Longitudinal Analysis of the Third-party Authentication Landscape

Anna Vapen, Niklas Carlsson, Nahid Shahmehri
Linköping University, Sweden. firstname.lastname@liu.se

***Abstract*—Many modern websites offer single sign-on (SSO) services, which allow the user to use an existing account with a third-party website such as Facebook to authenticate. When using SSO the user must approve an app-rights agreement that specifies what data related to the user can be shared between the two websites and any actions (e.g., posting comments) that the origin website is allowed to perform on behalf of the user on the third-party provider (e.g., Facebook). Both cross-site data sharing and actions performed on behalf of the user can have significant privacy implications. In this paper we present a longitudinal study of the third-party authentication landscape, its structure, and the protocol usage, data sharing, and actions associated with individual third-party relationships. The study captures the current state, changes in the structure, protocol usage, and information leakage risks.**

I. Introduction

Today, users are frequently asked to create personalized website accounts. In exchange, websites typically offer personalized experiences and services, including customized news feeds, faster online purchases, or the ability to comment on articles, connect with friends, or socialize with other web users. To simplify the registration process many websites offer single sign-on (SSO) services. With SSO, the user can use an existing account with a third-party website, called an *identity provider (IDP)*, to login to the website, which in this case is called a *relying party (RP)*. In addition to providing third-party authentication, these services are also increasingly used for data sharing between the involved websites [21].

The data sharing is typically made possible by the use of the OAuth protocol, which besides SSO also authorizes the RP to act on behalf of the user on the IDP (e.g., by posting to the user's account on the IDP) and to import user data from the IDP to the RP. For example, a user using Facebook as its IDP on a RP website must typically agree on an app-rights agreement that allows the RP to import varying amounts and types of personal information from Facebook, allows the RP to post information on the user's Facebook profile, or both. Similarly, a user using Twitter as their IDP on the same or different RP website may need to allow the RP to post certain information to the user's Twitter stream. These types of cross-site data sharing can have significant privacy implications for users. For example, there have already been many studies that leverage these types of information and various forms of data mining and statistical tools to glean insights into users and their behaviors [4], [11], [13].

To understand cross-site leakage risks and how they are changing, we have performed a longitudinal study of the third-party authentication landscape, its RP-IDP relationships, and the data sharing associated with different relationships and structures. In prior work we presented an initial snapshot of the landscape as it existed in Apr. 2012 [20] and characterized the information shared between websites at that time [21]. In this paper we extend that work with a longitudinal characterization (Apr. 2012 to Apr. 2015) that brings together both structural and information sharing aspects. In particular, we develop a structural model of the landscape and use the model to capture longitudinal trends associated with the participants, individual relationships, and the cross-site data sharing. To capture the evolving privacy risks associated with structural, protocol, and data sharing changes in this third-party landscape, we break our analysis into four parts.

First, over a three-year period, we manually identified the RP-IDP relationships starting from the 200 most popular websites on the web (according to Alexa) and used these relationships to characterize how the structure of the third-party authentication landscape has changed. In this part, we build a conceptually simple model of the third-party landscape and discuss privacy aspects based on the most commonly identified structures and high-level trends. Important observed trends include the tendency for high-degree IDPs with many RPs to get more RPs using them over time. There are also many RPs that have started to use multiple IDPs, increasing the risks for IDP-to-IDP leakage via the RPs. We have also discovered nested structures with sites being both RP and IDP, through which a lot of information can potentially flow.

Second, we performed a protocol-based analysis of each RP-IDP relationship. Our analysis captures general trends (e.g., a shift towards richer data sharing protocols such as OAuth from pure SSO services based on OpenID), identifies how individual sites have added/removed/changed protocols over the measurement period, and investigates potential relations between protocol and IDP changes with changes in website popularity. In general, RPs tend to use combinations of 2-3 very popular OAuth providers. IDPs supporting several protocols are now mainly used with OAuth by RPs. Sites with a stable popularity rating tend to be stable in their IDP usage (i.e., they do not change IDPs), while RPs that remove IDPs tend to become less popular at the same time.

Third, we characterized the cross-site information sharing in this landscape. After classifying the data sharing and privacy

This is the authors' version of the work published at *Proc. NDSS Workshop on Understanding and Enhancing Online Privacy (UEOP@NDSS)*, San Diego, CA, Feb. 2016. It is posted here for your personal use, not for redistribution. The official publication is available here: http://dx.doi.org/10.14722/ueop.2016.23008.

risks of individual RP-IDP app-right agreements presented to the users, we discuss the privacy risks associated with different classes of RPs and some of the major IDPs. In the context of our landscape structure, we analyze the risks on both a single-hop basis and in the context of potential multi-hop information leakage flows. Our results show significant differences in the risks associated with different website classes and how these differences can impact multi-hop information leakages under different sub-structures of the third-party landscape.

Finally, we present results from a targeted login study of the adoption of Facebook's app-rights agreements. When using Facebook with an RP, a user is asked to authorize information sharing between the two parties. In this study, we examine the nature and quantity of this information, both before and after Facebook changed its API specifications. While the API change was intended to encourage RPs to select more privacy preserving data sharing settings, the change lead to very small changes for end user privacy, with RPs still typically importing and posting rich user data.

The remainder of the paper is organized as follows. Section II presents our methodology and datasets. Section III presents a longitudinal analysis of the structural properties of the third-party authentication landscape. Section IV presents a protocol-based analysis, and Section V analyzes longitudinal changes in the app-rights and cross-site information sharing between RPs and IDPs. Related work is discussed in Section VI, before we present our conclusions in Section VII.

II. Methodology

Over a three year period, between April 2012 and April 2015, we have performed a series of manual measurement campaigns to identify and characterize RP-IDP relationships. In total, ten snapshots of the third-party authentication landscape were collected. For each snapshot we first manually identified all RP-IDP relationships that originated from one of the websites on the top-200 most popular websites in the world (according to alexa.com) either at the time of the first snapshot (April 2012) or the current snapshot. The aggregate set of identified relationships allows us to track either the dynamics associated with a fixed set of websites (as exemplified by the "original" top-200 set) or the currently most popular websites (as exemplified by the "current" top-200 set at the time of each measurement campaign).[1]

The use of manual identification allows us to minimize and even eliminate the number of falsely identified relationships. However, since there always is a risk that we might miss some relationships, the set of identified relationships must be interpreted as a lower bound. To sanity check the manual methodology and glean some insight into how tight this bound may be, we also performed crawls with different automated crawlers [20], each with significantly different identification accuracy, and had other student volunteers perform parallel identification campaigns. While the crawl-based identification resulted in many false positives and a non-negligible set of false negatives, the candidate relationships identified by the crawlers allowed us to (after manual validation/rejection) validate that we manually identified all RP-IDP relationships. These tests convinced us that the methodology appears to provide a relatively tight lower bound of the number of RP-IDP relationships associated with the websites examined during a snapshot campaign.

