Survey on Privacy Preservation and Identity Management Systems

Uzair Ahmad Khan, Khurram Mustafa
Department of Computer Science
Jamia Millia Islamia
New Delhi, 110025

Abstract

Internet has become an integral part of our life, users are dependent on internet for various day to day tasks starting from emails, online purchases, health records, financial details etc. Most of the interaction on internet if not all, require user to prove their identity and they also need to know with whom they are interacting with (i.e. the identity of other party involved in communication). Digital identity has an important role in internet economy. Identity management systems have been primarily designed to store and manage entity (person or service) and their attributes. Different identity management system model has been designed and produced in the market. While few consider users privacy aspect as the key driver others consider business aspect, government aspect, security aspect etc. Over the last two decades identity management has evolved, starting from the centralized(silo), to federated to user centric identity management system. Literatures are found for all these models and their respective identity solutions in the market. This paper is an attempt to review identity management system which has been designed considering privacy as the main aspect. This paper will also highlight some of the challenges in the field of identity management system and its future avenues.

Keywords: IDM, Privacy, Single Sign on, Identity, Federated identity management, authentication, authorization, security.

1. Introduction

The online world has brought revolution in terms of easier online information retrieval, online purchases, monetary transactions, learning etc. Though the advantages of online revolution are immense, its darker side can’t be ignored. Industries have witnessed the instances of privacy breaches in the recent past which has caused not only huge financial loss to the companies but also hurt its repute and users’ confidence in the online system. The recent privacy breach [32] reported the loss of around 18 million data records of the office of personal management USA, which has integration with homeland security and federal bureau of investigation, this data contains quite sensitive information such as biometrics data about user. The user’s distrust in online system can have direct impact on the blooming internet economy. Identity theft is another issue which can hurt an online user when the confidential details provided by him at different portals are aggregated and thus the identity of the person can be revealed. One can imagine the consequences of unprotected identity of online users. The online systems are not robust against securing the sensitive details of the users which can lead to disturbance not only to the influential users but also to the normal online user. Privacy preservation and identity management are crucial aspects in any online transaction and therefore online platform needs appropriate infrastructure to ensure the protection of identity and privacy of identity related data.

Websites such as online retailers, social networks and search engines often publish aggregate statistics about their users to realize valuable social and economic utilities. These published statistics gets updated over time when new data arrive. Such practices are ubiquitous and we name a few examples below. Sites such as Amazon, IMDB, Delicious and Flickr recommend popular items or content to users to enhance their browsing experience and engage their interests.

At the first stance it seem harmless to release aggregate information about users. However, previous work has shown that sensitive information can be extracted about an individual user with such statistical disclosures. In particular, sites that continually update the published statistics over time can give even more leverage to the adversary and result in more severe privacy leakages.

As we have seen that identity is quite an important and sensitive entity while transacting over internet. IDM has become an interesting and challenging area of research in industry and academia[35,37,39]. Identity is one of the driving force to the internet economy [36]. According to author in [66] “Identity
management is described as the framework and system used in computer or communication systems to control identity”.

2. Literature survey

2.1. Identity:

Identity is defined as the unique characteristics of an entity. The identity which is used for identification purpose is called an identifier [2]. Author in [9] states that “Digital identity is the electronic representation of users’ or organizations’ personal information”. An entity (person or organization) can have various attributes (e.g. age, salary, social security number, credit card number). Depending on the context and situation only subsets of users attributes are needed to represent a person both in the physical and the digital world, these are called (digital) partial identities [44].

In the era of internet, the digital identity of users is used for online activities like communication, sharing contents or doing transactions across various organizations. The service providers (SP) use the entity identifiers for authentication purpose [1].

A person typically uses different partial identities for work, others for leisure activities (e.g., doing sports, or with the family), or dealing with companies (e.g., a bank, a bookstore) [45]. Some partial identities containing the information which other communication partners typically know about a person. Some information is static (e.g., birthday) while other might change dynamically (e.g., interests) [45].

2.2. Identity Management System (IDM):

In a typical online interaction there are three main parties involved with respect to identity management system. 

Entity- Person, Organization, service. User Agent- Typically a web browser in case of human user. Identity provider (IDP) - which stores identity information about an entity. Service provider (SP) - The online website that user want to interact with (e.g if user is accessing bankofamerica.com then this will be service provider).

