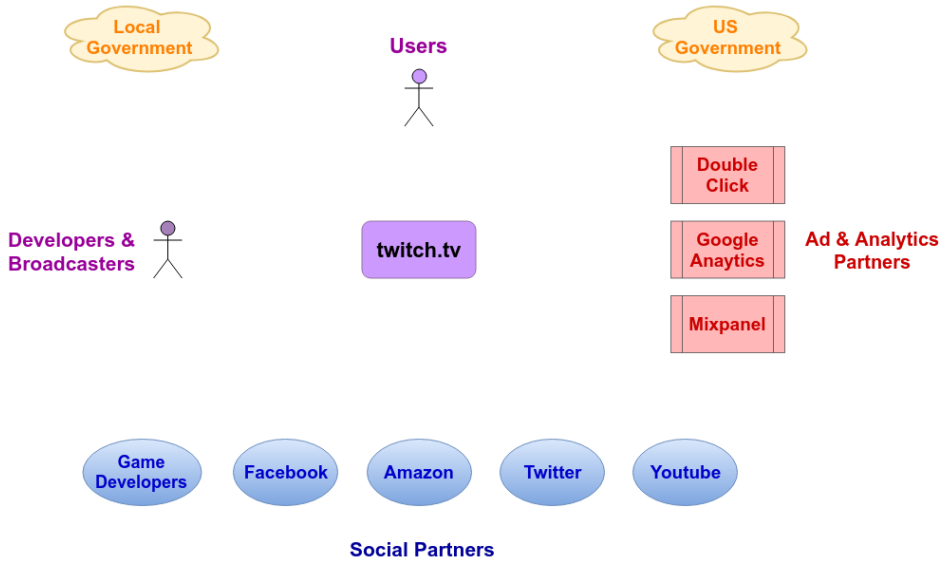

Figure 9: Entities involved in data flow with Twitch in the center

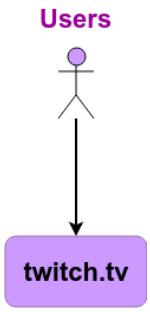

Figure 10: Data flow between Twitch and its users

its users' data the same way its partners do. This is especially the case for *Social Partners*, as they do not receive monetary compensation for their services in contrast to *Ad and Analytics Partners* like e.g. Google Analytics. A graphical visualization of Twitch's partner entities is shown in Fig.11. To facilitate user information exchange with its partners, Twitch offers a feature in the user's profile settings called *Recommended Connections*. On this page, the user can choose which social network, gaming console, or other partnered service they want to connect with their Twitch account. An example of the *Recommended Connections* page is given in Fig.16 (appendix).

Twitch partners and simple users are complemented by the "advanced" users. These are either *broadcasters* or *developers*. Broadcasters are the ones providing streaming content to Twitch users. Developers are the ones who

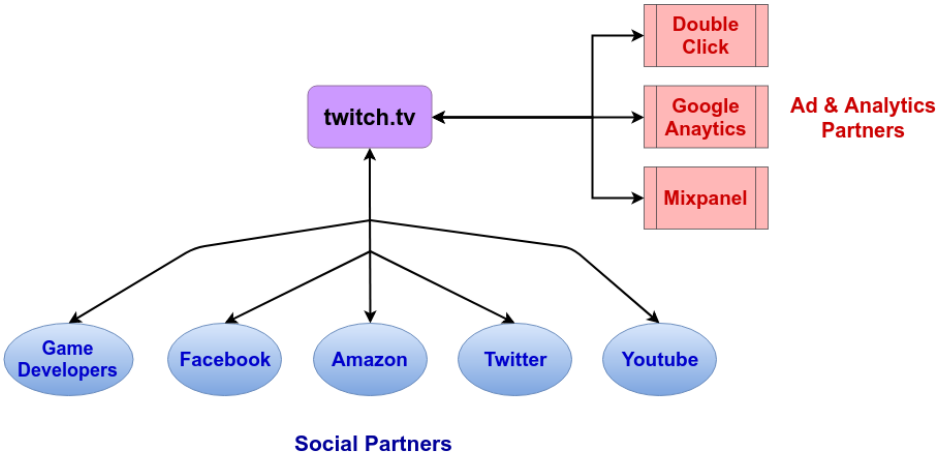

Figure 11: Data flow between Twitch and its social, advertisement and analytics partners

either need to embed Twitch services in an external environment or to deliver various features and extensions to the broadcasters [5].

