Twitch Analyzed - Following the Data Journeys

Twitch.tv - a global streaming service

With modern technology such as smartphones, the ubiquity of fast internet access and the youth's widespread interest in digital media and video games, markets as well as the e-sports and gaming scene are growing rapidly. This is also true for the landscape of competitive e-sports, the video gaming industry itself, and hardware, software or service providers such as betting sites or digital marketplaces. Twitch.tv (Twitch) is an important hub for this digital evolution: it is an online streaming platform accounting for 355bn minutes of watched streams from all over the world in 2017. It provides broadcasters with the needed technology to stream games, professionally produced e-sports events or even outdoor activities, while viewers can enjoy the broadcasted content and comment on it in live chat. Naturally, it is an ideal location for brand marketing and product placement.

Growing Privacy Concerns

Data is often referenced to as the "oil" of the 21st century, as it is a driving force behind industries. Data gives us, through tools of computer science, the power to optimize chains of production, discover new correlations in given contexts and create self sufficient systems. Because this also entails the personal tendencies of internet users, data can often be directly used for advertising and marketing purposes. The process of collecting data from increasingly many areas of human life and of transforming this information into economic value is called 'datafication'. "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" - this is how Wolfie Christl, a digital rights activist based in Vienna, sees the current development in the area of data collection. Social networks have traditionally been generators of suitable data, and have been extensively criticised as such. As our society tries to keep them in check, privacy matters are of rising concern for users of digital services - an understanding for the value of user data is starting to develop.

Analyzing Twitch

As digital natives and avid gamers ourselves, we want to scrutinize how a platform like Twitch operates with regards to datafication. While it is impossible to monitor what exactly Twitch does to gather data about its users and in what ways this data is used, shared and monetized, we still aim to provide a broad picture of what happens behind the scenes. In order to do so, we will introduce the most common data collection and tracking techniques, focusing on the topic of cookies, particularly third-party cookies, which concern any user of the world wide web. This lays a foundation which is not only helpful for a better insight into tracking on large corporate websites, but will guide our understanding of Twitchs privacy and data policies. We will thoroughly analyze these policies, which are the starting points to lay out the network of business and advertising partners, the various forms of data flow from and to the Twitch platform including user and payment data; a further consideration being the Twitch streaming API. We hope to untangle and clarify an area which is mostly unnoticed by the majority of users, while assessing possible misuses of data by this network and potential

risks connected with using the Twitch platform.

Cookies

The original purpose of cookies is to provide an important mechanism to websites: the notion of state. Since http, the protocol used to request and deliver websites, is agnostic to state, features such as logging in, saving preferences or dynamic shopping carts are impossible to implement. To work around this limitation, web browsers can save pieces of state data on behalf of websites – this kind of data is called a cookie. Even third-party cookies, which are set not by the website a user is currently on, but some external element, have very legitimate uses: without them, there would be no "Like on Facebook" buttons on articles, no external survey polls etc.

But these mechanisms also work quite nicely to track a user around the Internet: a web server can recognize the user on every page on which external content of this web server is present. The reach of services like Facebook is quite extreme in this regard. Understanding the technical possibilities for tracking and keeping them in mind while dissecting cookie and privacy policies is crucial to be able to grasp the actual consequences of the provided information.

Usage of cookies

While there are six different types of cookies that Twitch states it might use, we quickly notice that in fact Twitch does not really use most of them. We were able to determine this by adding the plugin Ghostery to our browser, which is a free-of-charge service to ''track the trackers'', i.e. to find out what cookies are being used to track our online activities. The only types of cookies which are set when entering Twitch are essential and advertising cookies (as classified by the Twitch cookie policy). This leads us to the conclusion that Twitch, like many other big Internet companies nowadays, simply states in its policies that they are allowed to use any kind of legal tracking mechanism, just in case they might want to in the future. Now, the problem we have as users is that we do not know exactly how many and what kind of cookies are tracking our online behavior. There might have been two of them yesterday and there could be 20 tomorrow, without any change to Twitchs cookie policy.

Data types

So let's focus on what data is currently being collected:

At this point, the company distinguishes between data that users provide to Twitch themselves, such as their user name and their email address, and data that Twitch collects without making it obvious to its users. The latter sort of data includes metadata such as your IP address, the type of device you use to get online, the type of your browser machine and operating system (Android, iOS, etc.) as well as more complex data.

