A Longitudinal Characterization of the Third-Party Authentication Landscape

Oscar Järpehult Fredrik Josefsson Ågren Madeleine Bäckström Linn Hallonqvist Niklas Carlsson
Linköping University, Sweden

***Abstract*—Many websites offer users to authenticate using third-party identity providers (IDPs) such as Facebook or Google. As part of the signup process, these websites often ask the user to give them additional permissions with the IDP (e.g., some data sharing or authorize some actions) that can have significant privacy implications. Motivated by the increased scrutiny of Facebook and other popular IDPs (e.g., due to the 2018 Cambridge Analytica scandal), we present a longitudinal analysis of the IDP usage and permissions changes over the past nine years (2012-2021) as well as a large-scale characterization of the current state. Our longitudinal analysis identifies trends and characterizes changes in both the IDP usage and permission agreements of different subsets of websites. For our large-scale analysis, we develop and share a Selenium-based measurement framework that we use to collect datasets. Using this data, we study the IDP usage across popularity ranges, the permissions used in the wild, and highlight differences between websites using different IDPs and those that do not. Our analysis shows increased IDP usage, especially among the most popular websites, and that the permission requests on average are becoming more modest but also brings forward significant exceptions that may need further scrutiny.**

I. Introduction

Many modern websites offer additional service only to users authenticated via a third-party *Single sign-on (SSO)*. With SSO, a user wanting to sign up for a service offered by website A can authenticate themselves using their existing account with a third-party service B. In this case, website A is called a *relaying party (RP)* and service B an *identity provider (IDP)*.

Today, almost all SSO is implemented using the *OAuth protocol*. In addition to authentication (for basic SSO), OAuth supports authorization of data sharing between the IDP and RP as well as authorization of the RP performing actions on behalf of the user on the IDP. When first using an IDP to sign up with an RP, the user is therefore asked to agree on a set of application permissions requested by the RP. Here, we call such a permission agreement an *app-rights agreement*. As an example, a user signing up for an RP using popular IDPs such as Facebook or Twitter, may be asked to allow the RP to access some subset of the user's profile on the IDP (information from IDP to RP) or to post information on the user's IDP profile (RP performing actions on behalf of user).

While third-party authentication (and authorization!) can help websites provide better service, their widespread use also comes with security and privacy implications. For example, recent Facebook outages have shown how reliance on a single IDP hindered 100s of millions of people from accessing services that required Facebook login [1]. Second, the more RPs using the same IDP, the more damage can be achieved by an attacker successfully compromising the user's IDP account (as the attacker can control the user's profiles across all RPs).

ISBN 978-3-903176-48-5 © 2022 IFIP

Third, generous app-rights agreements may provide RPs access to sensitive information that may compromise the user's privacy or may provide the RP permissions that could allow the RP to make actions that could negatively impact the user's reputation. Finally, multi-step cross-site information leakage has been demonstrated, in which the information from one IDP can be leaked via the RP to a different IDP [2].

A comprehensive characterization of the third-party authentication landscape seen in 2014 was provided by Vapen et al. [2], [3]. However, since then many of the major IDPs (e.g., Facebook, Google, Twitter) have seen a lot of scrutiny due to their data sharing practices with applications, including the RPs discussed here. One major contributor to the added scrutiny was the 2018 Cambridge Analytica scandal [4], [5], in which Cambridge Analytica was found to have collected user data for 50 million Facebook users without their consent and then used the data for targeted political advertising during the 2016 US election. This scandal resulted in Facebook's CEO, Mark Zuckerberg, having to testify in front of Congress and to publicly apologize for their role. While the scandal increased public awareness of the information that several big companies have access too, no prior work has studied how the third-party authentication landscape have changed over this time.

In this paper we present the first longitudinal study of the third-party IDP landscape that spans both an extensive period before and after these events, as well as the first large-scale measurement study of the app-rights agreements observed in the wild. The study is based on a combination of manually collected information from 500+ websites followed over time and 14,000+ websites that we crawled using an automation tool developed within the project. The manual data collection augment existing data from 2012-2015 with new snapshots from 2019-2021, spanning a nine-year period. We next summarize the main contributions.

- Our longitudinal analysis (§ II) of the IDP usage identifies trends, characterizes long-term changes, and quantifies the churn in the IDP usage, who acts as an RP, as well as the IDP usage of different subsets of websites.
- Our longitudinal analysis (§ II) of the app-rights agreements highlights interesting changes of the app-rights seen for individual IDPs and highlights differences in how the different IDPs are used by the RPs.

- We develop and share a Selenium-based measurement framework (§ III-A) for identifying RP-IDP relationships and extracting app-rights agreements. The tool is shown to have very high precision (99.3%) and good recall (66.8%), motivating its use for large-scale data collection.
- Our crawl-based analysis (§ III) spans 14,526 websites. We compare the IDP usage across websites with different popularity (defined using the top-1M lists of Alexa, Majestic, Tranco) and study the app-rights agreements seen for the most popular English-speaking IDPs.
- Our crawl-based analysis of the current RP landscape (§ IV) includes a PCA-based comparison of the website characteristics of RPs and non-RPs, highlights the main characteristics of RPs using the most popular IDPs, and summarizes how RPs most commonly are implemented.

The measurement tool and datasets are shared with the paper.[1] Our findings shows that the IDP usage is increasing, that the IDP usage is by far the highest among the popular domains and among certain website categories (e.g., News and file sharing websites), that Facebook and Google remain the dominating English-speaking IDPs, and that Apple quickly has gained usage since its introduction in 2019. Our RP-based analysis shows that some RPs are much more likely to use certain IDPs (e.g., all but three of the English-speaking news websites that are RPs use Facebook as one of their IDPs), that RPs often use more third-party services in general (than non-RPs), and that they are more likely to pay for an X.509 certificate from DigiCert than using a free certificate from Let's Encrypt (opposite is true for non-RPs).

Our analysis of the app-rights agreements confirms that there is a positive trend in data sharing practices, which has resulted in RPs getting access to less sensitive information on average. For example, the permission practices of Facebook and Google improved substantially between 2015 and 2019. The data permissions still allow substantial information sharing, especially for RPs using Facebook and Twitter. Apple (followed by Google) provides the most restrictive permissions, typically giving the RP access only to minimal profile information, whereas Twitter often gives write permissions to the RP. While Google did well, it provided one of the most extreme cases, as one RP (645voyager.com) requested full access to Google Drive, which for some users could include highly sensitive information that the RP potentially could access or delete given such permissions.

II. Longitudinal IDP usage

A. Manual data collection

TABLE I
Most frequently observed IDPs (Apr. 8, 2021).

