Privacy Token: A Mechanism for User's Privacy Specification in Identity Management Systems for the Cloud

María Elena Villarreal[*], Sergio Roberto Villarreal[†], Carla Merkle Westphall[‡], Jorge Werner[§]

Network and Management Laboratory
Post-Graduate Program in Computer Science
Federal University of Santa Catarina
Florianopolis, SC, Brazil
Email: `maria.villarreal@posgrad.ufsc.br`[*], `sergio@lrg.ufsc.br`[†],
`carla.merkle.westphall@ufsc.br`[‡], `jorge@lrg.ufsc.br`[§]

Abstract— **With the increasing amount of personal data stored and processed in the cloud, economic and social incentives to collect and aggregate such data have emerged. Therefore, secondary use of data, including sharing with third parties, has become a common practice among service providers and may lead to privacy breaches and cause damage to users since it involves using information in a non-consensual and possibly unwanted manner. Despite numerous works regarding privacy in cloud environments, users are still unable to control how their personal information can be used, by whom and for which purposes. This paper presents a mechanism for identity management systems that instructs users about the possible uses of their personal data by service providers, allows them to set their privacy preferences and sends these preferences to the service provider along with their identification data in a standardized, machine-readable structure, called privacy token. This approach is based on a three-dimensional classification of the possible secondary uses of data, four predefined privacy profiles and a customizable one, and a secure token for transmitting the privacy preferences. The correct operation of the mechanism was verified through a prototype, which was developed in Java in order to be incorporated, in future work, to an implementation of the OpenId Connect protocol. The main contribution of this paper is the privacy token, which inverts the current scenario where users are forced to accept the policies defined by service providers by allowing the former to express their privacy preferences and requesting the latter to align their actions or ask for specific permissions.**

Keywords–Privacy; Cloud Computing; Identity Management.

I. INTRODUCTION

Cloud Computing offers infrastructure, development platform and applications as a service, on demand and charged according to usage. On the one hand, this paradigm gives users greater flexibility, performance and scalability without the need to maintain and manage their own IT infrastructure. On the other hand, it aggravates the problem of application and verification of security and causes users to lose, at least partially, control over their data and applications [1].

With the increasing amount of personal data stored and processed in the cloud, including users' Personally Identifiable Information (PII), economic and social incentives to collect and aggregate such data have emerged. Consequently, secondary use of data, including sharing with third parties, has become a common practice among Service Providers (SPs) [2]. However, since users only interact directly with SPs, which do not provide clear policies to warn them about how their PII

can be used, they are usually unaware of secondary use of data and the existence of third parties.

According to the privacy taxonomy defined in [3], secondary use consists in the use of data for purposes other than those for which they were initially collected without the consent of the subject, e.g., the use of personal data collected on social networks for offering personalized advertising. This practice, thus, may violate the privacy of the user and cause damage since it involves using information in a non-consensual and possibly unwanted manner [3]. Nonetheless, whether certain action violates the privacy of a user depends on the perception of such user and his or her willingness to share given types of data. This, therefore, raises the need of collecting and respecting the privacy preferences of users.

An important aspect of the implementation of privacy in the cloud is Identity Management (IdM), which allows Identity Providers (IdPs) to centralize user's identification data and send it to SPs in order to enable the processes of authentication and access control [4]. IdM systems, such as OpenId Connect [5], allow the creation of federations, i.e., trust relationships that make possible for users authenticated in one IdP to access services provided by various SPs belonging to different administrative domains. An example is when users authenticate in different services with their Facebook accounts. In this case, Facebook acts as an IdP.

Even though there are several approaches that are intended to allow users to define their privacy preferences and organizations to express their practices, they are poorly adopted by both users and companies because they do not offer practical methods. In addition, most of them do not consider the decentralized nature of federated cloud environments. Consequently, IdM systems do not offer effective mechanisms to collect user's privacy preferences and to send them to the SP and, therefore, users are still unable to control how their PII can be used, by whom and for what purposes [1].