Having selected a set of websites to track, and identified their RP-IDP relationships, we next characterize the websites themselves as well as each of the identified relationships. First, at the time of each measurement campaign, we extracted the Alexa list of the top-million most popular websites (not only the top-200). This allows us to characterize websites based on their popularity dynamics (e.g., based on if they are becoming more or less popular) even if they may only be in the top-200 set during a subset of our snapshots. We also classify each website based on the services they provide, as per the classification suggested by Gill et al. [8]. This classification allows us to investigate differences in the dynamics associated with different classes of websites.

During the initial snapshots, relationships were classified based on the IDP that provided the third-party authentication service. Since August 2013, we have also more carefully classified each relationship based on the protocol and potential information flow associated with each relationship. Protocol usage has been identified by carefully analyzing the third-party communication during an authentication attempt.

For all relationships with one of the top-three English-speaking IDPs (Facebook, Google, and Twitter), we have also recorded information about the app-rights agreements between RPs and users, including (i) the information that the RP will obtain from the IDP, and (ii) the actions that the RP will be allowed to perform through the users' IDP account. For this data collection, we initiated the account creation process associated with each such relationship, and recorded the details on which the user is asked to agree. Finally, to better understand the impact of Facebook changing their app-rights agreement API, we performed complementary recordings of the permissions associated with all Facebook-related relationships, before and after this change.

III. Structural Dynamics

In our three-year analysis of the third-party authentication landscape, we have identified commonly occurring sub-structures and changes in the overall structure. To simplify our discussion, let us use a graph abstraction in which participating websites are represented using nodes, and RP-IDP relationships (with their cross-site information flows) are represented using edges between these nodes. Under this abstraction, three base structures can be considered: (i) IDPs and their relationships, (ii) RPs and their relationships, and (iii) hybrid nodes that simultaneously act as both RP and IDP. These three base cases are illustrated in Figure 1.

Before analyzing each of these three base structures individually, we note that the above graph structure has increased in size over time. For example, between April 2012 and April 2015, there has been an increase from 180 to 213 relationships in the "original" top-200 set, and from 180 to 193 when considering the "current" top-200 set.

[1] For the last four snapshots we also identified relationships for any other websites that had been in the top-200 set at some prior point during our series of snapshots. However, the focus of our analysis is on the dynamics associated with the "original" and the (ever changing) "current" top-200 set, respectively.

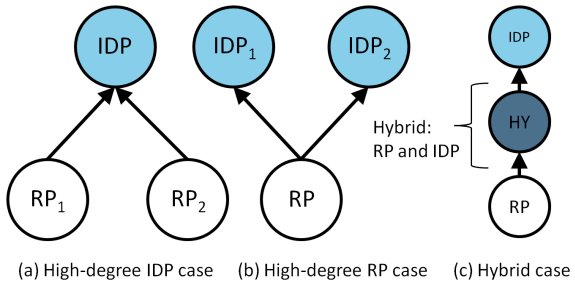

Fig. 1. Base structures in the third-party authentication landscape.

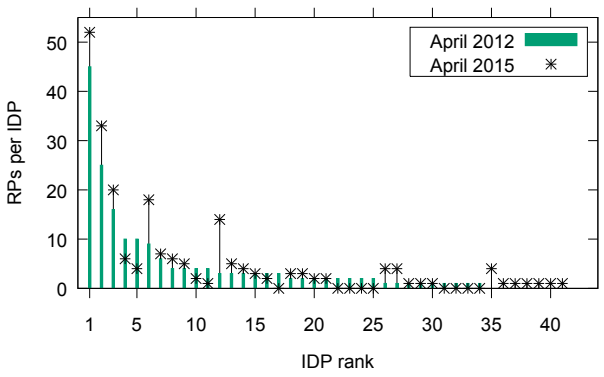

Fig. 2. IDPs (ranked by the number of RPs using them in 2012) and their number of observed RPs per IDP.

A. Popular IDPs

The IDP usage is skewed towards a few very popular IDPs. Already in our original dataset, from April 2012 [20], we observed a few dominant players (e.g., Facebook, Google, Twitter, and QQ) that were responsible for the majority of the RP-IDP relationships, whereas specialized tech IDPs (i.e., sites whose primary service is being an IDP and that may provide stronger or privacy preserving authentication, for example) were uncommon. Since then, the skew has become stronger and many specialized tech IDPs have gone out of business. Out of the 34 unique IDPs observed in April 2012, nine are no longer used (five of these are specialized IDPs, of which four have gone out of business). In the April 2015 snapshot (32 unique IDPs) these have been replaced by seven new IDPs.

The increasing skew towards the top IDPs is illustrated in Figure 2. Here, we show the number of RPs per IDP for the "original" top-200 set, as seen in both April 2012 and April 2015, as a function of the IDPs' original rank in April 2012. Most notably, the top-three IDPs (Facebook, Google, and Twitter, respectively) all show a significant increase in the number of RPs using their services. The other two IDPs with significant increases are QQ (original rank 6) and Weibo (rank 12). Table I separates the usage of these five IDPs further. As seen here, their combined usage has gone up from 54% of the relationships to 64% of the relationships. The results are similar when comparing the "current" top-200 sets.

Thus far we have taken a global perspective. We next break down the IDP usage based on primary language. Figure 3 shows the change in number of RPs per IDP, for IDPs with different numbers of RPs (and languages) in the "original" top-200 set. We observe small but stable usage for Japanese

TABLE I. USAGE OF TOP IDPS IN "ORIGINAL" TOP-200 SET.

Num. relationships with	April 2012	April 2015
Facebook	45	52
Google	25	33
Twitter	16	20
QQ	9	18
Weibo	3	14
Non-top IDPs	82	76
% rels. with top IDPs	54.44%	64.32%
% RPs using top IDP(s)	86.96%	90.48%

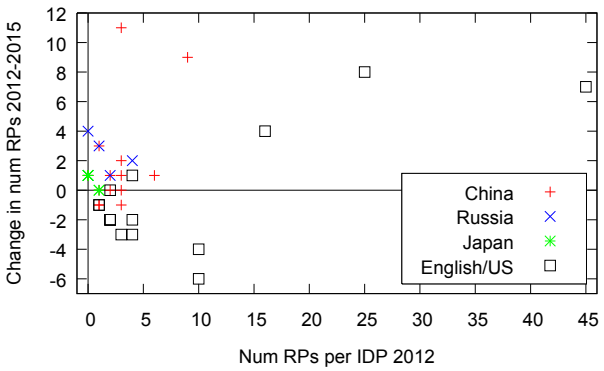

Fig. 3. IDP usage increase/decrease for different language regions.