Rank	All websites (505)		English speaking (297)	
1	google.com	131	google.com	100
2	facebook.com	128	facebook.com	94
3	apple.com	59	apple.com	48
4	qq.com	33	twitter.com	19
5	weibo.com	30	microsoft.com	8
6	twitter.com	25	linkedin.com	8
7	vk.com	13	github.com	7
8	linkedin.com	9	vk.com	5
>8	Others	76	Others	24
Uniq IDPs	37	–	23	–
Total	–	504	–	313

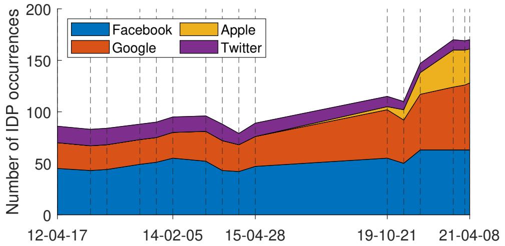


Fig. 1. IDP usage over time for top-200 websites at each snapshot.

For our longitudinal analysis, we augment a historic dataset collected by Vapen et al. [2]. In addition to 10 snapshots collected between Apr. 2012 and Apr. 2015, we add another 6 snapshots collected between Oct. 2019 and Apr. 2021. The combined dataset contains 16 snapshots collected over a nine-year period (Apr. 2012 to Apr. 2021). Furthermore, we selected the dates of the last few snapshots so to analyze the pairwise changes observed at different time scales (e.g., approximately 1 week, 1 month, 6 months, 1.5 years, 6 years, 9 years).

For each snapshot, the union of (1) the websites on the current Alexa top-200 and (2) all websites already in the dataset were manually inspected. During data collection, every website on this list was classified based on the primary service it provides, IDPs were manually identified, and the app-rights agreements of each identified RP-IDP relationship were manually extracted. In total, the dataset includes 505 websites.

B. Top IDPs

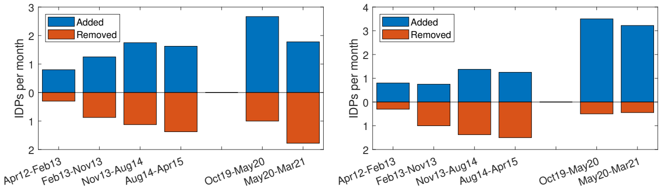
Table I shows the usage frequencies of the eight most commonly used IDP across all websites as well as the English-speaking subset. Here, we use data from the latest snapshot (Apr. 8, 2021). Note that Google, Facebook, Apple are by far the most used IDPs in both sets, and that they together with Twitter (fourth most popular English-speaking IDP) makes up 78.0% of all observed RP-IDP relationships and 83.4% of the relationships seen in the English-speaking subset. In the remainder of the paper, we focus on these top-four IDPs.

In addition to their current dominance in the English-speaking part of the web, this choice allows for easier comparison of the manual and automated results. (All four domains allowed us to implement effective identification of RP-IDP relationships and the extraction of the used app-rights agreements.) Several of these companies are also of special interest due to the privacy scrutiny they currently are facing.

C. IDP usage trends

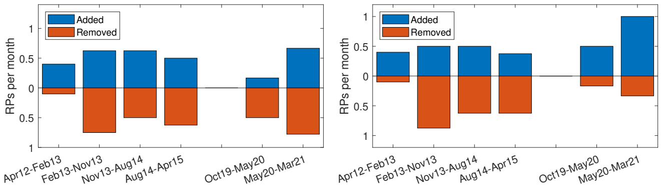
Figure 1 breaks down the usage of the top-four IDPs and how their usage has changed over time. Here, we show timelines for the top-200 set observed at each snapshot. (The results when tracking the original top-200 set is similar.) The overall

[1] Measurement tool and datasets can be found here: https://www.ida.liu.se/~nikca89/papers/networking22.html



(a) Top-200 from 2012-4-17 (b) Top-200 from earlier snapshot

Fig. 2. Short-term monthly IDP churn.



(a) Top-200 from 2012-4-17 (b) Top-200 from earlier snapshot

Fig. 3. Short-term monthly changes in the set of RPs.

increase has been substantial, with Google and Apple seeing the biggest individual increase in usage. While Apple was not observed as IDP until in the first 2019 dataset (when Apple introduced SSO [6]), it has since seen a substantial increase in usage. Twitter, on the other hand, has seen decreased usage.

D. IDP Churn

While the IDP usage has increased, several websites have reduced the number of IDPs they use or stopped being RPs.

IDP churn rates: Figure 2 summarizes how the short-term churn has changed over time. Here, we show the average number of IDPs added (blue) and removed (red), normalized per month, over different roughly 9-month windows. For Figure 2(a) we use the websites from the original top-200 list and for Figure 2(b) we use the top-200 list associated with the earlier of the two snapshots being compared. In both cases, we observe significantly higher rate of IDP additions (than removals) the last two years, compared to 2012-2015.

Changes in RP set: Figure 3 presents the corresponding statistics for the average number of added (blue) and removed (red) RPs, normalized per month, for the same top-200 sets and time periods. Interestingly, the original top-200 set (Figure 3(a)) sees a reduction in the number of RPs, whereas we see an increase when looking at the recent top-200 lists (Figure 3(b)). This observation implies that the websites that have entered the top-200 list are more likely to become RPs than websites fall of the top-200 list. This observation is consistent with our crawl-based results (c.f. Section IV) that shows that IDP usage is highest among the most popular domains but does not answer whether the increases/decreases in popularity are related to their IDP usage or not.

E. RP-based trends

There are substantial differences between the IDP usage among different website categories. Table II shows a heatmap of the fraction of websites associated with each snapshot and website category with at least one IDP. In general, news (e.g., bbc.com, nytimes.com) and file sharing (e.g., github.com, mediafire.com) websites are the most frequent users. However, in contrast to most other categories (most of which have seen increased IDP usage), these two categories have seen reduced IDP usage. Yet, they are still the biggest users of IDPs. The biggest increase in usage is seen among commercial websites (e.g., ikea.com, amazon.com, walmart.com).

TABLE II
LONGITUDINAL IDP USAGE FOR DIFFERENT WEBSITE CATEGORIES.

