Protecting End User’s Privacy When using Social Login through GDPR Compliance

Carlos Villarán and Marta Beltrán

Department of Computing, ETSII, Universidad Rey Juan Carlos, Madrid, Spain

Keywords: GDPR, Identity and Access Management, OpenID Connect, Privacy, Social Login.

Abstract: Social login allows end-users to identify and authenticate in different applications and services using their social network providers (Facebook, Twitter, Google, LinkedIn) instead of using specific accounts and passwords. This kind of single-sign-on approach relies on federated identity management specifications that significantly simplify login processes. However, this kind of solution also implies new threats for end user’s privacy, because identity providers (social network providers) have access to sensitive information that allows them to perform processing without explicit consent (to profile or track their users, for example) or that can be shared with third parties. This paper proposes the inclusion of new capabilities within the authentication flows, intending to mitigate these privacy threats guaranteeing compliance with the General Data Protection Regulation (GDPR) through transparency and efficient use of already existing mechanisms and technologies such as back-channel logout or consent receipts. Furthermore, the integration of these capabilities in OpenID Connect flows has been validated with a real prototype of the proposed solution.

1 INTRODUCTION

Social network providers have become identity providers. Federate identity management specifications such as SAML (OASIS Security Services (SAML) Technical Committe, 2005), OAuth (Internet Engineering Task Force (IETF), 2012) or OpenID Connect (OpenID Foundation, 2014) enables single-sign-on procedures. Once an end-user has logged at the social network, this account can be used to authenticate to third-party applications and services smoothly. This approach reduces the number of accounts and passwords an end-user has to deal with, improving at the same time usability and security.

Although identification and authentication procedures are transparent, and users’ experience is comfortable (it consists of a click in a button), this kind of social login may have a significant impact on privacy (Scott et al., 2016). End-users provide sensitive information to identity providers (static personal data, dynamic contextual data), and these providers may share this information with other providers or third parties (without users’ knowledge or consent), may leak it involuntarily or may use it to profile or track them with commercial objectives.

The protection of end-users’ privacy in these scenarios and the compliance with current regulations are still an open problem because aforementioned federated specifications do not include specific capabilities or features regarding data protection.

The main contributions of this work are: 1) A novel approach for protecting end-users’ privacy when using federated mechanisms to solve identity management. The proposed solution enables compliance with the GDPR with independence from the identity provider and the application or service provider. 2) The identification of the specific capabilities that must be added to social login implementations to guarantee each of the rights set out in this regulatory framework. 3) The adoption of the proposed capabilities through the use of an unified and standard web portal and well-known and widely adopted technologies and mechanisms, easily integrable with current federated specifications used for social login. Mainly, back-channel logouts and consent receipts.

The rest of this paper is organized as follows. Section II provides an overview of the background on underlying concepts and the related work. Section III discusses the primary motivations for this work with some examples and potential use cases. Section IV describes the proposed approach to add privacy capabilities to federated identity management specification capable of guaranteeing GDPR compliance. Section V details the proposed solution implementation,
validation and evaluation. Finally, Section VI summarizes our main conclusions and the most interesting lines for future research.

2 BACKGROUND AND RELATED WORK

2.1 On Federated Identity Management

Federated identity and access management involve an end-user who needs to be authenticated or authorized to access an application or service. Instead of forcing this application or service to solve authentication/authorization by itself, this last agent and the end-user rely on an external server, delegating the primary responsibilities to this third-party provider. Identity providers can work within trust federations, summarizing claims about the authenticated user and her privileges in a standard format, a capacity or token.

The most adopted federated specifications are SAML for authentication, OAuth for authorization and OpenID Connect/Mobile Connect for authentication and authorization (with only one flow, this is why it is the basis of almost all social login solutions).

Using the most complex flow as an example (figure 1), the proposed by these last specifications, when an end-user needs to access an application or service (step 1), this agent, the Relying party or RP, redirects her browser to the Identity Provider or IdP (step 2). This provider is responsible for authenticating the end-user or obtaining her consent to perform the requested interaction if she has logged in already (steps 3 and 4).

After performing the authentication, the end user’s browser is redirected back to the application or service with and Authorization Code (steps 5 and 6). This code is used to request a token (step 7 and 8) that claims about the authenticated end-user (the ID token) and an additional token that can be used to determine the scope of the user’s privileges and to obtain more information about her.