In [13] Clarke defined IDM as the process, policies and technology which authenticates the users and provides access and privileges to them. In application centric Identity Management (IDM) model, each application stores the Personal Identifiable Information (PII) of entities (which can be users or services). In application-centric IDM, the SP keeps track of the entities using the services. In an IDM, the identity provider (IDP) manages the entities identity information and authenticates the user; while the service provider (SP) provides the entities the access to the services [Bhargav 2007]. Trusted interactions between users and SPs are desired, but it is challenging to provide such trusted interaction in an environment where client platforms are infected by malware.

Local user-centric identity management is a novel concept that provides a solution to this challenge, and one of the solution is an OffPAD, which is an offline personal authentication device [Josang 2015].

The two categories of Identity Management Systems (IdMSs) are centralized or decentralized. A centralized IdMS can be maintained easily as compared to a decentralized IdMS, however, a centralized IdMS permits the identity provider to monitor all activities on the system, which is off course not in the interest of users privacy [7].

In cloud environment, IDM is more challenging as an entity can be associated with multiple accounts or in other words an entity can have multiple digital identities. One solution to this issue is provided by [1], authors proposed an entity centric approach for IDM in the cloud. Another challenge in IDM is dispense the functionality of IDPs among IDPs and SPs [7]. To elaborate this, there should be secure and privacy-preserving mechanisms to retrieve user identification information from different SPs. In addition to this, only the information which is required to access the services are to be provided to the SPs.

In [47] Cameron has analyzed IDMS and the reason for their failure and adoptability in the market and he came up with certain guiding principles which are essential for the success of IDMS, These principles are as follows.

2.2.1. Minimal Disclosure for a Constrained Use:
An IDMS should disclose less PII (personally identifying information) and restricts the usage until unless it is utmost necessary for the transaction to take place.

2.2.2. User Control and Consent:
If at all any transaction requires disclosure of the identity information, it must seek users consent, so that user can take informed decision and have sense of control over the PII’s.

2.2.3. Justifiable Parties:
An identity management system must be designed in such a manner that identifying information is disclosed only to parties having a necessary and well justifiable need.
2.2.4. Directed Identity:
An identity management system must support global identifiers for use by public entities and local identifiers for use by private entities.

2.2.5. Pluralism of Operators and Technologies:
An identity management system should be flexible enough that it must support interoperability of multiple identity technologies run by different identity providers.

2.2.6. Human Integration:
An identity management system must employ unambiguous machine-human communication mechanisms that prevent identity-based attacks (for instance, impersonation and phishing).

2.2.7. Consistent Experience across Contexts:
An identity management system must provide a simple, consistent experience to users while supporting multiple operators and technologies.
User Control and Consent and Minimal Disclosure for a Constrained Use are what can be termed as confidentiality properties. Cameron argues that users should control attribute dissemination. In particular, identity management systems should provide users with information, such as an attribute-use policy, that enables them to make informed decisions about attribute dissemination. Second law, states the mechanisms of dissemination of coarser form of identity attribute.

Existing work in the direction of identity management systems focuses primarily on individual systems, each of which focuses on one of three general types of functionality mentioned in [58].

2.2.8. Single sign-on:
These systems issue authentication assertions to multiple service providers after a single user authentication. Examples include Passport, Shibboleth (http://shibboleth.net), OpenID (http://openid.net), and Facebook Single Sign-On.

2.2.9. Federated identity:
These systems manage multiple distinct identities for a single user and issue authentication assertions on the basis of any of these identities. Examples include CardSpace, and Client-Side Federation [2], Project Liberty (http://projectliberty.org), Higgins (www.eclipse.org/higgins), PRIME (www.prime-project.eu).

2.2.10. Anonymous credentials:
These systems provide authentication assertions that don’t reveal the users identity to a service provider. Examples include Idemix [7,40] U-Prove, and P-IMS [50,51]

3. Privacy preservation
Privacy is the control provided to the entities to reveal or hide their identifier information at their discretion. In cloud environment, the users must be provided a handle to control their identification over the cloud or the cloud service provider. But in the cloud environment, securing the personal information of entities must be done meticulously as the cloud service providers can store and disseminate entities identifiers at multiple locations [8]. In [4], authors have proposed to disseminate sensitive data while preserving the privacy aspects. In the privacy preserving scheme [4], authors have proposed bundling the data and meta-data, evaporate it in unfriendly environment and apoptosis of attacked bundle (s).