Data flow

So there is indeed some data that Twitch gathers from you as a user. But shouldn't it be safe to provide your payment information to such a well-known company like Twitch? While scanning their privacy policy for important information, we were able to extract the main partners and entities that Twitch shares your data with.

Data flow between Twitch and its users

On the one hand, we have the simple user of services offered by Twitch. The user obviously does not receive any personal data on other users or Twitch partners, and only serves as an origin to the information Twitch collects. This is illustrated in Fig. 1. 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 becomes significantly easier as soon as the user connected their Twitch account to one or more of the partners.
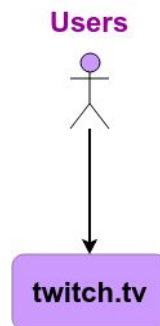

Fig. 1: *Data flow between Twitch and its users*

Data flow between Twitch and its partners

As Twitch partners want something in exchange for the provided information, the platform usually shares the newly acquired data on the registered user with its partners (Fig. 2). The privacy policy does not directly state that Twitch is obligated to do so. In the times of capitalism though, it seems likely that Twitch has to share its users' data the same way its partners do. This is especially the case for social network partners, like e.g. Facebook and Twitter, as they do not receive monetary compensation for their services in contrast to ad and analytics partners like e.g. Google Analytics.
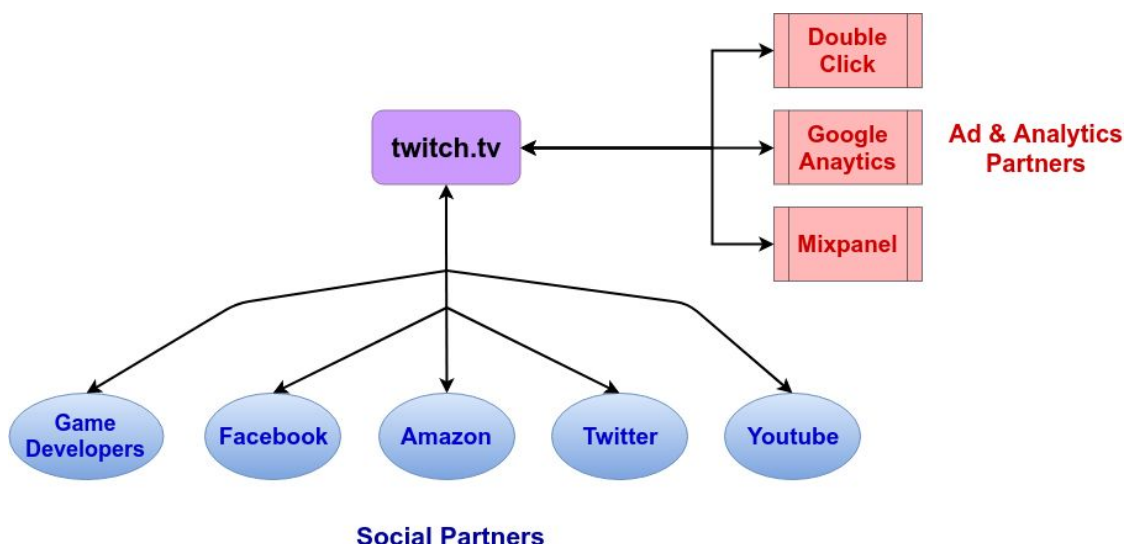

Fig. 2: *Data flow between Twitch and its partners*

To make sharing information with its partners easier, Twitch offers a feature in the user 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.

Data flow between Twitch, broadcasters and developers

In addition to Twitch partners and simple users, there are also the "advanced" users. These can be either broadcasters or developers. Broadcasters are the ones providing the streams users watch. Developers are the ones who embed Twitch services in an external environment (like e.g. a gaming console) as well as produce various features and extensions for streaming channels.
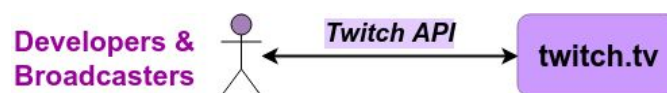