2.2 On Social Login and Privacy

In these federated specifications, social network providers, technology companies, and mobile network operators have found a mechanism to provide a new service to their users: single sign-on for applications and services. For example, Facebook, Twitter, Google or Apple allow their users to sign in to a third-party provider instead of creating a new account specifically for that application or service. As mentioned before, this simplifies registrations and logins for end-users and improves their quality of experience. For providers, this kind of authentication and authorization also has significant benefits; they avoid bothering users creating new accounts, losing fewer users during the registration or account resetting processes, which increases the number of visitors.

On the other hand, identity providers offer this social login service, very often free, because users’ data is precious. The collected information allows providers to improve their content and services (through personalization), their ability to place targeted advertisements, etc. Therefore, not keeping any information about the authentication activities of their users (an easy way to become GDPR-compliant) is not an option.

Different researches have emphasized in the privacy threats that this social login poses to end-users. Some of these works apply different threat modelling methodologies and techniques (Robles-González et al., 2020) to conclude that the main privacy threats observed when using social login are personal data leakage, lack of control over personal data at the identity provider, lack of transparency in the sharing of personal data, end-user profiling or end-user location tracking.

Previous works try to mitigate all these threats proposing specific countermeasures or mitigations. The first threat can be mitigated or even avoided improving the security of the specifications and its implementations at the identity provider infrastructure. Different works have focused on this research area such as (Mainka et al., 2017), (Fett et al., 2017), or (Li and Mitchell, 2020).

Other works such as (Bodnar et al., 2016), (Vilarreal et al., 2017) or (Navas and Beltrán, 2019), propose solutions for the second and third threats, trying to minimize the personal data required when registering at an identity provider and improving the transparency of information-sharing processes as well as the control provided to end-users. These works rely on concepts such as privacy scopes to be agreed by end-users when sharing their personal data with third parties privacy tokens to exchange privacy preferences or reputation schemes designed to quantify the degree of privacy protection achieved by the different providers. Even there are works proposing experiments to explore end users’ privacy awareness when using login methods or developing tutorials and awareness campaigns to inform them on the pros and cons ((Moey et al., 2016), (Farzaneh Karegar and Fischer-Hübner, 2018)) or helping the users to scan vulnerabilities when they use OpenID via web plugin (Li et al., 2019).
On the other hand, (Fett et al., 2015), (Isaakidis et al., 2016) and (Hammann et al., 2020) focus on guaranteeing unlikability in order to avoid user profiling or tracking (threats fourth and fifth in the provided list).

It has to be pointed that the last group of works ((Asghar et al., 2016), (Halpin, 2017)) proposes a complete change in the concept of federation, letting identity providers issue credentials and tokens that end-users store locally. These users use these credentials and tokens when they need to access applications, and services without any interaction with identity providers. Therefore, keeping control over their personal data.

3 MOTIVATION AND USE CASES

Different research results show how their sensitive information's privacy is an important issue for end-users. However, most of them rarely make an active effort to protect this information; on the contrary, they tend to give it away voluntarily if they can get a benefit in return. This is known as the privacy paradox, and it is one of the main motivations of this research because it has been rarely considered in previous works. Some providers may have already created the infrastructure needed to be compliant with GDPR and to provide their users with tools to exercise their rights. But these tools are rarely known by users, understandable, usable, complete, interoperable or well integrated with the technology that supports the authentication processes.

We think that the capabilities necessary for GDPR compliance must be built into the IdP, from design, following standard frameworks (to avoid end-users to handle one new solution for each specific provider), and provided in its default implementations, not just when users request them. Other limitations identified in these previous works can be summarized in:

- Ease of adoption and integration. Many of the mentioned works propose solutions that involve significant changes to current specifications or their implementations. This makes their adoption very difficult, so they are not usually deployed in production environments. Another important adoption barrier is that in many cases, they are based on the development of custom solutions and components rather than standard technologies or mechanisms widely used by development teams that have to integrate identity management into their products.
- Compliance with regulation. Previous works usually propose improvements that increase the levels of privacy provided by social login mechanisms but do not guarantee that these comply with current regulations such as the GDPR in the case of Europe. This makes its adoption very complicated in many contexts in which this type of data protection regulations has been obligatory for a couple of years now.
- Regarding the compliance with the regulation, the end-user’s consent to carry out any processing activity is essential. However, most of the previous works in this research area do not address this consent’s management at all.

The main goal of this research is to overcome all these limitations. Current identity providers are often failing to fulfil explicit end-user requests (without undue delay) to guarantee their rights under the GDPR because of the already mentioned limitations of specifications and their implementations. This failure may cause a complaint to supervisory authorities and a result of a large fine or reputation loss.