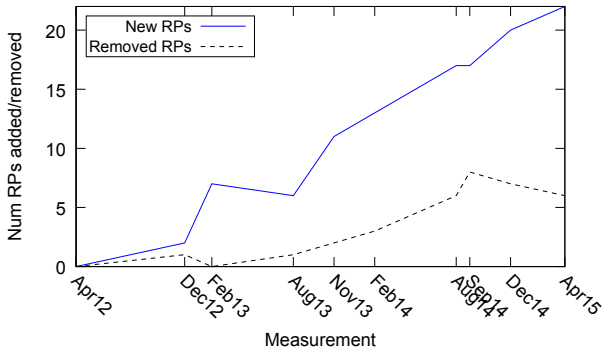

Fig. 4. Number new and removed RPs in each snapshot, considering the websites of the "original" top-200 set and comparing with the RPs in the first April 2012 snapshot.

IDPs (green stars), and significant increases of IDP usage for Russian (blue crosses) and Chinese (red crosses) IDPs. Here, Weibo.com and QQ.com are the biggest winners. For the English-speaking IDPs (black squares) we observe noticeable rich-get-richer effects. For example, only one of the nine IDPs with ten or fewer RPs has seen an increase, whereas all three IDPs (Facebook, Google, and Twitter) with 16 or more RPs in 2012 have seen substantial increases. The IDPs that have seen the biggest drop are Yahoo and Live.

B. RPs and their IDP usage

There has been an increase in the number of RPs. For example, the number of RPs in the "original" top-200 set has increased from 69 to 84 between April 2012 and April 2015. However, when discussing the increase in the number of RPs, it is important to note that there is some churn in the sites that act as RPs. Whereas 22 sites from the "original" top-200 set have become new RPs (by adding IDPs), seven sites (former RPs) have removed all their IDPs. This is illustrated in Figure 4, which shows the number of new RPs and removed

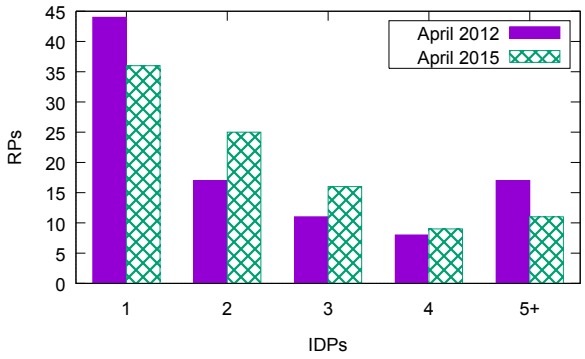

Fig. 5. IDPs per RP, April 2012 and 2015.

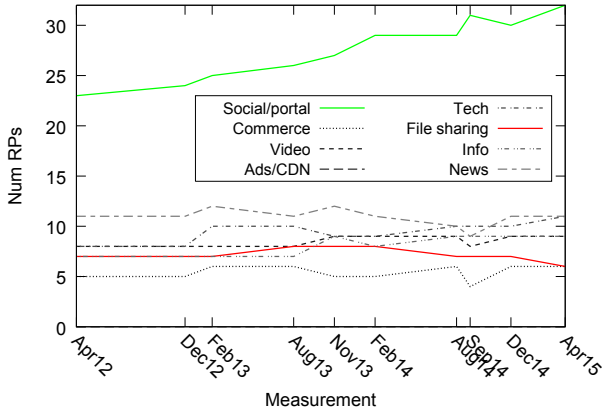

Fig. 6. Number of RPs for different categories in "original" top-200 set.

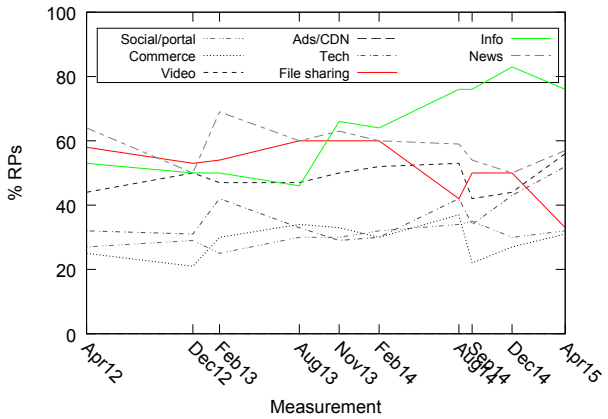

Fig. 7. Percentage of RPs for different categories in "current" top-200 sets.

RPs compared to the original top-200 set as seen in the first snapshot. In addition to these sites, other sites have acted as RPs in during intermediate snapshots. The 22 new RPs (31% increase) and seven removed RPs (10% decrease) over this three-year period can be compared to the 33.5% of sites in the first (April 2012) top-200 set that have been replaced in the "current" top-200 set by April 2015.

Over time, we have seen fewer RPs with many IDPs. The number of RPs with a single IDP has also decreased. Instead, increasingly many RPs use 2-3 IDPs. Figure 5 shows a breakdown of the number of IDPs per RP as observed in April 2012 and April 2015, respectively, for the "original" top-200 set. We note that the single IDP case is still the most common. The majority of these RPs use Facebook as their only IDP. The most common cases with 2-3 IDPs, are RPs that use a combination of Facebook, Twitter and Google, with Twitter-Facebook being the most common pair [21].

When discussing the RPs, it is important to understand which sites decide to be RPs. In prior work [20], we showed that the majority of News (65%), Filesharing (58%), and Info (50%) sites in the "original" top-200 set used IDPs. Since then, Social sites have seen the largest increase in the "original" top-200 set (green curve in Figure 6) and Info sites have seen the greatest increase (today 76% RP coverage) when considering the "current" top-200 set (green curve in Figure 7). Filesharing sites, on the other hand, have seen the largest reduction, across both sample sets (red curves in Figures 6 and 7).

C. Hybrid nodes

We have also observed hybrid (HY) nodes that act as both IDP and RP. In our dataset we have observed three distinct hybrid cases. First, we have observed popular IDPs (e.g., Yahoo and LinkedIn) with many RPs and one or a few IDPs. This example structure is similar to the high-degree IDP, but with an additional connection to one or two high-degree IDP(s). Second, we have observed semi-popular RPs with large sets of IDPs (e.g. Livejournal) that also act as IDPs themselves. This example structure is perhaps best seen as an RP with additional connections to other RPs.

Third, we have observed highly nested structures in which HYs connect with other HYs, forming relationship "chains". Such nested structures have been particularly common on the Chinese web. While the Chinese web have had many HYs across all snapshots, this part of the web has changed from having the most nested structures on the web, to having even more nested structures, and then (recently) back towards a simpler structure with fewer HYs. The Russian web appears to be going through a similar cycle, and is very nested in our most recent measurements. In general, Russian IDPs are often combined in larger groups and used together with Facebook.