Fig. 4: *Data flow between Twitch, broadcasters and developers*

The main difference between simple users and advanced users is how much data they have to share with Twitch and how much 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 their personal information such as their postal mail address and their billing details. In exchange for this, they do not only get advanced functionality of the platform, but also gain access to information on other users, whenever their extension is being used. "Interaction by a viewer with an Extension or with the channel on which the Extension is activated will provide the developer that operates the Extension with viewer information" is what the privacy policy states. The user information gathered by this mechanism is limited to your metadata, such as previously mentioned in the 'Data Types' part. This means that your user name and ID are only disclosed to the developers if you actively click "Grant Access" within the extension or you install an extension on your own channel. From a developer's point of view, the communication between developers and Twitch takes place over the so-called Twitch API, which is a developer platform created by Twitch. The explained data movements are illustrated in Fig. 4.

Data flow between Twitch and governmental structures

Last but not least, we have the U.S. government as well as the local government of the country where Twitch is being used (Fig. 5). The reason why we mention the U.S. government is because Twitch headquarters are based in San Francisco, California. The arrows have a dotted notation, because Twitch does not see governmental structures as their direct partners. Sharing data with these entities is only listed in one of the paragraphs of Twitch's privacy policy, so that some users might not be aware of the consequences of this relationship.
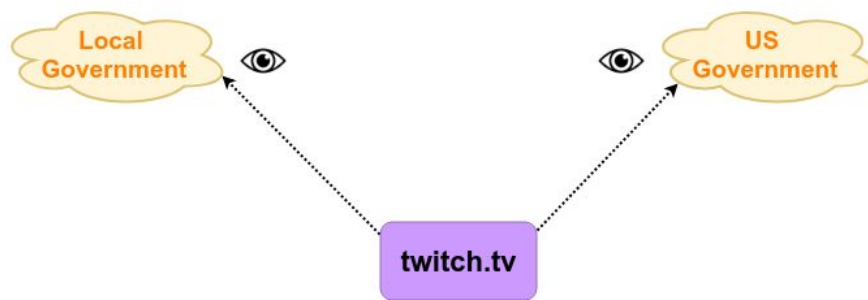
Fig. 5: *Data flow between Twitch and governmental structures*

The Twitch data universe

For the full picture, we summed up the data journeys within 'The Twitch Universe' in Fig. 6. It represents the information flow between Twitch and the entities they exchange data with.
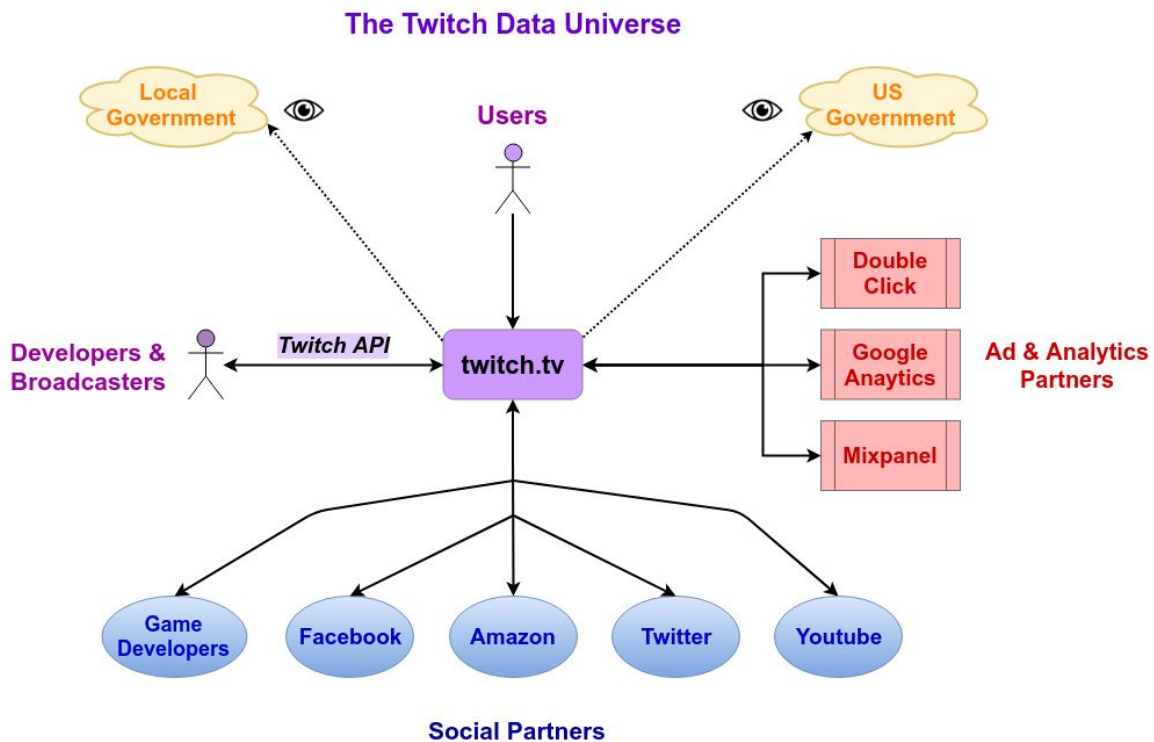