The most important use cases that may benefit from the proposed solution, are those exemplified in the following scenarios:
Scenario 1. An end-user wants to Facebook connect as her primary login alternative at different applications. This end-user is concerned about her privacy, and she would like to be informed: What types of personal data is Facebook processing? Which is the lawful basis for this processing? How long will this provider be storing this data? Will Facebook share this data with other organizations? What visibility will have the user about this data sharing? Furthermore, the user wants the right to access (to obtain a copy) and to rectify (to correct inaccuracies) this data.

Scenario 2. A European end-user wants to use Sign In with Apple to perform social login at different applications and services. Nevertheless, she wants to object to direct marketing or restrict it (or other types of processing at the identity provider or a specific relying party), take her data to another identity provider (including historical data about information shared with RPs), or erase her data at any moment.

Scenario 3. A SaaS company relying on GSuite would like to simplify employees’ authentication using Google Sign-In to perform social login at all the corporate apps and services (collaboration, marketing, development, etc.). However, this company requires precise control over shared information and detailed traceability of information shared with RPs. For example, to perform an audit or to demand accountability in case of a data breach.

4 A MODEL FOR GDPR COMPLIANCE WITHIN SOCIAL LOGIN SOLUTIONS

4.1 GDPR Roles and Rights

The GDPR (Parliament and Council, 2016) is the current data protection regulation in Europe, designed to guarantee greater privacy protection and rights to European citizens and to harmonise data protection laws across all European countries. It is currently considered the world’s strongest data protection regulatory framework.

This regulation distinguishes five different roles regarding data protection. The first is the Data Subject, the natural person who is the owner of the personal data which identifies him or her or enables him or her to be identified. Within a social login solution, the Data Subject is an European end-user.

The second is the Data Controller, the natural person or legal entity that decides the purposes and mechanisms of the processing of personal data. The controller is accountable for GDPR compliance. The third is the Data Processor, a third party (natural person or legal entity) that processes personal data on behalf of the Data Controller. The processor must fulfil the conditions specified by the controller in a signed Data Processing Agreement (the mechanism used to guarantee that obligations stated in GDPR are complied with).

Within social login solutions, the Data Controllers are the IdP and RP and the Data Processor are also the IdP and the RP or even a third party not included in the authentication flow.

The fourth is the Supervisory Authority, a public organization in an European country responsible for monitoring compliance with GDPR: advising companies, auditing their compliance, managing complaints from Data Subjects, etc. Since identity providers operate in multiple European countries, these companies usually choose to appoint a Lead Supervisory Authority for the purpose of reporting. In this way, they simplify compliance management, reporting, etc.

The last one is the Data Protection Officer (DPO), that is not always present (it is not mandatory for certain companies). This officer is the person responsible for ensuring compliance with GDPR, advising company management and staff about the best strategy to follow and it implementation. Identity providers must appoint a DPO because they process personal data of European citizens at a large scale.

All end-users have GDPR rights that the IdP and RP have to guarantee, but currently, the Data Subject has to request the Data Controller this fulfilment explicitly. The Data Subject rights within the GDPR are the right to be informed, right of access, right to object, right to rectification, right to erasure (right to be forgotten), right to restriction of processing, right to data portability and right not to be subject to a decision based solely on automated processing. The last one does not apply in social login scenarios.

The right to be informed (articles 13 and 14 and recitals from 60 to 62) allows end-users to know who is the IdP or RP and how to contact them (their DPO); the purposes and interest of the performed data processing; recipients and international personal data transfers and the retention period.

The right of access (article 15 and recitals 63 and 64) allows end-users to request the IdP and RP an intelligible copy of the personal data they have collected about them and other supplementary information. Without undue delays and without being charged any fee.

The right to object (article 21 and recitals 69 and 70) enables end-users to oppose data processing when
there are no legitimate grounds. The right to rectification (article 16 and recital 65) allows end-users to request a personal data correction to the IdP or RP when this personal data is incomplete or inaccurate.

The right to erasure, also known as the right to be forgotten, (article 17 and recitals 65 and 66) can be exercised when the end-users’ personal data is no longer required at a provider, when the end-users withdraw their consent when data processing is illicit or due to legal compliance and after exercising the right to object when there are no other legitimate grounds.

There is an alternative to requesting the erasure of personal data: restrict the processing in certain circumstances (article 18). This right can be exercised to limit the way that the IdP or the RP use data if the end-user has a particular reason for asking for this restriction. The restriction is usually in place temporarily.