	All	Social	Tech	Video	News	Info	FileShare	Commerc	Ads	CDN
4/17/2012	0.26	0.19	0.24	0.31	0.58	0.43	1.00	0.09	0.00	0.00
12/4/2012	0.25	0.21	0.20	0.25	0.50	0.38	1.00	0.08	0.00	0.00
2/21/2013	0.24	0.18	0.30	0.25	0.52	0.38	1.00	0.10	0.00	0.00
8/21/2013	0.25	0.18	0.24	0.24	0.48	0.41	1.00	0.13	0.00	0.00
11/18/2013	0.25	0.18	0.19	0.24	0.50	0.47	0.91	0.13	0.00	0.00
2/5/2014	0.25	0.20	0.23	0.24	0.46	0.41	0.91	0.13	0.00	0.00
8/11/2014	0.24	0.20	0.18	0.22	0.41	0.50	0.73	0.17	0.00	0.00
9/30/2014	0.24	0.20	0.19	0.20	0.41	0.45	0.82	0.17	0.00	0.00
12/22/2014	0.25	0.20	0.21	0.19	0.47	0.43	0.56	0.24	0.00	0.00
4/28/2015	0.25	0.22	0.20	0.21	0.47	0.45	0.41	0.24	0.00	0.00
10/21/2019	0.27	0.28	0.21	0.20	0.32	0.28	0.45	0.34	0.00	0.14
3/20/2020	0.27	0.28	0.23	0.20	0.36	0.28	0.45	0.34	0.00	0.14
5/18/2020	0.28	0.29	0.28	0.17	0.35	0.23	0.35	0.36	0.00	0.13
3/8/2021	0.29	0.28	0.32	0.19	0.39	0.21	0.43	0.37	0.00	0.11
4/1/2021	0.30	0.29	0.34	0.20	0.39	0.21	0.43	0.37	0.00	0.11
4/8/2021	0.30	0.29	0.33	0.20	0.39	0.20	0.43	0.37	0.00	0.11

Among the other categories, usage was highest among Tech websites offering technical services (e.g., microsoft.com, adobe.com), Social media websites (e.g., facebook.com, instagram.com, whatsapp.com), Video streaming services (e.g., youtube.com, netflix.com), and Info websites where users can find information (e.g., google.com, imdb.com). The usage was smallest among ad-related services (e.g., adcash.com, onclickads.net) and CDNs (e.g., fbcdn.net).

Current per-category status: Figure 4(a) shows the fraction of websites of each category that use one of the four English-speaking IDPs. Notable observations include: (1) Facebook and Google dominate in all classes except File sharing (Apple and Twitter both see their highest relative share here), Ads (no IDPs), and CDN (only one RP and it uses Google and Apple). (2) All RPs classified as Tech use Google as IDP. This matches the high openID usage (protocol no longer in use) among these sites back in 2012. (3) All RPs classified as News use Facebook. This observation, combined with the category's overall high IDP usage, is interesting given the significant criticism that several news outlets have directed towards Facebook's data sharing.

Except for categories Social (fairly even spread between RPs using 1, 2, or 3 IDPs) and Video (almost half use only one IDP), most website categories typically use two IDPs (in most cases Facebook + Google). This is illustrated in Figure 4(b).

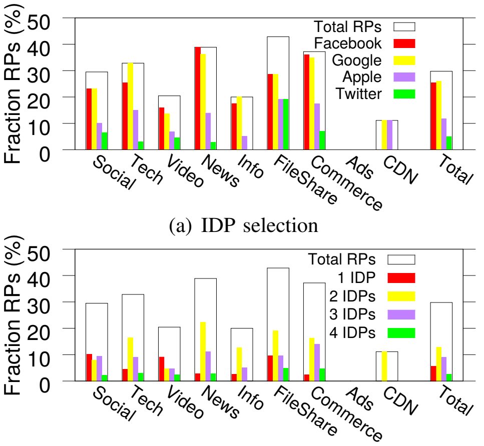
F. App-rights agreement changes

When a user selects to use an IDP, the RP presents the user with an app-rights agreement in which the user needs to agree on some information sharing between the RP and the IDP, and in some cases also on giving the RP some rights on the IDP.

We next summarize how these app-rights changed over time. For this analysis, we only had access to data since 2014-9-30.

Permission classification: For easy compassion with prior work, we use a similar classification as Vapen et al. [2]:

- *Basic Information (B):* Relatively non-private information that often is found online; e.g., name and email address.
- *Personal Information (P):* Personal information (e.g., country, gender, friend list), including of sensitive nature (e.g., religion, sexual orientation, political views).



(a) IDP selection

(b) Number of English-speaking IDPs

Fig. 4. Per-category breakdown of the current IDP usage, as seen 2021-4-8. (Here, we include all websites that have been in the top-200 at some snapshot.)

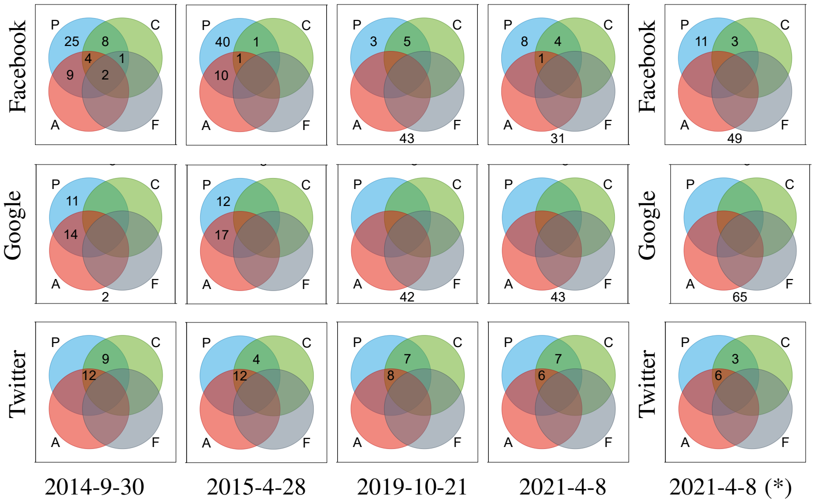


Fig. 5. Permissions at different dates, shown as Venn diagrams. For the first 4 columns we use the top-200 list from 2012-4-17. For column 5 (marked *) we use that date's top-200 list.

- *Created content (C):* This permission class includes directly or indirectly created content which are made by the user; e.g., likes, check-in history, images.
- *Friend's data (F):* This class consists of data of other users; e.g., friends of the user. This data belongs to another user that potentially are non-consenting.
- *Actions on behalf of the user (A):* This class contains the rights for the RP to export data to the IDP and to perform actions on behalf of the user. Actions that could be made are posting information on the user's IDP feed or timeline but also sending messages to other IDP users. Data that could be exported include images, information about the user's actions, and what music the user just listened to.

Figure 5 shows the app-rights changes between Sept. 2014 and Apr. 2021 for the RP-IDP relationships associated with Facebook, Google, and Twitter. (We omitted Apple here, as they were not around for the first two data points shown, and all current Apple agreements only request basic information (B).) Each circle in the Venn diagrams illustrates one of the four non-basic permission classes (P, C, A, F). Any agreement that only requested basic (B) information is shown outside the union of the circles. The first four columns use the websites from the top-200 list on 2014-9-30, while the fifth column uses the top-200 list from the final collection date (2021-4-8).