The main difference between simple users and advanced users lies in the amount of data they have to share with Twitch and the amount of data they receive from Twitch. While simple users are not obliged to share any personal data except for their user name and email address, they also have no access to other users' data. Developers and broadcasters, on the other hand, are obliged to share extended personal information such as their postal mail address and their billing details. In exchange for that, they do not only get advanced functionality of the platform, but they also gain access to information on other users. From a developer's point of view, the communication between developers and Twitch takes place over the so-called *Twitch API* [25, 26]. The corresponding data flow is visualized in Fig.12.

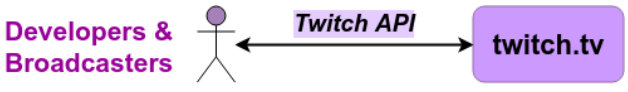

Figure 12: Data flow between Twitch, broadcasters and developers

The last two major entities taking part in the data flow are the *U.S. Government* and the *Local Government* of the country where Twitch is being

[5]The developers are frequently employees of gaming companies that want to embed Twitch streaming content into their gaming consoles/software. An example for such a console is the Blizzard Battle.net Desktop App: [24].

used (Fig.13). The reason why the U.S. Government is listed is because Twitch headquarters are based in San Francisco, California. The relationship between Twitch and these two [6] entities is mostly unilateral. This is indicated by directed arrows from Twitch to the respective government in our graphic. The arrows are featured in a dotted notation, because Twitch does not view governmental structures as their direct partners. The possible data sharing with these entities is only listed in one of the paragraphs of Twitch's Privacy Policy. Hence, some of the platform's users might not directly be aware of the consequences of this relationship. The circumstances, under which user data might be shared with governmental structures we have previously covered in the Privacy Policy section.

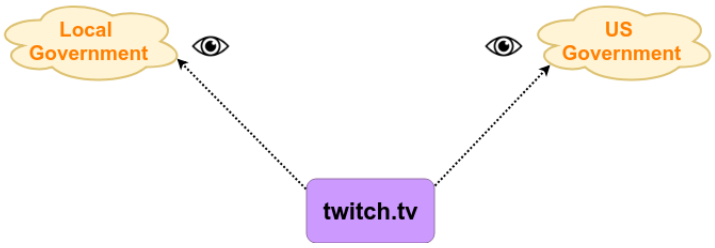

Figure 13: Data flow between Twitch and governmental structures

We colored 'The Twitch Universe' accordingly to represent the data flow between the different entity classes. The full version of the data flow scheme can be seen in Fig.14.

In addition to the described data flow possibilities, Twitch states that in the event of a merger or sale, the entire user data saved on the platform's servers might need to be completely transferred to another company's storage servers. Such a procedure could result in major changes to the Privacy Policy. In this case, we recommend the user to timely take the policy changes into account, in order to make sure their data is treated as confidentially as before.

5.4 Experimental Findings

In order to determine the extent of tracking measures Twitch uses, we decided to conduct several experiments based on the information we were able to extract from its policies. We started with a 'clean' run to determine what sort of Twitch experience an ordinary user gets. For this purpose, we kept the allowance for all kinds of cookies to be placed on our experimental computer, switched off the VPN service, and took some other measures to make sure we behave like an ordinary user.