Finally, end-users have the right to data portability (article 20). Therefore, to obtain data that the IdP or RP store and reuse it for their own purposes, for example, move it to a new provider.

4.2 Privacy Capabilities Required for GDPR Compliance

Taking into account the rights discussed in the previous section, the capabilities that the IdP must incorporate to ensure that they are fulfilled can be summarised in three:

1. Centralised information repository where the end-users can see their personal data and other information related to them and can exercise their rights. This repository has to be user-friendly, up-to-date and concise in order to help users to understand their rights and assist them in the process of exercising them. Furthermore, the IdP and RP can easily comply with their obligations of facilitating the end-users’ right exercise (recital 59), serving information about the end-user data processing (recital 61) and providing secure remote access to end-users’ personal data (recital 63). The IdP and RP have, in a centralized manner, all the necessary information to inform their users or to quickly transfer the information to other IdPs or RP if needed.

2. Ability to revoke tokens, forcing the logout of an RP from an IdP, without needing end-user intervention.

3. Use of consent receipts, standard documents that summarize what personal data has been given, to whom (IdP, RP or third party), and for which purposes. These receipts must be stored in the centralized information repository, written in a machine-readable format and available for downloading.

4.3 Improvements in Social Login Implementations

This work proposes a similar solution for the right of access, showing all the needed information for this right in the unified web. All the information for both rights also gives the chance to understand better how the IdP and RP manage the end-user personal data. As explained in the information right, after the end-user clicks on the right of access, he or she is informed about this right, pointing that all the information is on the main screen of the unified web.

This work proposes that once the end-user has selected the right to object on the main screen,
its explanation is shown. In the next screen, the end-user can select the data processing he wants to object. Some of these data processing could not be chosen if legitimate grounds exist that overrides end-users right, but the web has to clearly explain those grounds. Once the end-user has selected the data processing he wants to object; the IdP has to manually check the request or use an automatic process. After that, the IdP (or RP) has to change the scopes of the ID token. First, this work proposes to perform a back-channel logout (Foundation, 2020) with those RPs that have in place the data processing the end-users are objecting to. As it has been mentioned before, this capability forces the logout of an RP from an IdP, without needing end-user intervention. Before starting this flow, the RPs must be registered in the IdP, including their back-channel logout URI (step 0 of figure 2). In the first step of the back-channel logout flow, the IdP sends an HTTP POST to the RPs back-channel logout URI with a logout JSON Web token or JWT (step 1 of figure 2). This logout token includes the logout event ("events": "http://schemas.openid.net/event/backchannel-logout": ) and a reference to the affected end-user. This reference can directly identify the user, including the subfield, the end-user session (sid field) or both. The RPs, after validating the logout JWT has to perform the end-user logout (step 2 of figure 2), taking into account the recommendations of (Foundation, 2020). Each RP responds to the IdP (step 3 of figure 2) with an HTTP 200 OK if the process was successful or with an HTTP 400 Bad request otherwise. Besides, the RP can respond an HTTP 501 Not implemented if the logout process fails. A similar approach should be used with the right to restrict processing.

Regarding the right to rectification, within the social login scenario, the personal data is stored in the IdP, so the data correction must be done in the IdP. This work proposes that the unified web lets the end-user directly correct them in the IdP. After selecting this right on the main screen of the unified web, the end-user can read an explanation of this right. It should include that this process could force a logout in those RPs that uses the personal data that need to be corrected (to avoid these RPs using an obsolete version of this data). After that, the end-user can correct his data. Once he has finished and moved to the next screen, the IdP can validate if this personal data is correct. This validation can be done with an automated process or manually. An example of an automated process can be entering a code sent via email or SMS to check that the email or mobile phone, respectively, is correct. A manual process can be performed by an agent calling the end-user to check the correction. After the validation screen, the user returns to the main screen. If the modified personal data is also stored in JWT, the IdP could wait for the JWT expiration and correct the personal data in the reissuing process. If the expiration time is too long, the IdP has to force a back-channel logout following the same flow as in the right to object. The end-users can see on the main screen when the personal data has been updated.