Werner and Westphall [6] proposed a privacy-aware identity management model for the cloud in which IdPs and SPs interact in dynamic federated environments to manage identities and ensure user's privacy. The model, while allowing users to choose and encrypt the data that can be sent to the SP, does not define a mechanism for determining users' privacy preferences and allowing them to control the use and sharing of their PII.

In order to complement the aforementioned model, this paper presents a mechanism for identity management systems that instructs users about the possible uses of their personal data by service providers and allows them to set their privacy preferences. These preferences are converted into a standardized, machine-readable structure, called privacy token, which is then sent to the SP along with other authentication data.

The remainder of this paper is organized as follows. Section II describes basic concepts relevant to the understanding of the proposal and Section III presents the main related work. In Section IV, the proposed mechanism for user's privacy preferences in IdM systems is introduced and a prototype implementation of the mechanism is described. Finally, conclusion and future work are presented in Section V.

II. BASIC CONCEPTS

This section presents the definitions of concepts considered important to the understanding of the proposal of this paper.

A. Identity Management (IdM)

IdM is implemented through IdM systems such as OpenId Connect [5], and is responsible for establishing the identity of a user or system (authentication), for managing access to services by that user (access control), and for maintaining user identity profiles [7].

Typical identity management systems involve three parts: users, identity providers, and service providers [7]. The user visits an SP, which, in turn, relies on the IdP to provide authentic information about the user. These systems enable the concept of federated identity, which is the focus of this work and allows users authenticated in various IdPs to access services offered by SPs located in different administrative domains due to a previously established trust relationship [8].

Some important IdM concepts are described next, as defined in [4][9][10]:

1) Personally Identifiable Information (PII): information that can be used to identify the person to whom it relates or can be directly or indirectly linked to that person. Thus, depending on the scope, information such as date of birth, GPS location, IP address and personal interests inferred by the tracking of the use of web sites may be considered as PII.

2) PII Principal: natural person to whom the PII relates.

3) Identity Provider (IdP): party that provides identities to subjects and is, usually, responsible for the process of authentication.

4) Service Provider (SP): party that provides services or access to user's resources and, for that, requires the submission of valid credentials.

B. Privacy

In this work, which focuses on IdM systems and federated cloud environments, privacy is considered to be the right of a user to decide if his or her PII can be used, by whom and for what purpose [3][10][11].

1) Privacy policy: set of statements that express the practices of the organizations regarding user data collection, use, and sharing.

2) Privacy preferences: preferences and permissions of a user for the secondary use of his or her PII, i.e., they determine by whom and for what purpose a PII can be used.

There are several approaches that are intended to express policies and privacy preferences, and the ones considered most significant for this work are described in the next section, along with other relevant privacy-concerned studies.

III. RELATED WORK

Platform for Privacy Preferences (P3P) [12] is a protocol designed to inform users about the practices of collecting and using data from websites. A P3P policy consists of a set of eXtensible Markup Language (XML) statements applied to specific resources such as pages, images, or cookies. When a website that has its policies defined in P3P wants to collect user's data, the preferences of that user are compared to the corresponding policy. If this is acceptable, the transaction continues automatically; if not, the user is notified and can opt-in (accept) or opt-out (reject). This work provides a basis for collecting user preferences, but it requires every user and SP to define their policies in this language and does not meet the needs of federated cloud environments.

Enterprise Privacy Authorization Language (EPAL) [13] is a formal language designed to address the industry's need to express organizations' internal privacy policies. An EPAL policy defines a list of hierarchies of data categories, user categories and purposes, as well as sets of actions, obligations, and conditions. These elements are used to formulate privacy authorization rules that allow or reject actions. Nevertheless, as it is specific for internal corporate policies, it does not consider user's preferences and is not suitable for privacy in federated identity environments.

Purpose-to-Use (P2U) [2] was proposed to provide means to define policies regarding the secondary use of the data. It is inspired by P3P, but allows the specification of privacy policies that define the purpose of use, type, retention period, and price of shared data. This language, although it enables user-editable and negotiable policies, is complex for users as it assumes that they have privacy policies and are able to define them in P2U. It also requires the SPs to have their policies defined in the same language.