[6] Or more, depending on how many countries are involved at once

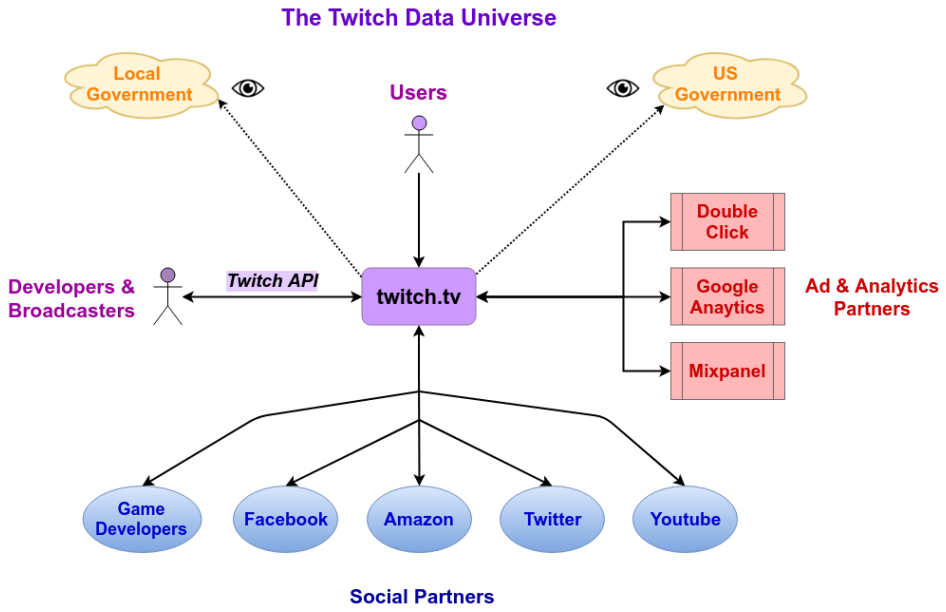

Figure 14: The Twitch Data Universe

Within this 'clean' run, the plugin Ghostery[7], which we preinstalled in our Google Chrome browser, indicated right away that there is a wide range of cookies used by the website. The particular cookies tracking us were:

1. Advertising Cookies
    - Amazon Associates
    - Datalogix
    - ScoreCard Research Beacon
    - Krux Digital
    - Google Publisher Tags
    - Google IMA
2. Analytics Cookies
    - Google Analytics
    - INFOnline

Interestingly, Ghostery did not list any Essential Cookies used on the homepage of `twitch.tv`, even though Twitch lists *twitch_session_id* as one of the

[7]Available under [27]

Essential Cookies they are using. Our first conclusion was that Ghostery might filter out the essential ones, while reducing the set of cookies shown to solely active trackers. To give ground to this assumption, we conducted the same experiment on other websites while using Ghostery. Our finding was that Ghostery, in fact, does show Essential Cookies after all, and even classifies them as such.

Having found this out, we decided to check whether there are any cookies at all that are essential for Twitch to work properly. For this purpose, we went into our browser settings and checked the box to not allow any cookies and other trackers on our computer. After this, we once again tried visiting the homepage of Twitch. We instantly noticed that the page did not load at all. This leads us to the conclusion that Twitch does use Essential Cookies, just as the platform states itself in its Privacy and Cookie Policies. Hence, it seems that the plugin we used for determining the cookies does not always recognize the Essential Cookies for some reason.

When looking into the list of cookies locally stored on our computer, we were able to find a wide range of cookies listed under the label `twitch.tv` in our browser's settings (Fig.15).