This work proposes that after selecting the right to erase or the right to be forgotten, the end-user can read an explanation of the right. This explanation has to include information about how much time is needed to complete the process, including backups of personal data. In the next screen, the end-user can select the personal data he wants to delete. Some personal data could be not selected when the end-user does not have the right to erase that personal data, but an explanation of the grounds needs to appear. After that, the end-user return to the main screen and the IdP can process the request with undue delay using an automated process or manually. Once it is confirmed that the end-user has the right to delete the personal data, the IdP performs a back-channel logout in those RPs that uses the personal data. After that, the IdP can delete the personal data following their own procedure. Also, it has to remember to erase the data from backups if feasible because backups can contain this type of information. The end-users can see the deletion of their personal data on the main screen of the unified web.

With the right to portability, the end-users can obtain a structured and processable file with their personal data. This work proposes that after selecting this right on the main screen, the end-users can read an explanation of this right. After that, the end-users can select between transferring the information to other integrated entities with the IdP or downloading a file, in a standard format, with the content. Finally the end-users return to the main screen.

5 IMPLEMENTATION AND VALIDATION

5.1 Prototype

The prototype implementation is based on HTML, JavaScript and PHP in a XAMPP server over Windows 10. It also uses JSON as structured format for data (compatible with current social login specifications and implementations). It has been implemented within an OpenID Connect IdP, the basis of all current
The unified web portal shows the information in tabs. The first tab includes the end-users’ personal data, consent receipts, retention period and source. The second tab shows performed data processing and grounds. The third tab incorporates recipients and international transfers. The fourth tab includes IdP and DPO information. The last tab contains a log of the exercised rights, including the affected personal data or data processing, when was the request, and when was the resolution. At the bottom of this web, there is a form where the user can make a new request.

5.2 Validation and Discussion

From a functionality perspective, the proposed solution provides IdPs mechanisms to fully comply with the GDPR. This is due to the following reasons:

- Guarantees lawfulness and fairness: The proposed unified web portal allows end-users to exercise their applicable data protection rights in the same standardised way in all their IdPs. The three scenarios introduced in section 3 can be solved with GDPR compliance thanks to the information available within the portal, the permission to edit (to update or to erase, including backups) or download data (to access or to move to another provider) and the capacity of objecting to or restricting data processing (including the ability to revoke alive tokens already provided to RPs).

- Facilitates security, transparency and accountability: If the end-user authentication performed by the IdP is assumed to be secure, the proposed mechanisms are secure, because only properly authenticated and authorised users are allowed to access the unified web portal. The additional information included in this portal, such as explaining the users rights or the DPO contact data improves transparency. Finally, the logging of exercised rights and the recording of consent receipts increase accountability.

The prototype demonstrates that the proposed solution constitutes a valuable tool to allow individuals to exercise their rights in social login scenarios without requiring changes in current federated identity management specifications. But this kind of user-centric privacy-enhancing technology requires further research in two aspects. The first is to find incentives to IdPs to provide this kind of tool. The most essential incentive should be to foster the competition of digital services, but current providers do not perceive that they need to offer this kind of control to their users to win this competition. It is much more likely that their motivation, at present, is related to avoiding the fines mentioned in section 3. The second is to find incentives to end-users to reclaim this control, understand the value of their data, and be able to use a tool such as the one proposed in this paper to exercise their rights and enjoy the mentioned value.

6 CONCLUSION

GDPR is a legal document that provides no technical guidance to the entities that have to comply with it. This is one reason why current social login solutions (provided by non-European companies in most cases) are not aligned with this data protection regulation, although many of the users of these services are European citizens.

This paper has introduced three main privacy capabilities, easy to implement and to integrate with current authentication flows, that can be used by the IdP of these social login solutions to guarantee GDPR-compliance. The first, a simple, unambiguous, intuitive and tailored to end-user needs unified and standard web portal containing all the information regarding applicable data protection rights (right to be informed, right of access, right to rectification, right to erasure, right to restrict processing, right to data portability and right to object). The second, the ability to revoke alive tokens when the exercise of these rights makes it necessary. And the third, the use of standard consent receipts to keep a user-friendly record of end-user consents for data collection and processing.

A prototype of a solution based on the proposed
mechanisms has been implemented and used to validate their functionality, demonstrating how an IdP using OpenID Connect can easily guarantee GDPR compliance (no change in the current specifications is required and standard technologies and mechanisms can be used), following the principles of lawfulness, fairness, security, transparency and accountability.

REFERENCES


Scott, C., Wyne, D., and Boonthum-Denecke, C. (2016). Examining the privacy of login credentials using web-based single sign-on—are we giving up security and privacy for convenience. In 2016 Cybersecurity Symposium (CYBERSEC), pages 74–79. IEEE.