Basso et al. [14] define a UML profile to assist in the development of applications and services that need to be consistent with the statements of their privacy policies. The authors identify privacy elements, such as policies and statements, through which organizations can define their policies for collecting, using, retaining, and releasing data; and organize their relationships into a conceptual model. This model is then mapped to a UML profile defined by stereotypes, attributes, and constraints that allow modeling statements of actual privacy policies. Although this profile helps application developers, it does not offer practical means for users to set their privacy preferences and transmit them to SPs.

Chanchary and Chiasson [15] performed an online survey to understand how users perceive online tracking for behavioral advertising. They demonstrated that users have clear preferences for which classes of information they would like to disclose online and that some would be more prone to share data if they were given prior control of tracking protection tools. The authors also identified three groups of
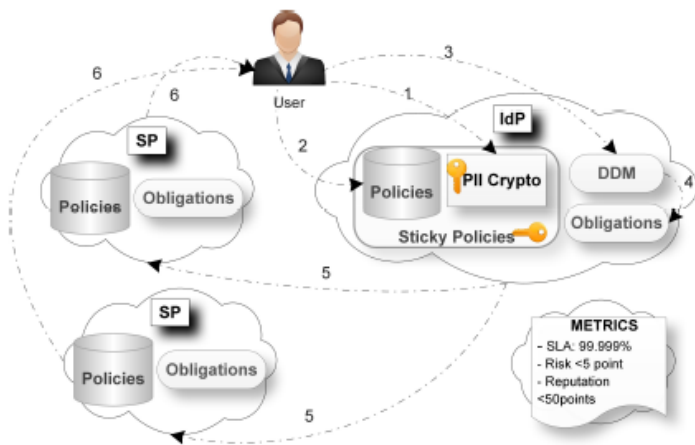
Figure 1. Interaction model between user, IdP and SP proposed in [6].

users according to how their privacy attitudes influenced their sharing willingness. These groups are used as a basis for the privacy profiles of our mechanism and are presented next:

1) Privacy Fundamentalists (30.4%): consider privacy as a very important aspect and they feel very strongly about it.

2) Privacy Pragmatists (45.9%): consider privacy as a very important aspect but also like the benefits of abdicating some privacy when they believe their information will not be misused.

3) Privacy Unconcerned (23.6%): do not consider privacy an important aspect or do not worry about how people and organizations use their information.

Werner and Westphall [6] present an IdM model with privacy for the cloud in which IdPs and SPs interact in dynamic and federated environments to manage the identities and ensure the privacy of users. They propose predefined, customizable privacy settings that help users to declare their desired level of privacy by allowing them to choose the access model, which can be anonymous, pseudonymous, or with partial attributes, and warning them about the reputation of the SP.

The interaction model defined in [6] and shown in Figure 1 proposes the registration in the IdP of the user's attributes and credentials, which may be encrypted (step 1), as well as the privacy policies to regulate the use and dissemination of their PII (step 2). Both the data and the policies are encapsulated in a package called sticky policies, which is sent to the SP along with a data dissemination model and obligations that must be fulfilled by the SP. The idea of the sticky policies is that PII are always disseminated with the policies governing their use and dissemination so that the user's privacy preferences are met by any SP. If the policies of the SP and the sticky policies are compliant, a positive reputation assess is generated for the SP; otherwise, a low reputation score is returned. The authors, however, do not define a mechanism for collecting these preferences, converting them into a machine-executable structure and sending them to the SP.

IV. PROPOSAL FOR A PRIVACY PREFERENCES SPECIFICATION MECHANISM

The proposal of this paper consists in a mechanism to incorporate to the OpenId Connect protocol a privacy token, which allows users to have a profile with their privacy preferences that is always sent to the SP along with their data. These profiles are based in a three-dimensional representation of the possible uses of PII.

The proposed mechanism allows users to choose a predefined privacy profile or to create a personalized one by choosing to opt-in or opt-out of each privacy preference. This profile is then transformed into a secure JSON Web Token (JWT), similar to the ID and access tokens already used by the OpenId Connect protocol.