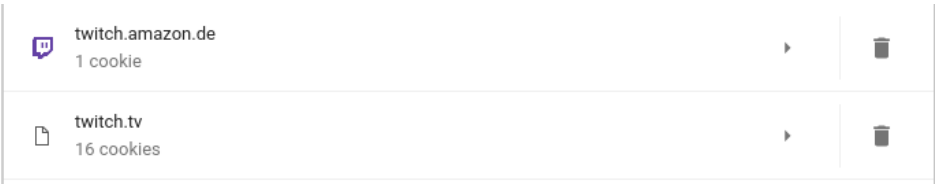

Figure 15: Section of our browser's cookie list with a drop list dedicated to twitch.tv

The number of cookies listed was 16. Our first guess was that some of these cookies could be outdated and derive from some previous Twitch usage on our experimental PC. We were proved wrong when we looked at the creation time stamp of each and every one of these cookies. All of the cookies seem to have been created within the time scope of our research, which consisted of several working units spread over one month. The cookies listed can be seen in Fig.17 (appendix).

At a closer look, most of the cookies proved to be structured similarly to the way we described it in the Technical Background (Section 3, Page 3). An example is given in Fig.18 (appendix).

Another change in visual experience we noticed was that we got different advertisements shown when visiting a sample Twitch channel with cookies switched on and off. When cookie services were on, we had the ad by `pokerstars.de` played, whereas with cookies off we viewed an ad by

`Epicenter.gg`. We were able to reproduce the outcome of this experiment by switching the 'incognito mode' on in our browser instead of disabling cookies. Apparently, the 'incognito mode' automatically switches off cookies and thus produces the same result. Besides the different advertisements, the streaming channel itself, which is shown on the homepage of `twitch.tv` on entrance did change as well, depending on whether we had cookies allowed or not. We were not able to identify a particular pattern though, as the welcoming streams were sometimes different and sometimes the same depending on the time of the day when we accessed Twitch. We assume that Twitch does use cookies to determine which stream should currently be shown to a particular user. It seems, however, likely that the algorithm behind this decision is complex and thus does not always produce a different result than if there were no cookies in use.

While extending our experimental research on advertising, we noticed cases when our first visit on a Twitch streaming page caused an Amazon Prime advertisement to be shown, while any subsequent refreshing of the same page did not feature any ads at all. This might be a convenience feature, as Twitch users tend to refresh the page of the stream they are currently watching whenever the stream starts lagging or does not seem to load. When testing Twitch streams for an extended period of time, we noticed that such service behavior is quite frequent. The reason for this is that the Internet connection of each and single streamer may cause interruptions from time to time. This is especially the case in countries which do not provide solid Internet quality.

In order to test whether Twitch optimizes the advertisements 'on the flow', we visited some of the commercial websites we thought Twitch might be partnered with to produce more income for the companies. Considering that Twitch is a streaming platform that is mostly used for video gaming content, the likes of *MSI* and *Razer* [8] seemed fitted for this experiment, as these companies produce some of the popular gaming hardware. Nevertheless, after visiting their respective websites and also some others, we did not notice any change in Twitch's advertisement policy in favor of either *MSI* or *Razer*. There are two possible explanations for that, namely:

1. Twitch is either not partnered with the specific companies we experimented with or simply not to an extent where it would optimize the user's ad experience in these companies' favor.

2. Twitch does optimize its advertisements in favor of such companies, but

[8] `https://www.msi.com/index.php` and `https://www.razer.com/eu-en`, respectively.

the algorithm is too slow on conclusions or too complex for any changes to take place within the timeframe of our experiment. This explanation seems likely in regard to modern machine learning algorithms, as their outcome is influenced by many different input parameters and not just by one or two, like visiting *MSI* and *Razer* webpages for instance.

Another case of cookie usage that we noticed, is that whenever we log in to the Twitch services with an account that is connected to *Amazon Prime* services, we do not get any advertisements shown. This behavior is totally expected, considering that this is one of the widely advertised features of using both Amazon Prime and Twitch. Nevertheless, it is a valid example of how Twitch uses its Essential and Functionality cookies to improve user experience on its website. We mentioned gaining access to secured areas or features (like Amazon Prime features) of the website in the section on Essential Cookies (Section 5.1.1, Page 14).