Longitudinal changes using fixed top-200 set: Overall, we have seen a big shift, especially for Facebook and Google, towards agreements requesting increasingly less sensitive information (e.g., P, F, C) or authorization to make actions (A). For example, in 2014-2015 all agreements except two agreements involving Google asked for more information than just the basic info (B), whereas in 2019-2021 this class was responsible for all relationships involving Google and most (43/51 and 31/44) of all Facebook relationships. For Google, this can in part be explained by the shutdown of the consumer (personal) version of Google+ in April 2019 [7]. For Facebook, the changes may be driven by changes Facebook made to their API specifications in part aimed to reduce the amount of unjustified sharing of information. Here, a big step was taken when Facebook asked all its RPs to upgrade from using version 1.0 to version 2.x of their API no later than May 1, 2015. This change was seen already a few days before the change (Apr. 28, 2015) when the number of RPs requesting friend list information had reduced from 3 to zero.

Interestingly, we still see 13 Facebook agreements and 13 Twitter agreements that ask for additional information in 2021. The most rights are requested by RPs using Twitter. For example, for the 2021 dataset, all Twitter agreements ask for permissions either for P+C (7) or P+C+A (6).

The website that had the P+C+A permission with Facebook in 2021 was livejournal.com. This website temporarily changed down to basic (B) permissions in 2019, just to change back up to requesting P+C+A permissions when using Facebook as IDP. The websites that still used the P+C+A permissions with Twitter in 2021 were mediafire.com, imgur.com, alibaba.com, 4shared.com, livejournal.com, and goo.ne.jp.

In contrast to Google and Facebook, Twitter has not changed their permission policies over the past nine years. Instead, it has consistently been the IDP that shares the most user data and allows the most actions (A) on behalf of the users.

Longitudinal changes using current top-200 set: The currently popular websites appear to be more cautious in the permissions they request than the websites that were most popular in 2012. For example, comparing the last two columns, we observe that the websites based on the most up-to-date top-200 list (i.e., last column) use Google and Facebook to a larger extent and are equally or less likely to request private (P) information or higher: Facebook (13/44 vs. 14/63), Google (0/43 vs. 0/65) and Twitter (13/13 vs. 9/9).

In addition to Google and Facebook actively having addressed some privacy concerns associated with the app-rights agreements, we expect that some RPs have become more aware of the privacy concerns of their users. While it is difficult to quantify which of these factors most have contributed to the trends we observe, we expect that at least some websites have re-evaluated which IDPs they partner with and what information they ask users to share. This change is perhaps most noticeable among newspaper websites.

III. Crawl-based IDP usage analysis

To study the IDP usage for a larger set of websites, we developed a Selenium-based crawler that automatically (1) identifies the IDPs used by each website, (2) extracts the app-rights agreements for any identified RP-IDP relationship, and (3) extracts a long list of website characteristics. We restrict our implementation and analysis to the four most popular English-speaking IDPs: Facebook, Google, Apple, Twitter.

A. Automated collection tool

High-level design: Our data collection framework is implemented in `Node.js`, is built using Selenium with the ChromeDriver, and allows for parallel data collection. The program starts by *setting up N parallel instances*, each responsible to evaluate one website at a time. When evaluating a website, each such instance first (1) *loads the website* in a new browser window, and then, when fully loaded, (2) *stores away information about the loaded page and the objects making up the website*, before (3) *collecting additional website information*, including IP address, geo-information about the IP address, a X.509 certificate for the website, supported TLS versions, source code, etc. After this, the tool (4) crawls the website. In this step, the framework (5) *close popups* that may block the rest of the page, and then iteratively (starting with the landing page) (6a) tries to *identify IDP buttons* either directly or by (6b) first *identify login buttons*, clicking these buttons to display any login options that may be available, and then again try to *identify IDP buttons*. When an IDP is found, the program (7) *extracts IDP data* and saves it. This crawling sequence is for each website of interest.

Details about how each step was implemented, and the optimizations we made to make the tool scalable and efficient are described in an extended version [8].

Additional website info: In addition to IDP information, the tool collects (1) the full website (all elements retrieved when visiting the website), (2) the X.509 certificate used by the website (fetched using the `get-ssl-certificate` and information of interest extracted using OpenSSL), (3) what TLS version were used and what versions are supported (extracted using a sequence of GET request for the page with each version of the TLS protocol specified), (4) IP related information (using a DNS module of `node.js`, the primary IP address is first looked up, followed by GeoIP2 [9] lookups to extract geographic information about the IP address), and (5) network transfer information logs (extracted using Chromedriver) containing information about every downloaded resource.

Limitations: Due to the complexity of the web, it appears practically infeasible to find a general solution that works on all websites. Next, we describe some known limitations. First, our tool only searches for IDP button located either directly on the landing page (one click away), or in a section accessible by first clicking another button (two clicks away). We have found cases were this is not sufficient. For example, fandom.com requires an element to be hovered to display the set of IDPs that can be used, and yelp.se disables IDP buttons until a box is checked. These cases are often highly website specific and were sufficiently rare that we selected not to address these.

Second, some websites use reCAPTCHA and similar mechanism to hinder automation tools. We did not try to work around this but acknowledge that we may not capture all IDPs available to an active user. Finally, while we kept track of failed requests (e.g., due to server-side errors, downtime, or another unexpected issue with the connection), we acknowledge that yet more retries could have improved the success rate further (at the expense of longer run times).

While we are satisfied with the crawler's performance (validation presented next), more work and longer run-times could improve the recall further. For example, we could increase the search depth and modify the scripts to make it harder to detect that it is a crawler (as some websites detect us as a crawler and present obstacles requiring human input). Targeted efforts to improve the recall of Twitter may also be beneficial.

B. Validation and data collection

Validation experiments: We evaluated the tool's accuracy against a manually collected ground truth dataset containing the top-256 websites from Alexa. We found small differences between using 1, 4, 8, or 16 parallel collection instances. Here, we report values for when using 16 parallel instances (used to collect the final dataset). While the tool missed 33.2% of the RP-IDP relations (recall of 66.8%), it had very high precision (99.3%) as it only had one false positive (Google as IDP).

The very good precision of the tool ensures that almost all identified RP-IDP relationships are RP-IDP relationships and that we can successfully retrieve the scope information for these relationships. However, the non-negligible fraction of missed relationship hinders us from reporting the exact usage of different IDPs. The identified RP-IDP relations can therefore be seen as a larger sample set of RP-IDP relationships that would be very time consuming to identify manually.