A. Classification of Possible Uses of PII

Due to the large amount of possible actions and methods for collecting and sharing data, it is unfeasible to thoroughly list them. Therefore, this paper proposes a generic model that, on the one hand, is useful for users to set their privacy preferences and, on the other hand, works as a reference for SPs to assess whether the business rules of their data collection applications meet these preferences.

For this purpose, possible uses of the PII were classified in a three-dimensional structure. The dimensions, along with their respective abbreviations, are described next:

1) Data type: category of the PII to which the preference refers. The attributes of this dimension are: *Personal Information (PI)*, which encompasses any kind of information that represents the PII principal, such as name, national identifiers, parents' names, home address, photo and credit card number; *Personal Characteristics and Preferences (PCP)*, which are considered to be the physical attributes of the PII principal and personal options like weight, religious or philosophical beliefs, and sexual orientation; *Location (LO)*, which refers to any information about where the user is or has been and his or her trajectories with any precision degree and obtained by any means, such as GPS, Wi-fi networks or telecommunications systems; *Activities and Habits (AH)*, which are any activities performed by the user and habits inferred from tracking, such as web sites visited, purchases, and behavioral profile; and *Relationships (RS)*, people with whom the PII principal is in a specific moment or interacts through means like social networks, emails, and instant messengers.

2) Purpose: purpose for which the PII can be used. The values of this dimension are: *Service Improvement (SI)*, *Scientific (SC)*, and *Commercial (CO)*.

3) Beneficiary: party that benefits with the use of the PII. The attributes are: *PII Principal (PP)*, *Service Provider (SP)* and *Third Party (TP)*.

The dimensions above define a structure in which each position represents a rule that expresses a user's privacy preference that must be respected by the SP. This way, each of these rules comprises three parts: the type of data the rule refers to, for what purpose it can be used, and for the benefit of whom it can be used. For example, a user can define that his or her location data can be used for the purpose of improving services for the benefit of the PII principal and, in another rule, define that the same type of information for the same purpose cannot be used for the benefit of a third party.

By using this classification, the privacy preferences can be collected in a detailed manner or through four predefined profiles, which are described in the next section.

B. User's Privacy Profiles

Four privacy profiles were defined based on the work in [15], presented in Section III, which classified users into three groups according to their privacy concerns. For offering more privacy options and as it had the highest percentage of users, the Privacy Pragmatist group was divided into two different profiles. Therefore, the proposed profiles are:

1) Privacy Fundamentalist: This profile is aimed at users who have very high concerns with their privacy and do not wish to share any kind of information. Some functionalities or services, however, may not work properly or at all when this profile is chosen.

2) Privacy Aware: This profile represents users who are concerned about their privacy but still want to enable services even though some functionalities are compromised.

3) Privacy Pragmatist: This profile is aimed at users who still want some privacy but also want to enable most of the services and functionalities.

4) Privacy Unconcerned: This profile is for users who are not concerned about their privacy or how their PII are used, hence any data can be disclosed for any purpose and in the benefit of anyone. All services and functionalities should work properly with this profile.

Beside simplifying the process of setting the privacy preferences, these profiles are clarifying for the users as they inform about levels of risks to privacy and the possible uses of their PII and, as a result, assist them in making a conscious decision. In addition, users have the possibility to customize their privacy preferences using any of the profiles above as a basis.

C. Privacy token

Once the profile is chosen or customized, the privacy preferences, along with additional information, are converted by the IdP into a JSON (JavaScript Object Notation) object, which is then used as the payload for creating a signed JWT, called privacy token. This token is encoded into a base 64 URL-safe string for easy transmission to the SP, without compromising performance. After receiving the token, the SP must validate it in order to verify its integrity.

The structure of the privacy token, illustrated in Figure 2, comprises three sections. The first one is the header, which declares that the data structure is a JWT and defines the security algorithm chosen and implemented by the IdP (in this example, SHA-256); the second section consists of the claims set, which is explained next; and the last section contains the signature of the token.

The claims set includes two parts. The first one defines the following claims inherited from the ID token: *sub*, which is the subject identifier, i.e., a sequence of characters that uniquely identifies the PII principal; *iss*, which identifies the authority issuing the token, i.e., the IdP; *aud*, which represents the intended audience, i.e., the SP; and *iat*, which declares the time at which the token was issued.