D. Structural Model

Motivated by the above observations and the most common sub-structures we have modeled the third-party authentication landscape as a two-layer directed graph structure, with edges indicating the direction of RP-IDP relationships. To capture the distinct role of IDPs and RPs we place IDPs at the top layer and RPs at the bottom layer. We then place the HYs at the layer that results in the smallest number of edges within the same layer. For example, a HY that is IDP more often than it is RP is placed in the upper (IDP) layer, and a HY that is RP more often than it is IDP is placed in the lower (RP) layer. Ties can be broken arbitrarily, but are placed in the IDP-layer for the purpose of our figures. This graph structure can be seen as a bipartite graph, with the minimum possible additional intra-layer HY edges (i.e., edges within a layer) added to this graph. In our figures, edges are directed from RP to IDP, with intra-layer HY edges in red. Furthermore, HYs are blue, RPs white, and IDPs are light blue.

The graph structure of the third-party landscape impacts how information can spread between nodes and the privacy

risks of RP-IDP users. For example, IDPs can export data to RPs, while RPs, if the protocol used allows it, may write, update or delete information on the user's IDP account.

E. Structural Changes

Tying back to the individual base structures, we note that the increasing number of relationships has resulted in more RPs and a much higher skew in the IDPs to which these RPs connect (e.g., Figure 2). In contrast, the degree distribution of the RP nodes has started to even out (e.g., Figure 5).

In general, we have observed that many sites that adopt IDP usage first tend to combine large groups of IDPs (e.g., possibly trying out the new technology or trying to attract users from different social networks), before later becoming more selective in their IDP selection.

Looking more closely at sub-graphs of the third-party authentication landscape, we have also observed differences and similarities between different regions of the web. Motivated by these observations, we conjecture that the Chinese and Russian web are in two different stages of a maturity cycle in which sites adopting third-party authentication in a region first tend to adopt many IDPs, before a second intermediate phase during which many HYs are formed as websites are competing for users, followed by a third phase during which RPs clean up among their IDPs, resulting in a few dominating and highly popular IDPs and far fewer HYs. Furthermore, we conjecture that the Chinese web is in the third phase of this maturity cycle, while the Russian web is in the second phase.

To illustrate how the HYs impact the structure Figures 8 and 9 show the RP-IDP relationships of all Chinese HYs for the April 2012 and April 2015 snapshots of the "original" top-200 set. Figures 10 and 11 show the corresponding relationships of all non-Chinese HYs.

While the Chinese April 2015 snapshot has more intra-layer HY edges compared to in April 2012, especially IDP-IDP edges, the current trend is towards fewer such edges and a simpler, less nested landscape. As an example, the Chinese web had 13 intra-layer HY edges (11 were IDP-IDP) in Feb. 2014, compared to 9 intra-layer HY edges in April 2015.

For the non-Chinese HY edges (Figures 10 and 11) we observe a combination of Russian HYs and a few English-speaking HYs. In general, as noted above, the Russian web is becoming increasingly nested. This is illustrated by the left-most cluster in Figure 11. The English-speaking web is relatively simpler and has not had many hybrid cases in any of our measurements. Furthermore, the two English-speaking HYs seen in April 2012 (Linkedin and Yahoo) are now IDPs. In contrast, AOL is now listed as HY, with several IDPs.

IV. Protocol-based Analysis

A. Protocol Usage

To better understand the RP-IDP relationships and their risks, we next analyze the protocol usage of these relationships. The protocols used in RP-IDP relationships are OpenID and OAuth, with a increasingly heavy skew towards OAuth.

Figure 12 shows the observed usage of these two protocols over time for the "original" top-200 set. In the larger sub-figure the protocol usage of each relationship is classified based on the protocols provided by the IDP in the relationship. With three IDPs in the set (Google, Yahoo and AOL) offering both OpenID and OAuth some relationships are listed as "both". Therefore, starting in August 2013 we have carefully examined and labeled each such relationship based on the authentication process. (See detailed breakdown in the inner sub-figure.) In the case of internal relationships, the IDP owns the RP, and the protocol choice is less relevant. The most common such case is between different Google sites. In relationships with unknown protocol it was not possible to determine the protocol by looking at the messages and APIs.

We note that there is a noticeable increase in OAuth usage, and an even more noticeable decrease in OpenID usage. This becomes even more apparent when looking more closely at the number of RPs using each IDP in April 2012 and April 2015, respectively, when broken down by protocol category. For example, Figure 13 shows the change in number of users for IDPs with different numbers of RPs in the "original" top-200 set. We note that the usage of most OAuth IDPs and Google (supporting both OpenID and OAuth) is increasing. The exception is Live.com, which is owned by Microsoft and used to be a popular IDP in China. We also see that most of the IDPs offering OpenID, as well as Yahoo and AOL (which support both OpenID and OAuth), decrease in usage. The only exception is a Japanese IDP for which a new OpenID relationship was discovered.

The richer information leakage risks of OAuth can result in several privacy challenges [21]. While OpenID was designed for SSO, OAuth also offer authorization, allowing, for example, an RP to act on behalf of a user on an IDP. Although both protocols support data transfer from IDP to RP, support for data transfer is both more developed and more commonly used in OAuth. Transfer of sensitive data between IDP and RP becomes increasingly dangerous to user privacy, if an RP supports several IDPs, since data can be transferred from one IDP to the RP, and then onward to another IDP [21].

B. Relationship and Protocol Churn

As discussed in Section III-B, there is a substantial churn in the set of RPs. Here, we take a closer look at the changes of the individual RP-IDP relationships of these RPs.

For this analysis, we classify each observed RP-IDP relationship associated with the "original" top-200 set along two dimensions: protocol usage and stability. As we only have observed relationships changing from OpenID to OAuth, not between other protocols, we will use four "protocol usage" types: (i) OAuth, (ii) OpenID, (iii) OpenID to OAuth, and (iv) Internal/unknown. We have also observed four general "stability" types: (i) *stable* relationships that the RP keeps over all ten snapshots, (ii) *new* relationships that the RP (or new RP) is adding after the first snapshot and then keeps for the remaining snapshots, (iii) *removed* relationships that were observed in the first snapshot but later removed, and finally (iv) *changing* relationships, in which the RP adds and/or removes an IDP several times over the snapshot sequence.

Table II summarizes the stability for the relationships of each protocol usage type. Due to its highly dynamic nature we have separated out the Chinese web in this analysis.

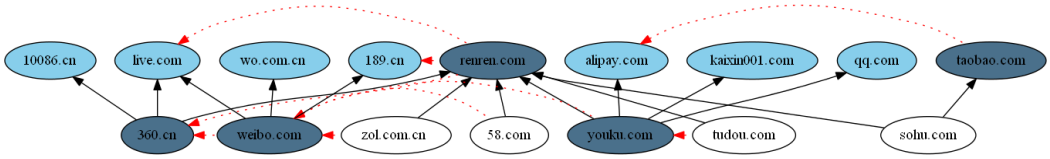

Fig. 8. Chinese HY web, April 2012.