6 Conclusion/Outlook

In conclusion, Twitch is tracking its users just as every major internet player does. The extent of the tracking does by no means exceed the basic mechanisms that informed users have come to expect from using corporate services. During our research, we could not find evidence of any violation of what is stated in the privacy policy, which itself is for the most part comprehensive and concise. The Twitch privacy policy also offers hints on how to opt-out of tracking on their platform, which revolves around individual opt-outs from all third-party tracking providers Twitch is using. The feasibility of this approach is highly questionable, since the providers might change arbitrarily without prior notice. Furthermore, opt-out processes are not standardized or streamlined in any way, such that the effort of doing so for multiple services is hard to estimate. An early effort, Do-Not-Track (DNT) requests are widely disregarded because there is no standard interpretation and no legal liability connected to them. Flat out disabling cookies, and therefore making tracking impossible, is no working solution, since Twitch becomes unusable in such a configuration. Particularly with the introduction of the General Data Protection Regulation (GDPR) in Europe, this may change in the future since consent to tracking is not necessary for Twitch to serve streams (lawful basis for data usage). Since opting out is no feasible solution, the only thing left to do for guaranteed protection is to not use the service at all. Todays online services are so closely interwoven with tracking partners and with user data fueling whole businesses that the act of using the Internet already means consenting to being tracked and user data to be utilized.

This imposes an array of ethical and practical questions which have to be discussed by modern society as a whole, the GDPR being the most recent political process. We strongly expect many more discussions on data privacy to arise in the future than it has been the case so far. Currently, the most viable compromise between usability and privacy seems to be to only register and login to services when it is deemed absolutely necessary, and, more importantly, to retain cookies for the current session only (usually this would mean until closing the browser). This obviously means that logins do not persist across sessions, this represents the trade-off for not being trackable across sessions.

References

[1] M. Salganik. Bit by bit: social research in the digital age. http://www.bitbybitbook.com/en/preface, 2016. [Accessed: 08.07.2018].

[2] W. Christl. Corporate surveillance in everyday life. http://crackedlabs.org/en/corporate-surveillance, 2017. [Accessed: 08.07.2018].

[3] D. Boyd and K. Crowford. Critical questions for big data. Information, Communication & Society, Vol. 15, Nr. 5, 2012. [Accessed: 08.07.2018].

[4] Blog on data science by twitch inc. https://blog.twitch.tv/tagged/data-science. [Accessed: 08.07.2018].

[5] Roy T. Fielding, James Gettys, Jeffrey C. Mogul, Henrik Frystyk Nielsen, Larry Masinter, Paul J. Leach, and Tim Berners-Lee. Hypertext transfer protocol – http/1.1. RFC 2616, RFC Editor, June 1999. http://www.rfc-editor.org/rfc/rfc2616.txt.

[6] A. Barth. Http state management mechanism. RFC 6265, RFC Editor, April 2011. http://www.rfc-editor.org/rfc/rfc6265.txt.

[7] Jon Postel. Transmission control protocol. STD 7, RFC Editor, September 1981. http://www.rfc-editor.org/rfc/rfc793.txt.

[8] Curseforge login via twitch. https://wow.curseforge.com/login?returnUrl=/. [Accessed: 25.05.2018].

[9] Adobe Systems Incorporated. Adobe Flash Platform Achieves Record Adoption. http://news.adobe.com/press-release/

adobe-flash-platform-achieves-record-adoption, January 2009. [Accessed: 14.05.2018].

[10] Electronic Privacy Information Center. Local Shared Objects – "Flash Cookies". https://epic.org/privacy/cookies/flash.html, July 2005. [Accessed: 14.05.2018].

[11] Soltani, Ashkan and Canty, Shannon and Mayo, Quentin and Thomas, Lauren and Hoofnagle, Chris Jay. Flash Cookies and Privacy. 2009.