When comparing the recall rate of the individual IDPs of consideration, we found that Facebook (70.2%), Google (69.1%), and Apple (63.8%) all had fairly similar recall rates, whereas the recall rate of Twitter (42.9%) was relatively lower. This suggests that the usage of the first three can be compared relatively fairly, but that comparisons of Twitter require extra care. Next, we present a crawl-based analysis that focus on comparisons between website classes and the characteristics of RPs using different IDPs. For this analysis, a high precision is desirable, and care is placed avoiding comparisons that are impacted by differences in the recall rates.

Data collection: The dataset analyzed here was collected over four days (March 12-16, 2021). The crawled websites were selected from four popularity brackets identified in the four top-1M lists: Alexa, Majestic, Cisco, and Tranco. From each list, we selected the websites that belonged to one of

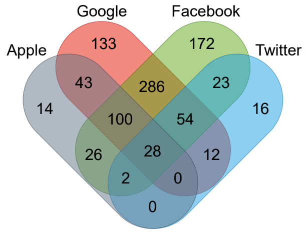
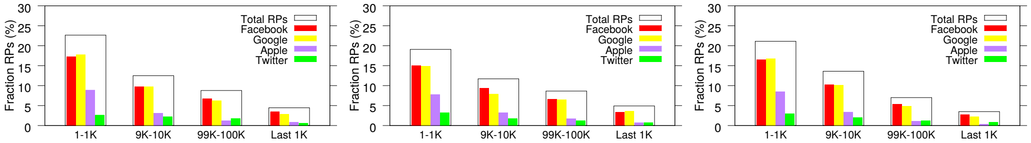
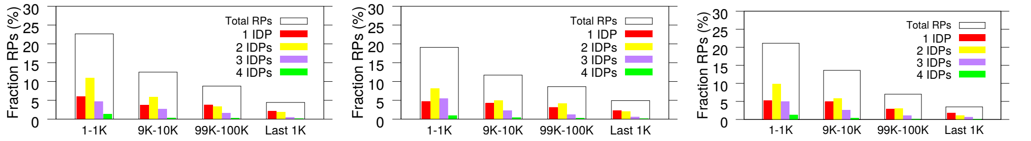


Fig. 6. Venn diagram with all observed IDP combinations.



(a) Alexa (b) Majestic (c) Tranco

Fig. 7. IDP selection among websites with different popularity.



(a) Alexa (b) Majestic (c) Tranco

Fig. 8. Number of IDPs used by websites with different popularity.

the following rank ranges: (1) 1-1000, 9,001-10K, 99,001-100K, and 999,001-1M. Here, it should be noted that we used ranks 497,490-498,491 for the last set of the Alexa list. (These are the 1,000 last entries in the list, as Alexa these days seldom include 1M entries in their top-1M list.) All ranks were extracted on 2021-3-8. After removing duplicates, the final list of domains included 14,526 unique websites that we crawled.

C. High-level IDP usage statistics

We successfully crawled 13,639 websites. Out of these, 909 were classified as RPs using at least one of the four IDPs of interest. Similar to the manual datasets, Facebook and Google were the most frequently identified IDPs. For example, 691 RPs used Facebook (76.0%), 656 used Google (72.2%), 213 used Apple (23.4%), and 135 used Twitter (14.9%).

Figure 6 shows a Venn diagram of the observation frequencies of each IDP combination. We make several observations. First, 574 (63.1%) of the RPs use at least two RPs. Of these RPs, 390 used two IDPs (42.9% of total RPs), 156 used three IDPs (17.1%), and 28 used all four IDPs (3.1%). Second, more than half (51.5%) of the RPs used both Google and Facebook. In most of these cases, no additional IDP was used with them (31.5%). Third, the differences in IDP usage are even greater when considering the cases where only one IDP is used. Here, Facebook is used in 172/335 (51.3%) of the cases and Apple in only 14/335 (4.2%) of the cases. Overall, Facebook was used alone most frequently (172/691 = 24.9%) and Apple the least frequently (14/213 = 6.6%). The big differences perhaps come from Facebook (together with Google) being the most established IDPs that may have historic advantage since many users already use them as IDPs for other services.

Fourth, Apple and Twitter were never used in combination unless (at least) Facebook also were used as IDP. This observation is interesting since it may suggest that RPs selecting to use Twitter and Apple typically are from quite different sets of websites. This hypothesis is supported by our app-rights analysis that shows that Apple is the most restrictive and Twitter the least restrictive (see Sections II-F and III-E).

D. Popularity differences

We next study to what degree websites of different popularity are more or less likely to be RP or using specific IDPs.

Popularity rankings: For this analysis, we used the websites rankings provided by (a) Alexa, (b) Majestic, and (c) Tranco. These three lists capture websites popularity in different ways [10]. Alexa's global top-1M list is the most popular website ranking. While the exact methodology used to generate this list is not public, the list is claimed to be based on the recent web activity of millions of people. Majestic top-1M list is based on data collected using Majestic's web crawler and the number of /24 IPv4- subnets linking to each website. Tranco [11] combines four existing lists (Alexa, Majestic, Cisco Umbrella, and Quantcast) and filters out unavailable or malicious domains. The list aims to improve rank stability and robustness against manipulation. Cisco Umbrella was not used for this analysis, since its ranking is not website based but rather compare the frequency that any FQDN is observed by Cisco's OpenDNS service.

Likelihood being an RP: Popular websites are more likely to be an RP than less popular websites. This is illustrated in Figure 7, where we show the fraction of websites of a particular popularity bracket that was an RP and used each IDP. For all three ranking lists, we observe a significant decrease in IDP usage as the popularity reduce, and the website on the top-1K are 5.1×, 3.9×, and 6.1× more likely to be RP than those in the bottom 1K.

The same trends were observed regardless of which IDP was considered. For all three ranking lists and all four IDPs (3×4 = 12 cases), there is a monotonically decreasing fraction of RPs in each popularity bucket when going from the top-1K towards the last ranked domains. These results clearly show that the most popular domains (who's IDP usage we have manually tracked) are the most likely users of IDPs. The above monotonicity property also holds when considering the fraction of websites that use 1, 2, 3 or 4 of the IDPs. (See Figure 8.) The relationship between popularity and likelihood of being an RP is statistically significant. To see this, consider any of the 27 monotonicity instances (i.e., 3 rankings × (1 overall + 4 IDPs + 4 number of IDPs)). The probability for such instance to be non-decreasing when assuming any order is equally likely is 1/24 = 0.04. The probability that all of them would be non-decreasing under such null-hypothesis assumption is therefore very small (e.g., if assuming independence – they are not – the p-value would have been $(1/24)^{27} = 5.4 \cdot 10^{-38}$).