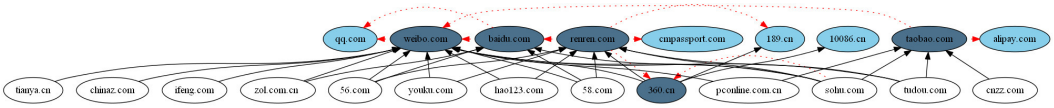

Fig. 9. Chinese HY web, "original" top-200 set, April 2015

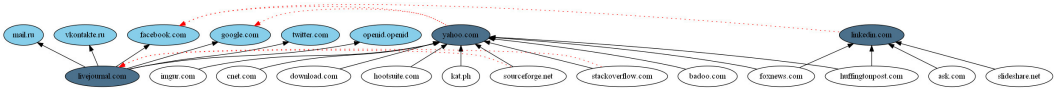

Fig. 10. Non-Chinese HY web, April 2012

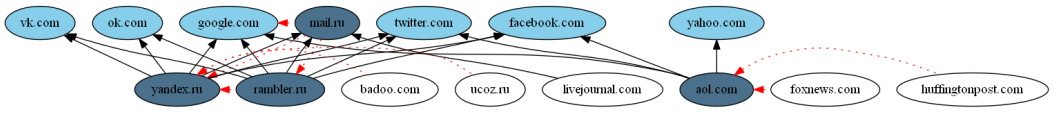

Fig. 11. Non-Chinese HY web, "original" top-200 set, April 2015

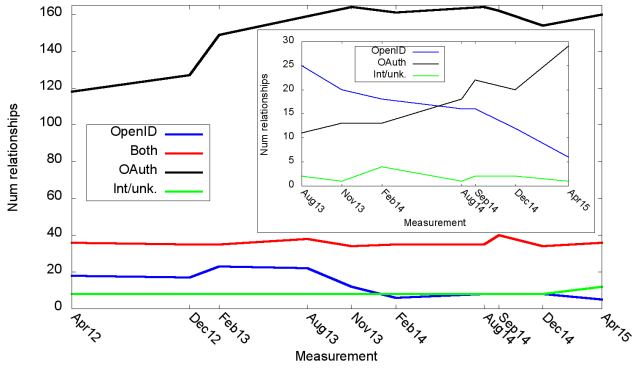

Fig. 12. Protocol usage breakdown, including detailed breakdown of "Both" relationships since Aug. 2013.

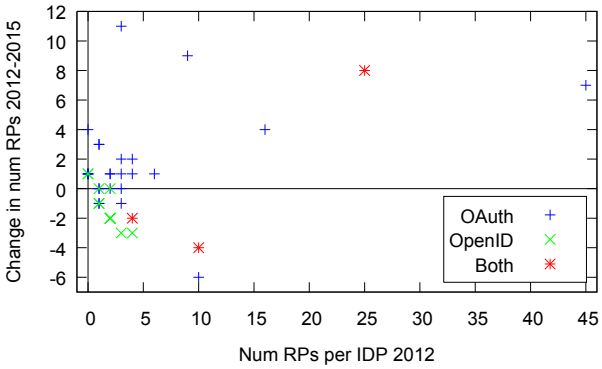

Fig. 13. IDP usage increase/decrease, April 2012 and 2015, protocols.

TABLE II. PROTOCOL USAGE AND RELATIONSHIP STABILITY OVER TIME.

Protocol	Total	Stable	New	Removed	Changed
OAuth	140	46%	33%	10%	11%
OAuth* China	102	25%	28%	15%	31%
OpenID	40	5%	15%	68%	13%
OpenID to OAuth	7	86%	0%	0%	14%
Internal/unknown	14	71%	7%	0%	21%

Consistent with the overall protocol changes in the third-party landscape, the OAuth relationships are considerably more persistent (*stable* + *new*) than the OpenID relationships. Also internal relationships and relationships changing from OpenID to OAuth are considerably more often stable.

Considering all Chinese relationships (primarily OAuth, some internal, but no OpenID), we found (in descending order) 32 *changing*, 26 *stable*, 29 *new*, and 15 *removed* relationships. The combination of high relationship churn, high (although decreasing) nestedness, and many large collaborating social networks makes the Chinese web an interesting component subject for additional analysis.

TABLE III. COEFFICIENT OF VARIATION (CV) OF POPULARITY RANK OF THE RPS IN DIFFERENT CLASSES OF RELATIONSHIPS.

Relationship type	Without China		With China	
	Num. rel.	Avg. CV	Num. rel.	Avg. CV
Internal/unknown	14	0.0496	14	0.0496
Stable OpenID	2	0.0602	2	0.0602
Stable OAuth	64	0.0842	90	0.877
New OpenID	6	0.0932	6	0.0932
Changing OpenID	5	0.0983	5	0.0983
New OAuth	47	0.0995	76	0.256
OpenID to OAuth	7	0.166	7	0.166
Changing OAuth	16	0.470	48	1.40
Removed OpenID	27	0.555	27	0.555
Removed OAuth	14	2.86	29	1.52

We have also investigated if there may be a relationship between RPs' add/removal activity and their relative popularity changes. Table III shows the average coefficient of variation (CV) of the popularity rank of the RPs (as observed over the ten snapshots) for the RPs in different relationship categories. The larger average CV, the larger (normalized) change in popularity rank. Motivated by the impact of the high popularity churn and dynamic nature of the Chinese web, we include sum-

mary results both with and without the relationships associated with the Chinese web included.

Focusing on the non-Chinese web, we observe some interesting trends that suggest that websites that make fewer changes tend to have more stable popularity (smaller CV values), whereas sites that make more changes (e.g., remove IDPs, change from OpenID to OAuth) typically see relatively bigger popularity changes. This can, for example, be seen by comparing the top-four categories in the table (with smallest CV values) with the bottom-four categories in the table (with largest CV values). Here, the "OpenID to OAuth" category contains all relationships (stable, new, added, etc.) for which the protocol has changed, but not the use of the particular IDP. Some of the popularity drops are due to sites changing their domain names, resulting in a popularity drop of the old domain. In cases where a site was removed or disappeared completely, we removed these from the calculations. Of the 94 sites that were RP in at least one of our measurements, the popularity of 65 sites decreased, 3 remained stable, and 26 increased their popularity. Of the sites becoming more popular, the majority used OAuth and added/kept their IDPs.