[12] Central District of California United States District Court. Lawsuit CV10-05484. https://www.wired.com/images_blogs/threatlevel/2010/07/CV10-5484-GW-JCGx-Complaint-Summons-Civil-Case-Cover-Sheet1.pdf, July 2010. [Accessed: 12.05.2018].

[13] Brian Tarran. Judge approves Quantcast and Clearspring settlement. https://www.research-live.com/article/news/judge-approves-quantcast-and-clearspring-settlement/id/4005437, June 2011. [Accessed: 12.05.2018].

[14] Parisa Tabriz, Google Director of Engineering. The Long Winding Road from Idea to Impact in Web Security. http://www.federalreserve.gov/boarddocs/speeches/1996/19961205.htm, February 2018. Keynote of Parisa Tabriz at the 2018 Network and Distributed System Security Symposium [Accessed: 12.05.2018].

[15] Adobe Corporate Communications. Flash & The Future of Interactive Content. http://www.federalreserve.gov/boarddocs/speeches/1996/19961205.htm, July 2017. [Accessed: 12.05.2018].

[16] The history of the do not track header. http://paranoia.dubfire.net/2011/01/history-of-do-not-track-header.html. [Accessed: 10.07.2018].

[17] Tracking preference expression. https://www.w3.org/TR/tracking-dnt/. [Accessed: 10.07.2018].

[18] Andreas Kroll. *Computational Intelligence*. 2016.

[19] The new face of gamers - results of a survey. http://twitch.wpengine.com/wp-content/uploads/2014/06/TheNewFaceofGamers1.pdf. [Accessed: 02.07.2018].

[20] Twitch: 355bn minutes watched in 2017. https://www.digitaltveurope.com/2018/02/12/twitch-355bn-minutes-watched-in-2017. [Accessed: 11.06.2018].

[21] Everything you need to know about twitch subscriptions. https://www.lifewire.com/twitch-subscriptions-4147319. [Accessed: 08.07.2018].

[22] Cookie policy by "Twitch Interactive, Inc.". https://www.twitch.tv/p/legal/cookie-policy. [Accessed: 22.05.2018].

[23] Privacy policy by "Twitch Interactive, Inc.". https://www.twitch.tv/p/legal/privacy-policy. [Accessed: 22.05.2018].

[24] Blizzard battle.net desktop app by "Blizzard Entertainment, Inc.". https://www.blizzard.com/de-de/apps/battle.net/desktop. [Accessed: 22.05.2018].

[25] Twitch developers platform by "Twitch Interactive, Inc.". https://dev.twitch.tv. [Accessed: 22.05.2018].

[26] Twitch developers guide and documentation by "Twitch Interactive, Inc.". https://dev.twitch.tv/docs/api. [Accessed: 22.05.2018].

[27] Ghostery plugin by "Cliqz International GmbH". https://www.ghostery.com. [Accessed: 22.05.2018].