Fig. 6: *Data journeys in the Twitch universe*

Consequences for the end consumer

On first sight it might not seem critical to provide your credit card details to Twitch for spending $2 on those neat, brand-new emotes to use them in chat during your favorite stream. While you may have anticipated that Twitch shares your data with Amazon, as there is a ton of advertising on the benefits you get when you connect your Amazon Prime account to Twitch, some other entities might be not quite what you had initially expected. Among

others, your sensitive billing information may now, without further ado, appear in the archives of the U.S. government or your chatting behavior on the PC of a Twitch broadcaster. Even your love for Logitech keyboards might end up on the servers of some advertising company. While this does not seem obvious, imagine that you watched a Logitech ad during some stream on Twitch and clicked on it or even decided to search for Logitech products on Amazon. Because Twitch, Amazon and some advertising partners are all interconnected, all of these advertising companies can potentially know what your shopping preferences are.

Criticism

Having done an extensive research on the tracking mechanisms Twitch uses and on its data sharing policies, it seems unlikely that Twitch itself misuses its users' data. The reason for this is that the company's profits are directly linked to its reputation among the user base. But what about its partners? What about the U.S. government? What about Facebook, which has recently been involved in some major privacy scandals? Twitch openly states that it does not guarantee that third-parties comply with its own privacy policy, even though data is being shared with them. Therefore, you would have to read the privacy policy of every single of Twitchs partners in order to understand what your data is being used for. Now, what is the privacy policy of the U.S. government?

Sources:

[1] https://abcnews.go.com/Technology/story?id=5817835&page=1, [Accessed: 11.06.2018].
[2]
https://venturebeat.com/2018/04/30/newzoo-global-games-expected-to-hit-180-1-billion-in-re
venues-2021, [Accessed: 11.06.2018].
[3] https://www.digitaltveurope.com/2018/02/12/twitch-355bn-minutes-watched-in-2017,
[Accessed: 11.06.2018].
[4] Ghostery Plugin by Cliqz International GmbH, https://www.ghostery.com,
[Accessed: 22.95.2018].
[5] https://www.youtube.com/watch?v=cyJosQBtzsw, [Accessed: 11.06.2018].
[6] Cookie Policy by Twitch Interactive, Inc., https://www.twitch.tv/p/legal/cookie-policy,
[Accessed 22.05.2018].
[7] Privacy Policy by Twitch Interactive, Inc., https://www.twitch.tv/p/legal/privacy-policy,
[Accessed 22.05.2018].
[8] "Corporate surveillance in everyday life", W. Christl, 2017,
http://crackedlabs.org/en/corporate-surveillance/#10, [Accessed: 19.06.2018].
[9]
https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-lon
ger-oil-but-data, [Accessed: 11.06.2018].
[10] "The Rise of Big Data", K. Cukier; V. Mayer-Schoenberger, 2013,
https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data, [Accessed: 19.06.2018].
[11] "Impact and value of big data across organisations and networks", A.-F. van Veenstra,
https://inclusivegrowth.be/downloads/calls/call-40-ceps/07-van-veenstra.pdf,
[Accessed: 19.06.2018].

Oleksii Kulikov
21 Jahre
Bachelor Informatik seit 2015

Jurek Olden
24 Jahre
Bachelor Informatik seit 2014

Tim Bierth
23 Jahre
Bachelor Informatik seit 2014