C. RP Behavior

When analyzing the overall behavior of RPs, across all their relationships, we have grouped 70 non-Chinese plus 24 Chinese RPs into six classes. First, there are 18+0 *RPs with stable IDP selection*, which keep their original IDPs across all snapshots. Of these, 13 RPs used OAuth only connections only, 4 became OAuth only after changing from using Google with OpenID to Google with OAuth, and 1 RP used a mix of IDPs with different protocols. Second, there are 10+5 *new RPs*, which became RPs by adding one or more IDPs. Of these 9+5 RPs used OAuth only, and 1 (non-Chinese) RP used a mix of protocols. Third, there are 9+2 *expanding RPs*, which extended their set of IDPs by adding more IDPs. Again, all but one (non-Chinese) RP used OAuth only. Four, there are 9 (non-Chinese) *RPs owned by their IDP*. Fifth, there are 19+17 *RPs with reduced or fluctuating IDP selection*. Of the 19 non-Chinese RPs in this category, 2 RPs removed multiple IDPs when these IDPs were going out of business or became unpopular, 7 appeared to add/remove bundles of popular IDP combinations, sometimes overlapping with the RPs' previous IDPs, 3 added domain specific (e.g., e-commerce) IDPs for particular audiences, and 7 had fluctuating behavior (both adding and removing IDPs). Of the 17 Chinese RPs in this category, 8 appeared to clean up their IDP selection and favored top IDPs such as QQ and Weibo, while 9 had fluctuating behavior. Finally, there are 5 (non-Chinese) *former RPs* that have removed all their IDPs.

V. Flow-based Analysis

A. Types of Information Flows

When a user logs in to an RP using an IDP, the user must agree to an *app-rights* agreement that outlines which *data* the RP is allowed to import (read) from the IDP, and which *actions* the RP may perform on the IDP (e.g., writing, updating and removing information). In general, OAuth allows richer data import to the RP than OpenID, and also allows actions which OpenID does not. To further categorize the information flow

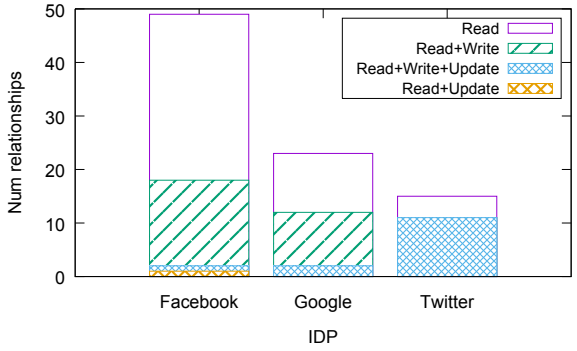

Fig. 14. Flow combinations broken down by top IDP.

risks, we have categorized app-rights agreements based on three main types of information flows:

1) **Read:** Information is transferred from IDP to RP at login. This flow occurs in both OAuth and OpenID, while the other types are for OAuth only. Since many popular IDPs are large social networks, these flows can be a rich information source.
2) **Write:** Information is sent from RP to IDP at any time, even when the user is not logged in at the RP or IDP. This information typically is less rich than in *read* flows, but is often made public (e.g. posted on Twitter) or published to a group of the user's contacts (e.g. posted to Facebook friends of the user).
3) **Update/remove:** Similar to the write flow, the RP can update the information on the IDP (e.g., update the user's profile, join groups, and take other actions on behalf of the user). While these flows can compromise the integrity of user data on IDP level, they do not involve explicit data transfer.

App-rights agreements typically combine different flow types. In prior work, we have classified the agreements based on the information type (e.g., generic vs. private) [21]. Here, we analyze the read/write/update rights agreed upon in these agreements and analyze how this impacts cross-site information sharing both on a single-hop and multi-hop basis.

B. Single-hop Analysis

For the top English-speaking IDPs (Facebook, Google and Twitter), we collected detailed information for the app-rights agreements for the relationships to these IDPs as observed by the English-speaking RPs in the Feb. 2014 snapshot of the "original" top-200 dataset. In total, there are 49 Facebook, 23 Google, and 15 Twitter relationships.

Four flow combinations were observed: (i) *read-only*, which imports (read) data from IDP to RP, (ii) *R+W*, which imports (read) data from IDP to RP combined with posting of information (write) to the IDP via actions, (iii) *R+U*, which imports (read) data from IDP to RP combined with updates/removal (update) of information on the user's IDP account, and (iv) *R+W+U*, which combines all three categories. In this set, we have not observed any case in which the RP does not import (read) data from the IDP.

Figure 14 shows the flow combinations broken down by top IDP. Facebook is using read-only to a large extent, followed by

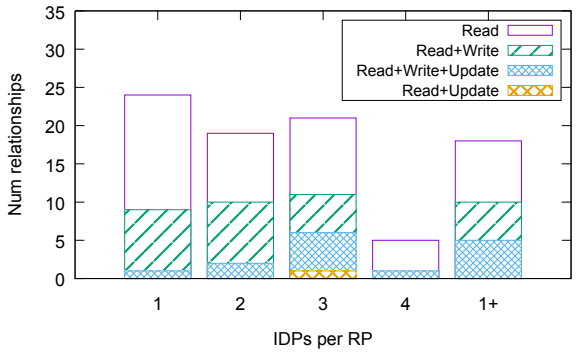

Fig. 15. Flow combinations broken down based on number of IDPs per RP.

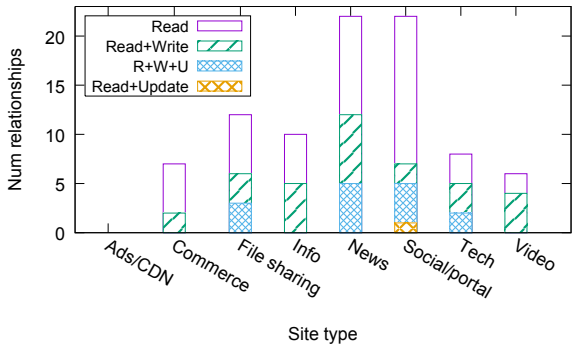

Fig. 16. Flow combinations broken down per RP site category.

R+W, and a few R+U relationships. Since Facebook has a rich set of user data and content created by the user (e.g., images and media content [21]), read-only lets the RPs using Facebook to import large sets of data. Furthermore, a large number of RPs are allowed to post on Facebook (R+W category). For both Google and Twitter the majority of relationships have both read and write rights, and in some cases also update rights. We note that the R+W+U cases (most common with Twitter) are the least privacy preserving.

Figure 15 shows the different flows used in relationships, categorized depending on how many IDPs in total the RP in the relationship has. In general, we see fewer differences here than when comparing across IDPs. This is to be expected, as the IDPs are relatively evenly spread across these three cases, with the exception of Facebook, which dominates the case where an RP only has one IDP. Not surprisingly, this case also has the most read only cases. More interestingly, we observe a significant number of writes (W) in cases with more than one IDP. This is a concern as these cases can result in IDP-to-IDP leakages via the RP. In Section V-C we analyze the risks associated with these cases more closely.

Finally, Figure 16 shows a breakdown of the different risk types for different website categories. Interestingly, News sites use R+W+U the most. A substantial fraction of the Tech, Info, Video, and Filesharing RPs have write (W) rights.