Appendices

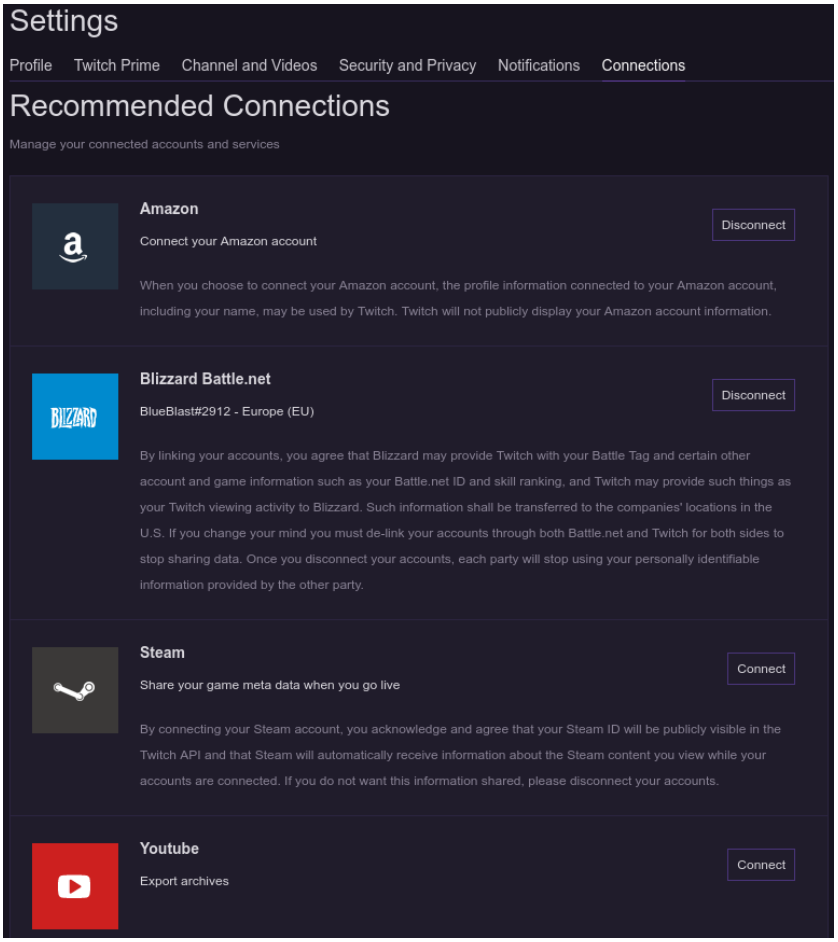

Figure 16: Management of account connections in Twitch user profile settings

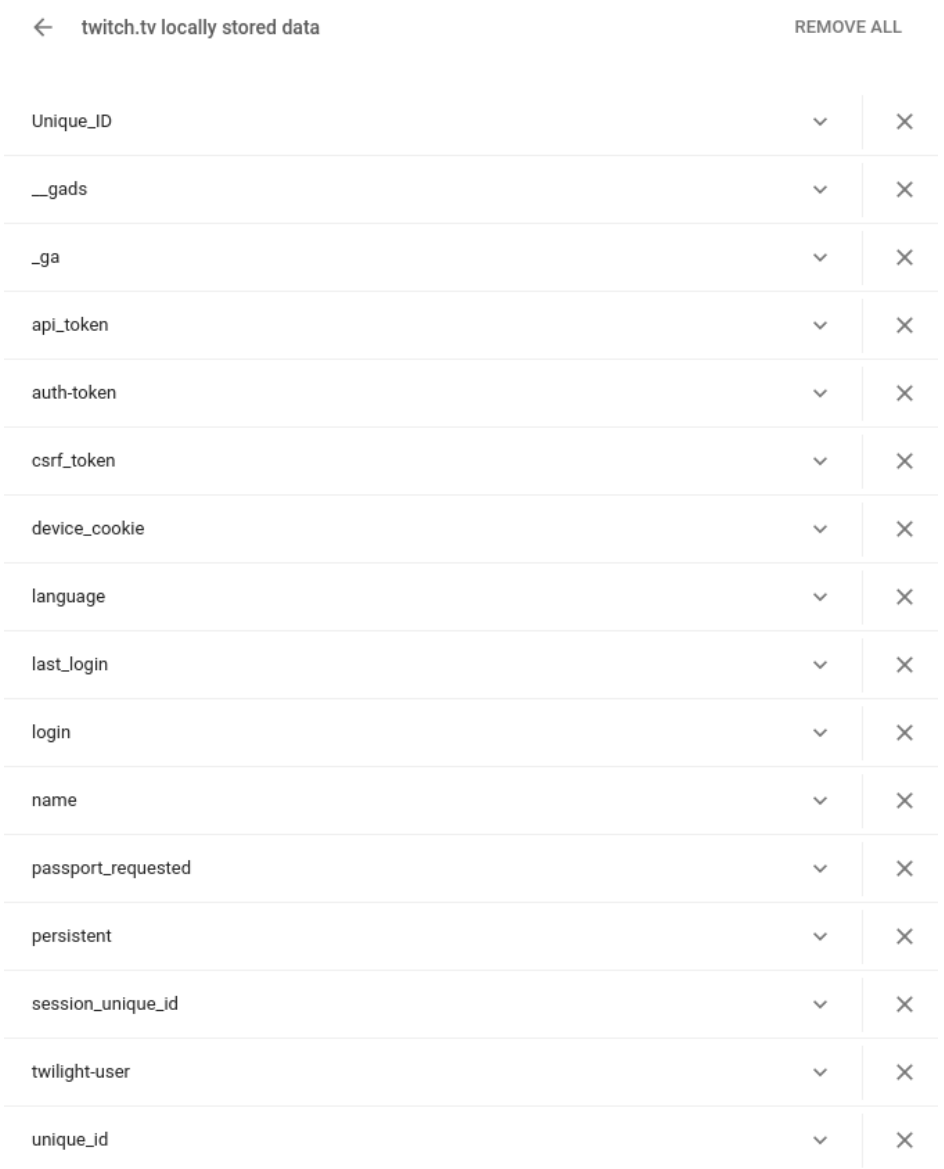