Relative usage of IDPs: While our tool does not find all IDP instances, given the consistently big differences in usage between the top-2 (Facebook and Google) and the other two

TABLE III
REQUESTED PERMISSION CLASSES BY RPS OF EACH IDP.

Category	B	P	F	C	A
All IDPs	1270 (74.9 %)	425 (25.1 %)	4 (0.2 %)	156 (9.2 %)	88 (5.2 %)
Facebook	561 (81.2 %)	130 (18.8 %)	4 (0.6 %)	20 (2.9 %)	7 (1.0 %)
Google	496 (75.6 %)	160 (24.4 %)	0 (0.0 %)	1 (0.2 %)	1 (0.2 %)
Apple	213 (100.0 %)	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)
Twitter	0 (0.0 %)	135 (100.0 %)	0 (0.0 %)	135 (100.0 %)	80 (59.3 %)

(Apple, Twitter), we argue that the recall rates (63.8-70.2% for Facebook, Google, Apple and 42.9% for Twitter) are sufficient to conclude that Facebook and Google are more likely to be selected over Apple and Twitter for almost all classes.

E. *App-rights comparisons*

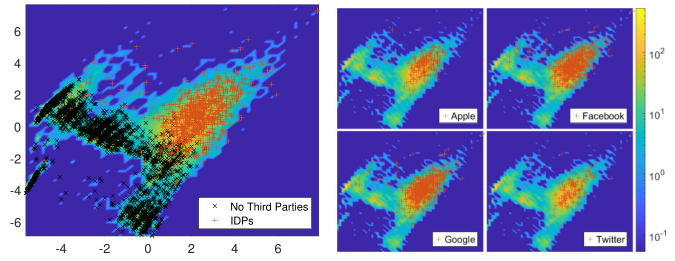
Table III shows the number of times each permission level was requested by an IDP. For Apple, only the basic profile was ever requested by the RPs. Apple only allows sharing of name and email. For Twitter, personal data (P) and content created by the user (C) were always requested, and more than half of RPs also requested to perform actions (A) on behalf of the user. Both Facebook and Google have lots of personal data about users that they may provide. Interestingly, compared to the manual top-200 dataset, RPs using Google in this bigger dataset (spanning also less popular domains) more frequently request private (P) information than RPs using Facebook. Yet, the highest permission classes (e.g., F, C, and A) are still considerably more frequently requested from Facebook.

On Facebook, the authorization requests to perform actions (A) on behalf of the user were either permission to post content or to update profile/page settings. The (single) RP that requested actions on Google was 645voyager.com, who requested full access to Google Drive. This is arguably a very serious type of permission request, since the RP then may be able to edit or delete any/all material created by the user.

IV. CRAWL-BASED RP ANALYSIS

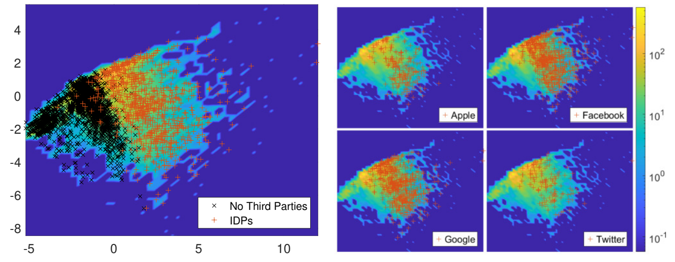
A. *PCA-based Comparison*

To glean some initial insights to whether RPs and non-RPs differ, we first use principal component analysis (PCA). For this analysis, we use two different sets of variables: continuous variables and categorical variables. The first set of parameters include (together with transformations): (1) total transfer size (log-transformed), (2) total resource size (log-transformed), (3) number objects downloaded when visiting landing page, (4) number of observed third parties, (5) number third-party objects, (6) Alexa rank (log-transformed), (7) Cisco rank (log-transformed), (8) Majestic rank (log-transformed), (9) Tranco rank (log-transformed), (10) IP address longitude, and (11) IP address latitude. The binary (categorical) variables we use are: (12) TLS 1.0 supported, (13) TLS 1.1 supported, (14) TLS 1.2 supported, (15) TLS 1.3 supported, (16) Certificate exist, (17) Certificate is valid, (18) IP address is within EU, (19) has keywords meta tag, (20) has robots meta tag, (21) has viewport meta tag, (22) has charset meta tag, and (23) has themecolor meta tag. The first five metrics (1-5) where all obtained from the initial page load (of the landing page).



(a) RPs vs non-RPs w/o 3rd-parties (b) Per-IDP breakdown

Fig. 9. Combined heatmaps of all observed websites and scatter plots of selected websites using the first two principal components based on PCA using all parameter categories (1-23).



(a) RPs vs non-RPs w/o 3rd-parties (b) Per-IDP breakdown

Fig. 10. Combined heatmaps of all observed websites and scatter plots of selected websites using the first two principal components based on PCA using only non-binary parameter categories (1-11).

Figures 9 and 10 show 2D plots combining a heatmap of all observed websites and a scatter plot of different subsets of websites. These two figures use the first two principal components based on using either all variables (Figure 9) and only the non-binary parameter categories (Figures 10), respectively. In the first sub-lots (Figures 9(a) and 10(a)) we plot all RPs and the set of non-RPs that does not use third-parties. Here, we note significant differences between the classes. For example, RPs tend to belong to one of two dominating clusters and this cluster is substantially separated from the main cluster with websites not using third-parties.

Second, we have observed only very limited differences based on who is the IDP (Figures 9(b) and 10(b)) and what app-rights are being used by the RP (included in extended version). These observations suggests that RPs tend to have some underlying characteristics that differ from websites not using third-parties.

B. *Website related RP characteristics*

We have found that RPs typically are larger, use more efficient compression (e.g., have smaller transfer to resource size ratio), and yet require more bytes to be transferred at the time of a page visit. Furthermore, the RPs typically consist of more web object, use much more third-parties, and load much more resources from the third parties. Table IV shows the median and 90^{th}-percentile, respectively, for these six metrics. Note that our observations easily hold for both statistics.

We have not observed any significant differences in the IDP selection of RPs from different top-level domains (TLDs). Figure 11 breaks down the RPs based on their TLD. As expected, .com is by far the most common TLD.