C. Multi-hop Information Leakages

In addition to the *direct* information leakage risks associated with each RP-IDP relationship (e.g., an RP can import/read data from the IDP, write to the IDP, and update/manipulate the user's IDP account) there are also *indirect* information leakage risks that depend on more than one RP-IDP relationship.

TABLE IV. SUMMARY OF RPS THAT ALLOW IDP-TO-IDP MULTI-HOP LEAKAGES.

From \ To	Facebook	Google	Twitter
Facebook	-	$6 \rightarrow 5$	$6 \rightarrow 6$
Google	$4 \rightarrow 4$	-	$7 \rightarrow 7$
Twitter	$4 \rightarrow 3$	$5 \rightarrow 5$	-

Consider for example the three cases in Figure 1. Perhaps the most obvious risk is in the hybrid cases (Figure 1(c)). In this case, the HY may relay information through a series of writes from its RP to its IDP, or from its IDP to its RP through a series of imports (reads). However, with few HYs, this scenario is relatively uncommon. Instead, we will focus on two other cases. First, referring to Figure 1(b), an RP may act as a relay from an IDP_1 to an IDP_2. In this case, the RP must have read/import (R) rights from IDP_1 and write (W) rights to IDP_2 for the same data type. In the second case, referring to Figure 1(a), an IDP may act as a relay from an RP_1 to an RP_2. In this case, RP_1 must have write (W) rights to an IDP and RP_2 must have read/import (R) rights for the same IDP and data type.

To allow a longitudinal analysis, we collected information for the app-rights agreements for the same set of relationships as used in the above single-hop analysis in both Feb. 2014 and Apr. 2015. After restricting ourselves to RPs and relationships that occurred in both snapshots we had a dataset with 49 RPs and 70 relationships (44 Facebook, 14 Google, 12 Twitter). All 70 relationships included read (R) rights. The write (W) rights where a bit more restrictive. Among the 44 Facebook relationships, 15 had write (W) permissions in Feb. 2014 and 11 had such permissions in Apr. 2015. The corresponding numbers for Twitter were 10 for both instances, and for Google the instances decreased from 7 and 6. All Twitter relationships with write privileges also had update/remove (U) privileges. Facebook and Google had only 1 and 2 relationships, respectively, with update/remove (U) privileges.

Using these numbers, the number of cases where RP-to-RP leakage can occur through Facebook has decreased from 645 to 473. For Google, there are 91 potential RP-to-RP leakage cases in both datasets, and for Twitter there are 110 such instances in both datasets. Considering created content, for which the privacy risks are typically higher than for write notifications, for example, the Facebook cases go down to 150 and 66, respectively, the Twitter cases remain the same (at 110), and the Google instances become zero.

Looking more closely at the IDP-to-IDP case, we observe RPs that allow leakages between all three IDPs. Table IV summarizes the possible IDP-to-IDP leakages (via some RP). Here, Feb. 2014 numbers are before the arrow and Apr. 2015 numbers are after the arrow. In general, we observe relatively small changes (with a slight decrease in the number of potential IDP-to-IDP leakages). As is perhaps expected, Facebook and Google are the most common information sources and Twitter is the most common destination in these risk instances.

D. Facebook Use-Case: API Changes and Information Flows

With the popular IDPs offering their individual app-rights APIs, many of the privacy risks in the landscape are controlled by the APIs and how RPs use them. The APIs specify what data and actions can be included in app-rights agreements, how data can be shared, and what actions can be performed.

Being used by the majority (61-65%) of the RPs in our datasets, it can be argued that Facebook is the most influential IDP. In the past two years Facebook has taken some steps towards providing end users with somewhat more control over what is shared with RPs. This has resulted in a series of API changes. Starting at version 1.0, Facebook introduced version 2.0 in April 2014, followed by 2.1 in Aug. 2014, 2.2 in Oct. 2014, and their current 2.3 version in March 2015.

As expected, the big step was from 1.0 to 2.x, which resulted in new restrictions to what RPs could do. In general, less created content can be shared and fewer actions are supported. With the exception of basic information and e-mail information, the data sharing and actions included in the agreements must also undergo a review conducted by Facebook. Users should also be allowed to opt out of all the data sharing and actions that require review. Furthermore, with the exception of users' messages from other users (which is classified as created content) and a partial friend list that only includes the user's friends who are using the same RP, no friend data can be shared.

For this step, RPs were told to change to 2.x no later than May 1, 2015, but were encouraged to make the change earlier. In this section, we take a closer look at the adaption of the new API, and how this impacted the information sharing seen in the final app-rights agreements presented to the end users.

For this analysis, we considered all RPs (across snapshots) that had Facebook as their IDP in Sept. 2014. For these RP-IDP relationships, we then recorded the app-rights four times: Sept. 2014, Dec. 2014, Apr. 2015, and May 2015. The last two snapshots were taken the week before and after Facebook's May 1 deadline. After removing 3 RPs that dropped Facebook as their IDP and 2 RPs for which there were authentication- or account problems we were left with 63 RPs.

In general, most RPs took their time to change to the new API, with many waiting until they were migrated by Facebook, more than a year after the new API was introduced. However, we also saw 3 *early adopters*. One RP was using the 2.0 API already in September, and two RPs started using the new API in December 2014. In April 2015, right before the forced API migration, 33 RPs had changed to the new API. Another 16 RPs had changed to the new API in early May 2015. However, even after the deadline 11 RPs had not changed their API. In the end, Facebook ended up rescheduling the API migration. While we have focused on RPs in the top-200 set it is likely that adoption was even lower for other sets and as we have seen in our crawl-based analysis of the third-party landscape [20] there are also many RPs using Facebook among the long tail of less popular websites. With many individual solutions, it is possible that performing this migration could have become more time consuming than expected.

To analyze potential app-rights changes and changes in privacy risks, we first classified each app-rights agreement into eight different risk classes [21], based on the type of information (e.g., basic, personal, created content, and friends' data) that the RP can import (read) from Facebook and the actions (write and update/remove) that RPs are allowed to perform on Facebook. Interestingly, none of the 63 RPs changed their app-rights (or were forced to change their app-rights) sufficiently for their risk class to change.

Secondly, we considered the number of cases in which the change in API forced a change in the app-rights agreements. Here, we observed 4 *pro-active* RPs that changed their app-rights to comply with the new API before changing their API, and another 22 that changed both their app-rights and API to comply with Facebook's new policy, but for which we cannot distinguish which of the two changes took place first, or if they were implemented simultaneously. On the positive side, the app-rights of 26 RPs (including the first early adopter) were already compliant with the new API restrictions and not much changed for the end user. Among the 11 late adopters, which did not make the API upgrade in time, none complies with the restrictions of the new API. These RPs may have contributed to the delay in the API roll-out.