Figure 17: Expanded view of twitch.tv cookies list

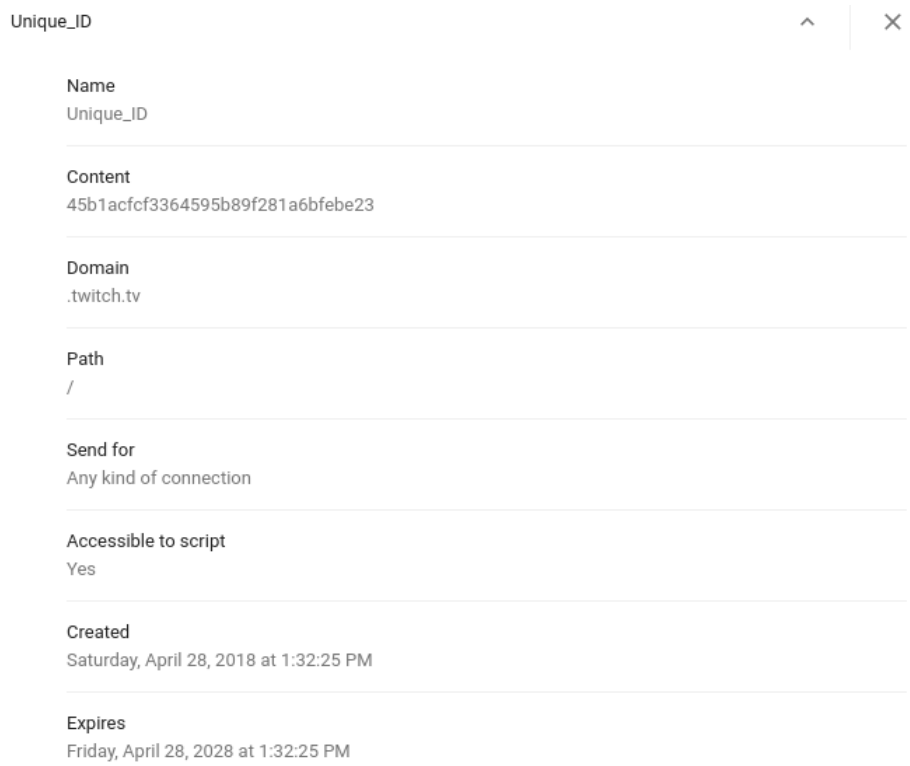

Figure 18: Structure of the *Unique_ID* cookie in the list of twitch.tv cookies