The second part of the claims set define the privacy preferences of the user. Each claim corresponds to a position of the structure presented in Section IV-A, i.e., a privacy preference, and has a boolean value. The structure of a claim is as follows: the first abbreviation represents the type of data, the second abbreviation refers to the purpose, and the last one

```
{
  "typ": "JWT",
  "alg": "HS256"
}
{
  "sub"            : "alice",
  "iss"            : "https://openid.c2id.com",
  "aud"            : "client-12345",
  "iat"            : 1488405983,

  "PI_SI_PP"       : true,
  "PI_SI_SP"       : false,
  "PI_SI_TP"       : true,
  "PI_SC_PP"       : true,
  "PI_SC_SP"       : true,
  "PI_SC_TP"       : false,
      ...
}
{
  D7SDQBpVCSRSqVUMP9PAungM0gh7JKjKgXYhUlKMr3Y
}
```

Figure 2. Structure of the privacy token.

represents the beneficiary. For example, if the value of the attribute *LO_CO_SP* is true, it means that location data can be used for commercial purpose in the benefit of the SP.

The privacy token must always be passed along with the ID token, for instance, when the ID token has expired and a new one is requested to the IdP, when passing identity to third parties or when exchanging the ID token for an access token. This is necessary to ensure that users' PII are always accompanied by the corresponding privacy preferences. This way, with the addition of the privacy token, the OpenId Connect modified flow presented in [6] would be extended, as shown in Figure 3, to encompass the following steps:

1) The user requests access to a resource in the SP;
2) The security manager at the SP asks for the user to authenticate in the IdP where she or he is registered;
3) The IdP asks for the user's credentials;
4) The user provides his or her credentials;
5) The IdP validates user's credentials and returns the ID token and the privacy token to the user, who passes it to the SP;
6) The SP sends the ID and the privacy tokens to the IdP for the proof of validation;
7) The IdP verifies the tokens and confirms their validity to the SP;
8) The SP verifies whether the preferences can be met. If not, the SP asks the user for permission;
9) If the user authorizes, the IdP generates a new privacy token according to the user's response;
10) The IdP sends the new privacy token to the SP;
11) The SP requests additional attributes to the IdP;
12) The IdP shows the data dissemination scopes supported by the SP for the user to choose;
13) The user chooses one of the scopes, and informs the IdP about the selected scope;
14) The IdP provides the data to the SP according to the selected scope;
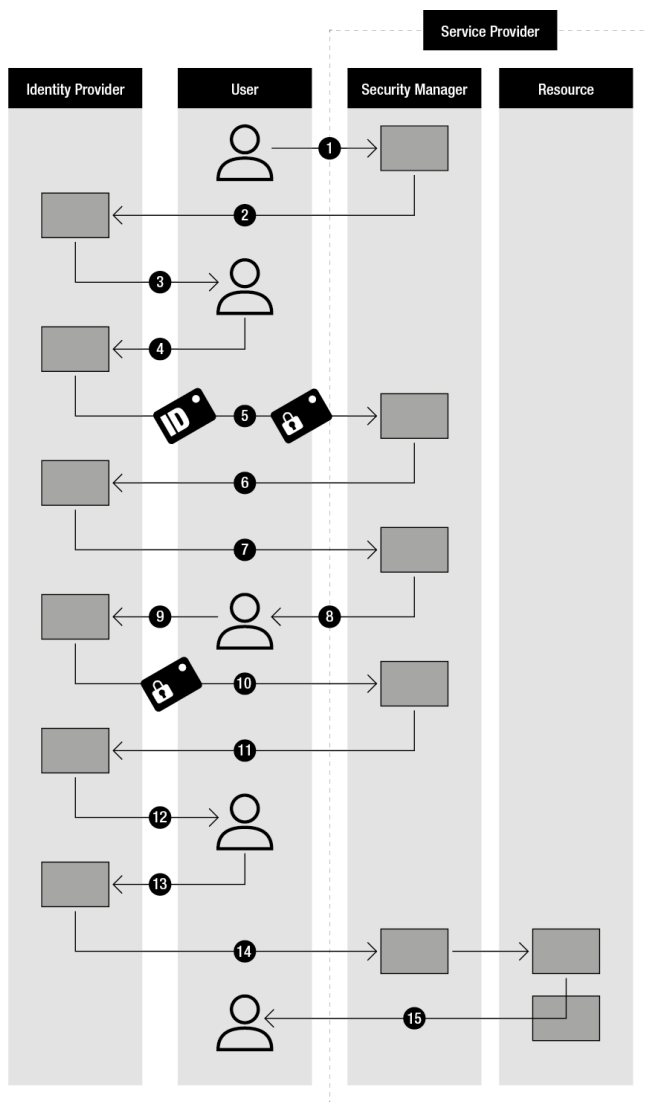15) The SP allows the user to access the desired resource.