TABLE IV
WEBSITE COMPARISON OF NON-RPS AND VARIOUS CLASSES OF RPS

		All	Non-RPs	RP	Apple	Facebook	Google	Twitter	1 IDP	2 IDPs	3 IDPs	4 IDPs
Median	Transfer size (kB)	737.3	629.7	1711.8	1768.4	1813.4	1666.9	1586.0	1650.5	1791.0	1750.0	1569.4
	Resource size (kB)	1655.1	1402.6	4173.7	4744.3	4395.7	4133.3	3734.2	3753.1	4350.5	4736.6	3680.0
	Transfer/Resource ratio	0.7	0.7	0.4	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.4
	Objects	41.7	37.4	84.9	80.3	88.9	86.0	87.0	76.8	93.7	80.0	86.0
	Third-parties	6.0	5.2	17.0	15.6	17.6	16.8	17.1	16.1	17.9	16.0	17.0
	Third-party objects	14.8	11.9	60.9	66.5	63.1	63.1	63.3	51.4	69.6	60.5	72.5
$90th$ percent	Transfer size (kB)	3934.7	3875.4	4639.5	4516.9	4989.9	4605.3	4614.3	4484.0	4671.8	4910.5	4255.2
	Resource size (kB)	7256.4	6996.6	9352.9	10450.4	9825.3	9383.0	9683.0	8804.8	9932.1	10516.4	9762.8
	Transfer/Resource ratio	1.0	1.0	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.6
	Objects	135.6	130.9	181.0	171.9	184.2	180.1	180.8	164.0	193.0	168.6	203.0
	Third-parties	25.9	24.2	40.3	35.4	41.0	39.7	40.3	37.5	43.0	34.4	35.2
	Third-party objects	98.0	90.9	157.6	154.8	168.8	157.7	152.0	137.5	172.6	145.0	158.6

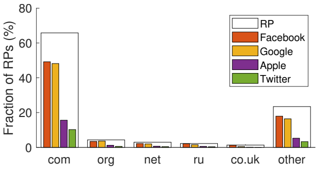


Fig. 11. IDP usage across TLDs.

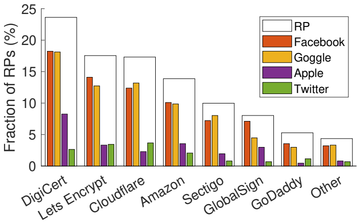


Fig. 12. CA selection of the RPs.

TABLE V
IDP IMPLEMENTATIONS.

	Front Page		Popup or Other Page	
IDP	Redirect	New Window	Redirect	New Window
Facebook	45 (6.5 %)	34 (4.9 %)	314 (45.5 %)	298 (43.1 %)
Google	33 (5.0 %)	28 (4.3 %)	324 (49.4 %)	271 (41.3 %)
Apple	4 (1.9 %)	3 (1.4 %)	128 (60.1 %)	78 (36.6 %)
Twitter	11 (8.1 %)	2 (1.5 %)	74 (54.8 %)	48 (35.6 %)

C. RP implementation characteristics

Table V shows how RPs implement sign-in with the IDPs based on where they are found on the page and if they are opened in a new window or via a redirect. Apple stands out, being much more rarely found on the front page (3.3% of the cases), while the other IDPs are placed on a front page at around 10% of all occurrences. All IDPs are most commonly opened via a popup or other page through a redirect.

Certificate selection: All identified RPs except one (pil.tw) presented valid certificates. Figure 12 shows the issuing CAs of the RPs, including broken down based on the IDPs the corresponding RPs used. We find that the top-7 issuers were responsible for the certificates used by 96% of the RPs, that DigiCert was most popular CA among the RPs (compared to Let's Encrypt among the non-RPs), and that RPs using Apple as IDP are more likely to use DigiCert as CA and less likely to use GoDaddy as CA (compared to the other RPs). These subtle differences are interesting since they may suggest some correlation between using free certificates (Let's Encrypt) vs. paying for their certificates, and the RP's IDP selection.

TLS version and downgrades: All identified RPs except pil.tw supported TLS. Out of these, all had either TLS 1.2 (41.8%) or TLS 1.3 (58.2%) as their highest supported version. While both these versions should offer good security properties, a more concerning threat is downgrade attacks targeting lower version that has not been fully disabled. Also here, RPs using Twitter and Apple stands out, as the set of RPs using Twitter has the largest fraction of RPs (58.3%) supporting deprecated versions, and RPs using Apple the smallest fraction (48.8%) supporting these versions. The corresponding numbers for Facebook and Google are 54.8% and 55.6%.

V. RELATED WORK

Security weaknesses: Almost all RP-IDP relationships today use OAuth. In our manual dataset, OpenID was last seen in 2019, when aol.com stopped using it. While OAuth [12] provides attractive security properties [13], [14] and its security properties have been formally analyzed [15], severe security weaknesses have been found in specific implementations [16], [17]. Several researchers have identified and studied security flaws or attacks against various SSO implementations (e.g., OpenID [18], [19], OpenID Connect [20], Facebook [19], OAuth in mobile apps [21]) or demonstrated flaws in many webpages' handling of authentication cookies [22].

User perception and risks: High use of third-party authentication can increase the risk of users giving their credentials to fake websites [23]. Malandrino et al. [24] proposed a client-side tool that maximizes users' awareness of their information leakage, while Sun et al. [25], [26] studied users' concerns and perceptions when using SSO. To help users make informed choices, Shehab et al. [27] designed a recommender system that bridge users' conceptual (mis)understanding of the risks with SSO [25]. However, users often do not take warnings seriously [28] and seemingly harmless information (e.g., user's music interests) can leak privacy-sensitive information [29].

Characterization: Some researchers have crawled for security vulnerabilities [30], [31] or evaluated the amount of information requested by website using IDPs (e.g., Facebook Connect [32]). However, only a few works have tried to characterize the IDP and app-rights usage [2], [3], [33], [34],

The most closely related work is by Vapen et al. [2]. In fact, the 2012-2015 snapshots used here were collected and shared by Vapen et al. In the work, they identified cross-site information sharing risks and studied differences in the app-rights associated with different classes of websites. However, their study only included two app-rights snapshots from 2014. Here, we use additional snapshots shared by Vapen et al. (from 2015) that we complement with our own snapshots (2019-2021) as well as a large-scale measurement using our Selenium-based crawler. Compared to Vapen et al., our study is both much longer (allowing us to study trends and other changes happening over different time scales) and captures much more RP-IDP relationships and RP properties.

Morkonda et al. [34] built a crawler to study the app-rights usage of the top-500 pages in four countries (when using Facebook, Google, Apple, and LinkedIn as IDP) but never validated their accuracy and had to rely on manual work to identify all IDPs. The early crawler presented by Vapen et al. [33] had much worse recall than ours (e.g., only found 36 out of 186 relationships, giving them a recall of 19%), and did not collect any app-rights agreements. Ours achieve much higher recall (66.8%), precision (99.3%), and automatically extracts the app-rights agreements for the identified relationships.

VI. Conclusions

This paper presents both a longitudinal study of the third-party authentication usage among the most popular websites and a large-scale characterization of the current IDP usage, the app-rights agreements used by today's RPs, and the RPs themselves. The longitudinal study spans a nine-year period, capturing IDP relationships and app-rights agreements both before and after important events such as the Cambridge Analytica scandal (2018) which have put significant pressure on Facebook and other IDPs in how they share data with RPs.