VI. Related Work

While third-party SSO service was originally seen as a means to identity management and authentication, these services are increasingly used for cross-site data sharing. While the sharing of rich personal information available on many of the popular IDPs is a privacy risk in itself, data mining and other statistical methods that leverage information from multiple sources can result in additional indirect privacy leakage. For example, researchers have demonstrated how public data (such as likes, twitter feeds, etc.) can be used to determine potentially private information [4], [23], [13] and to identify users based on behavior across several websites [9]. Others have demonstrated the privacy risks of old Facebook posts [2], explored the convenience-privacy tradeoff when using Facebook as IDP [7], and demonstrated cross-site leakage in the context of ad services and trackers [11]. Finally, within the context of identity management, Birrell and Schneider [3] present a privacy-driven taxonomy of the design choices.

There are also several known security problems related to both OpenID [17] and OAuth [16]. While the OAuth protocol itself provides attractive security properties [12], [5], severe security weaknesses have been found in specific implementations [16], [15]. The presence of vulnerabilities has prompted researchers to build automated scanning tools that crawl the third-party landscape for security vulnerabilities [1], [24]. Protocol related security problems that enable identity theft and blackmailing [22], and economic factors [6] have also been discussed in the literature. The high use of third-party authentication has also been shown to increase the risk of attackers fooling users into giving their credentials to a fake website or following non-trustful links [6].

Others have shown that users do not trust third-party identity providers, but still use them, without knowing whether they authenticate to a real provider or to a fake provider set up by an attacker [18]. To help users make informed choices when using OAuth, Shehab et al. [14] designed a recommender system.

This may help bridge the users' conceptual (mis)understanding of the risks with SSO [19], although it should be noted that users often do not take warnings seriously [10].

While each of the above works provide interesting insights into the third-party landscape, none of them provide a characterization of the structure, protocol usage, and information flows. In prior work we have provided a preliminary characterization of the structure and protocol usage for the April 2012 snapshot [20], in which the landscape was compared and contrasted with the third-party content delivery landscape. We have also provided an initial characterization of the information leakage in this landscape [21]. In this paper, we present a longitudinal characterization of all these aspects, and provide new insights into the privacy risks in this evolving landscape.

VII. Discussion and Conclusions

We have observed interesting changes in the third-party authentication landscape over the last three years. The usage of popular IDPs using the OAuth protocol has increased, whereas alternatives that provide better privacy have almost disappeared. We show that the landscape can be modeled as a bipartite graph with a few additional edges for hybrid nodes that act as both RP and IDP. These hybrid nodes are relatively rare with the exception of the Chinese and Russian areas of the web, which are more nested than the rest of the web, but can result in significant information leakage risks. In general, the top layer (with IDPs) is getting more skewed (towards the popular IDPs) and the bottom layer is expanding (more RPs). RPs increasingly use 2-3 selected IDPs, although the most common case is for an RP to use Facebook as their only IDP.

We have also modeled and characterized the information flows based on the observed app-rights agreements and the changes seen in these over time. In most cases we observe that the information leakage risks of the app-rights agreements on individual relationships remain fairly stable, even when the IDPs introduce bundling of app-rights or (as in the case of our Facebook use case) force RPs to change the API they use. Instead, most RPs appear to have certain information and actions they think their users may be willing to share. The biggest RP-related risks are associated with news, info, tech, and video sharing sites. These types of RPs often have both read and write rights, which can enable multi-hop leaks. Looking closer at the app-rights we provide insights into how much IDP-to-IDP leakage (via an RP) may result from the set of RPs using Facebook, Google, and Twitter, as well as the RP-to-RP leakage (via one of these IDPs) that is possible. While we evaluate these risks over a fixed set of relationships, we note that the increased skew towards popular IDPs increases these types of risks. Current/future work includes modeling of the multi-hop spread of information (beyond two hops) within the bipartite structure.

Acknowledgements

The authors would like to thank Anirban Mahanti for his contributions to this project and ELLIIT for financial support.

References

[1] G. Bai, J. Lei, G. Meng, S. S. Venkatraman, P. Saxena, J. Sun, Y. Liu, and J. S. Dong, "AUTHSCAN: automatic extraction of Web authentication protocols from implementations," in *Proc. NDSS*, 2013.

[2] L. Bauer, L. Cranor, S. Komanduri, M. Mazurek, M. Reiter, M. Sleeper, and B. Ur, "The post anachronism: The temporal dimension of Facebook privacy categories and subject descriptors," in *Proc. WPES*, 2013.

[3] E. Birrell and F. B. Schneider, "Federated identity management systems: A privacy-based characterization," *IEEE Security and Privacy*, vol. 11, no. 5, pp. 36–48, 2013.

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

[5] S. Chari, C. Jutla, and A. Roy, "Universally composable security analysis of OAuth v2.0," in *IACR Cryptology ePrint Archive*, 2011.

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

[7] S. Egelman, "My profile is my password, verify me!: the privacy/convenience tradeoff of Facebook connect," in *Proc. SIGCHI*, 2013.

[8] P. Gill, M. Arlitt, N. Carlsson, A. Mahanti, and C. Williamson, "Characterizing organizational use of web-based services: Methodology, challenges, observations, and insights," *ACM Transations on the Web*, vol. 5, Oct. 2011.

[9] O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira, "Exploiting innocuous activity for correlating users across sites," in *Proc. WWW*, 2013.

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

[11] D. Malandrino, A. Petta, V. Scarano, L. Serra, R. Spinelli, and B. Krishnamurthy, "Privacy awareness about information leakage," in *Proc. WPES*, 2013.

[12] 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.

[13] M. Pennacchiotti and A.-M. Popescu, "Democrats, republicans and Starbucks afficionados: user classification in Twitter," in *Proc. ACM SIGKDD*, 2011.

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

[15] 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.

[16] S.-T. Sun and K. Beznosov, "The devil is in the (implementation) details," in *Proc. ACM CCS*, 2012.

[17] S.-T. Sun, K. Hawkey, and K. Beznosov, "Systematically breaking and fixing OpenID security: Formal analysis, semi-automated empirical evaluation, and practical countermeasures," *Computers and Security*, vol. 31, no. 4, pp. 465–483, Jun. 2012.

[18] S.-T. Sun, E. Pospisil, and K. Beznosov, "What makes users refuse Web single sign-on? an empirical investigation of OpenID categories and subject descriptors," in *Proc. SOUPS*, 2011.

[19] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov, "Investigating users' perspectives of Web single sign-on: Conceptual gaps and acceptance model," *ACM Trans. Internet Technol.*, vol. 13, no. 1, pp. 2:1–2:35, 2013.

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

[21] ——, "Information sharing and user privacy in the third-party identity management landscape," in *Proc. IFIP SEC*, 2015.

[22] R. Wang, S. Chen, and X. Wang, "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 Symposium on S&P*, 2012.

[23] R. Zafarani and H. Liu, "Connecting users across social media sites: A behavioral-modeling approach," in *Proc. ACM SIGKDD*, 2013.

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