Figure 3. Extension of the IdM flow proposed in [6] with the addition of the privacy token.

The privacy profile that is used for generating the privacy token sent to the SP in Step 5 is chosen or customized by the user during the process of registration in the IdP. In order to offer more flexibility, users can change their choice at any moment requesting it to the IdP.

D. Prototype

In order to verify the correct operation of the proposed mechanism and serve as the base for a future extension of an implementation of the OpenId Connect protocol, a prototype was developed. It is a Web application implemented in Java that performs the processes of collecting the user's privacy preferences through four predefined profiles or a customized one and generating a privacy token from them, as described in Sections IV-B and IV-C, respectively.

The prototype comprises classes representing the IdP, the SP, the user, the user's privacy preferences, and the privacy token. The *User* object is defined by personal data collected through a registration form and the *PrivacyPreferences* at-

tributes are set with the values corresponding to the selected or customized privacy profile, which along with the *IdP* and the *SP* objects form the *PrivacyToken* object. The actual token is then created from this object with Nimbus JOSE+JWT [16], a Java library for the creation and verification of JWTs, and signed with Hash-based Message Authentication Code (HMAC) using SHA-256 algorithm. After generating the token, it is possible to see the output string that should be passed to the SP and to validate it, by verifying the signature.

Figure 4 presents the screen where the user can select a privacy profile. Aiming at usability, each profile is represented by a number, a name, a brief yet expressive description, and an icon. Also, colors are used to help differentiate the profiles and represent the levels of risks to privacy in each of them, being red for the profile with the highest risks and green for the one with the lowest risks. A *See details* button shows the complete profile, i.e., all the privacy preferences of the corresponding profile for more information about the possible uses of the user's PII.

The custom profile option comprises five sections, one for each data type and presents to the user options to opt-in or opt-out of each preference regarding the purpose and the beneficiary of the use of the PII belonging to the given data type. In this option, the user can choose one of the four profiles as the base for personalization.

V. CONCLUSION AND FUTURE WORK

In this paper, a practical mechanism that allows users to control how their PII can be used in a federated cloud environment was presented. The mechanism instructs them about the possible uses of PII by SPs, allows them to choose between four predefined privacy profiles or customize one, and sends their privacy preferences to the SP along with their authentication data in a standardized, machine-readable format.

To the best of the authors knowledge, existing work focuses either on low-level approaches, such as privacy policy languages, which can be executed by machines; or on conceptual, high-level specifications, such as UML profiles, which provide a better understanding about privacy requirements in the development of systems and applications. However, these approaches do not offer practical means for users to set their preferences and send them to the SP, and/or require the latter to express all their policies in a specific way.

The main contribution of this work is the privacy token, a secure JWT that inverts the current scenario where users are forced to accept the policies defined by SPs by allowing them to express their privacy preferences. These preferences are stuck together to their data and are used by the SP to align its actions or request specific permissions.

The mechanism does not require SPs to use any specific standards to express and implement their privacy policies. It is only expected for SPs to adapt their data collection systems to interpret and fulfill the preferences expressed in the privacy token, which they can already read and understand once it has the same format as the other tokens used by OpenId Connect.