Our findings can be split along several dimensions. We find increasing IDP usage, especially among the most popular domains, who also are the biggest users of IDPs. The usage is also high among certain website categories (e.g., News), Facebook and Google remain the dominating English-speaking IDPs, although Apple quickly has gained usage since its introduction in 2019. Our RP-based analysis shows that RPs often use more third-party services (than non-RPs), are more likely to pay for a X.509 certificates than non-RPs, and that the RPs that use the more privacy-aware IDPs (e.g., Apple and Google) are more likely to also use certificates and TLS connections with more desirable properties.

Finally, our app-rights analysis confirms a positive trend in data sharing practices, with RPs requesting access to less sensitive information. The permission practices of Facebook and Google improved substantially between 2015 and 2019. However, the permission practices can still differ substantially between RPs using Facebook and Twitter. Apple (followed by Google) provides the most restrictive permissions, whereas Twitter is most likely to give out write permissions. While RPs using Google often do not request many permissions, one RP (645voyager.com) requested full access to Google Drive.

References

[1] BBC, "Facebook down: Zuckerberg apologises for six-hour outage," 2021. [Online]. Available: www.bbc.com/news/technology-58800726

[2] A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, "Information sharing and user privacy in the third-party identity management landscape," in *ACM CODASPY*, 2015.

[3] ——, "A look at the third-party identity management landscape," *IEEE Internet Computing*, 2016.

[4] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach," https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election, 2018.

[5] J. C. Wong, "The cambridge analytica scandal changed the world – but it didn't change Facebook," 2019. [Online]. Available: https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook

[6] R. Brandom, "Apple announces new sign-in tool to compete with Facebook and Google," 2019. [Online]. Available: https://www.theverge.com/2019/6/3/18650885/apple-sign-in-sso-tool-data-collection-privacy-ios-13-wwdc-2019

[7] Google Support, "Frequently asked questions about the Google+ shutdown," 2021. [Online]. Available: https://support.google.com/googlecurrents/answer/9217723?hl=en

[8] O. Jarpehult, F. J. Agren, M. Backstrom, L. Hallonqvist, and N. Carlsson, "A longitudinal characterizing of the third-party authentication landscape (extended)," Technical report, 2022.

[9] "Geoip2 databases," https://www.maxmind.com/en/geoip2-databases.

[10] Q. Scheitle *et al.*, "A long way to the top: Significance, structure, and stability of internet top lists," in *Proc. IMC*, 2018.

[11] V. L. Pochat *et al.*, "Tranco: A research-oriented top sites ranking hardened against manipulation," *Proc. NDSS*, 2019.

[12] B. Leiba, "OAuth web authorization protocol," *IEEE Internet Computing*, 2012.

[13] S. Pai, Y. Sharma, S. Kumar, R. M. Pai, and S. Singh, "Formal verification of OAuth 2.0 using Alloy framework," in *Proc. CSNT*, 2011.

[14] S. Chari, C. S. Jutla, and A. Roy, "Universally composable security analysis of OAuth v2. 0." *IACR Cryptol. ePrint Arch.*, 2011.

[15] D. Fett, R. Küsters, and G. Schmitz, "A comprehensive formal security analysis of OAuth 2.0," in *Proc. ACM CCS*, 2016.

[16] S.-T. Sun and K. Beznosov, "The devil is in the (implementation) details: an empirical analysis of oauth sso systems," in *Proc. ACM CCS*, 2012.

[17] E. Shernan, H. Carter, D. Tian, P. Traynor, and K. Butler, "More guidelines than rules: CSRF vulnerabilities from noncompliant OAuth 2.0 implementations," in *Proc. DIMWA*, 2015.

[18] C. Mainka, V. Mladenov, and J. Schwenk, "Do not trust me: Using malicious IdPs for analyzing and attacking single sign-on," in *Proc. IEEE EuroS&P*, 2016.

[19] R. Wang *et al.*, "Signing me onto your accounts through Facebook and Google: a traffic-guided security study of commercially deployed single-sign-on Web services," in *Proc. IEEE S&P*, 2012.

[20] C. Mainka, V. Mladenov, J. Schwenk, and T. Wich, "SoK: single sign-on security—an evaluation of OpenID connect," in *IEEE EuroS&P*, 2017.

[21] E. Y. Chen, Y. Pei, S. Chen, Y. Tian, R. Kotcher, and P. Tague, "OAuth demystified for mobile application developers," in *ACM CCS*, 2014.

[22] K. Drakonakis, S. Ioannidis, and J. Polakis, "The cookie hunter: Automated black-box auditing for web authentication and authorization flaws," in *Proc. ACM CCS*, 2020.

[23] R. Dhamija and L. Dusseault, "The seven flaws of identity management," *IEEE Security & Privacy*, vol. 6, no. 2, pp. 24 – 29, 2008.

[24] D. Malandrino, A. Petta, V. Scarano, L. Serra, R. Spinelli, and B. Krishnamurthy, "Privacy awareness about information leakage: Who knows what about me?" in *Proc. WPES*, 2013.

[25] S.-T. Sun *et al.*, "Investigating users' perspectives of web single sign-on: Conceptual gaps and acceptance model," *ACM Trans. on Internet Techn.*, 2013.

[26] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov, "What makes users refuse web single sign-on? an empirical investigation of OpenID," in *Proc. SOUPS*, 2011.

[27] M. Shehab, S. Marouf, and C. Hudel, "ROAuth: Recommendation based open authorization categories and subject descriptors," in *SOUPS*, 2011.

[28] C. Herley, "So long, and no thanks for the externalities: The rational rejection of security advice by users," in *Proc. NSPW*, 2009.

[29] A. Chaabane, G. Acs, M. A. Kaafar *et al.*, "You are what you like! information leakage through users' interests," in *Proc. NDSS*, 2012.

[30] G. Bai *et al.*, "AUTHSCAN: automatic extraction of Web authentication protocols from implementations," in *Proc. NDSS*, 2013.

[31] Y. Zhou and D. Evans, "SSOScan: Automated testing of Web applications for single sign-on vulnerabilities," in *USENIX Security*, 2014.

[32] S. Egelman, "My profile is my password, verify me! the privacy/convenience tradeoff of facebook connect," in *Proc. CHI*, 2013.

[33] A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, "Third-party identity management usage on the web," in *Proc. PAM*, 2014.

[34] S. G. Morkonda, S. Chiasson, and P. C. van Oorschot, "Empirical analysis and privacy implications in OAuth-based single sign-on systems," in *Proc. WPES*, 2021.