With the development of this work, it is expected that the model will be implemented in IdM systems and used in federated cloud environments to enable user privacy allowing them to control their PII. Thus, it is also expected to increase

Figure 4. Prototype screen with the four predefined privacy profiles and the customizable one.

their trust in cloud SPs and, consequently, promote greater adoption of the paradigm.

As future work, we intend to verify and improve the classification of possible uses of PII based on privacy standards and case studies. We also intend to extend an implementation of the OpenId Connect to support the presented mechanism. Furthermore, it is proposed to assess the consequences for services, SPs and users of applying this mechanism in real federated cloud scenarios.

REFERENCES

[1] J. Zhao, R. Binns, M. Van Kleek, and N. Shadbolt, "Privacy Languages: Are We There Yet to Enable User Controls?" in Proceedings of the 25th International Conference Companion on World Wide Web, Montreal, Quebec, Canada. International World Wide Web Conferences Steering Committee, Apr. 2016, pp. 799–806, ISBN: 978-1-4503-4144-8.

[2] J. Iyilade and J. Vassileva, "P2U: A Privacy Policy Specification Language for Secondary Data Sharing and Usage," in Proceedings of the 2014 IEEE Security and Privacy Workshops, San Jose, CA, USA. IEEE Computer Society, May 2014, pp. 18–22, ISBN: 978-1-4799-5103-1.

[3] D. J. Solove, "A Taxonomy of Privacy," University of Pennsylvania Law Review, vol. 154, 2006, pp. 477–564.

[4] M. Benantar, Access Control Systems: Security, Identity Management and Trust Models. Springer, New York, 2006, ISBN: 978-0-387-27716-5.

[5] "OpenId Connect," 2015, URL: http://www.openid.net/connect/ [accessed: 2017-03-13].

[6] J. Werner and C. M. Westphall, "A Model for Identity Management with Privacy in the Cloud," in Proceedings of the 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, Italy. IEEE, Jun. 2016, pp. 463–468, ISBN: 978-1-5090-0679-3.

[7] G. Alpár, J. Hoepman, and J. Siljee, "The Identity Crisis. Security, Privacy and Usability Issues in Identity Management," Computing Research Repository, vol. abs/1101.0427, 2011.

[8] D. Temoshok and C. Abruzzi, "Draft NISTIR 8149: Developing Trust Frameworks to Support Identity Federations," 2016, NIST, Gaithersburg, MD, United States.

[9] E. Bertino and K. Takahashi, Identity Management: Concepts, Technologies, and Systems. Artech House, Norwood, 2011, ISBN: 978-1-60807-039-89.

[10] "ISO/IEC 29100. International Standard - Information Technology - Security Techniques - Privacy Framework," 2011, URL: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45123 [accessed: 2017-03-13].

[11] "OASIS Privacy Management Reference Model and Methodology (PMRM) Version 1.0," 2016, URL: http://http://docs.oasis-open.org/pmrm/PMRM/v1.0/PMRM-v1.0.html [accessed: 2017-03-13].

[12] "The Platform for Privacy Preferences 1.1 (P3P1.1) Specification," 2006, URL: https://www.w3.org/TR/P3P11/ [accessed: 2017-03-13].

[13] "Enterprise Privacy Authorization Language (EPAL 1.2)," 2003, URL: https://www.w3.org/Submission/2003/SUBM-EPAL-20031110/ [accessed: 2017-03-13].

[14] T. Basso, L. Montecchi, R. Moraes, M. Jino, and A. Bondavalli, "Towards a UML Profile for Privacy-Aware Applications," in Proceedings of 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), Liverpool, United Kingdom. IEEE, Oct. 2015, pp. 371–378, ISBN: 978-1-509001-552.

[15] F. Chanchary and S. Chiasson, "User Perceptions of Sharing, Advertising, and Tracking," in 11th Symposium On Usable Privacy and Security (SOUPS), Ottawa. USENIX Association, Jul. 2015, pp. 53–67, ISBN: 978-1-931971-249.

[16] "Nimbus JOSE + JWT," 2017, URL: http://www.connect2id.com/products/nimbus-jose-jwt/ [accessed: 2017-03-13].