Users have to maintain and remember multiple identities used for authentication and access to various websites. Users are overloaded with identities and suffer from password fatigue [15]. In [15], authors have proposed to make users able to manage and control their digital identities. Their work focuses on usability and privacy aspects of IDM systems. In [42] authors proposed a protocol messaging scheme which is used to protect all information using one-time passwords used in a dynamic multiple application.

Privacy-driven identity management systems are designed around three privacy properties: [53,57]

undetectability—Hiding user actions
confidentiality—enabling users’s control over dissemination of their attributes.
unlinkability—hiding correlations between identities and combinations of actions.

Identity management systems contain users’ personal information as identifiers and thus have direct impact of protecting the privacy. Privacy preservation of users’ identity is a serious concern as its leakage can lead to unsolicited mails or harassing phone calls by mischievous parties. Privacy can be safeguarded using policies, processes or technologies

Some of these are briefly discussed as:

i) W3C’s (P3P) Project [17] - A technological solution which enables that the privacy practices of websites are expressed in a standardized, XML-based format. However the drawback is that P3P doesn’t guarantee or enforce the privacy claims made by Web sites [41].

iii) OpenID [10] - federated login system.
v) PseudoID [16] - In federated IDM, PseudoID is a mechanism to protect users’ private login data for authentication.
vi) Smart card [42] - Security at the hardware level. Traditionally, the smart cards are used for single specific purpose. In [42] authors introduced the concept of utilizing the smart cards for multiple applications.

Other notable example include SPID Public Digital Identity System (SPID) [57], which is the Italian government framework compliant with the EU eIDAS regulatory environment, aimed at implementing electronic identification and trust service in e-government and business applications.

SPID authentication results in information leakage about customers of identity providers. To overcome this potential limitation, authors in [20] proposed a modification of SPID to allow user authentication by preserving the anonymity of the identity provider that grants the authentication credentials. This way, information leakage about customers of identity providers is fully prevented.

3.1. Entities in control:

The entities can have differentiated preferences for different types of personal information [7]. In federated IDM, securing the preferences of users distributed across organizations can be done using Automated-Trust Negotiation (ATN) techniques [7]. In [12], authors have presented an empirical study to demonstrate the relationship between identity and technology.

In [41], the authors proposed a user-centric approach to assist users while accessing online services. Authors in [24] have defined user-centric privacy protection principle as “Exposure of personal information must be minimized” which is one of the identity law proposed by Kim Cameron. In IDM this principle would mean that few parties should be involved in the management of identities used for online service access [Josang 2006]. Recently [65] have proposed a scalable, secure and user-friendly identity management solution. Their approach is based on keeping the IDM technologies to the user side instead of the server side.

Josang et al [65] proposed a device called OffPAD (Offline Personal Authentication Device). Different forms of authentication that are required for trusted interactions are provided by OffPAD.

The existing user identity management modes are:

- Silo model (Centralized identity management)
- Common Identity Domain Model
- Centralized Single-sign-on (SSO)
- Federated IDM or Federated Single sign on (SSO)

Federated systems were introduced to allow a seamless access to technology and services to the users [121]. In [7], authors have discussed Federated Identity Management which allows collaborative networking among multiple organizations while providing privacy to the entities involved. Federated IDM provide mechanism to access and manage user identifiers and other resources [7]. The federated IDM solution is a useful mechanism to authenticate users once and allowing them to access various federated organizations or group of Service providers SPs. Multiple silo domains are grouped to form a federated domain. The organizations in a federation are termed as circles of trust by Liberty Alliance [31]. The mechanism of federated IDM is an extension to SSO across organizations. For example, Facebook Connect, Live ID and OpenID 2.0 are able to offer one-click logins for relying parties. In federated Users trust identity providers to manage and protect their identity, so privacy concerns seems to be relatively minor [16].

Some of Federal IDM solutions are:

i) Liberty Alliance
ii) WS-Federation
iii) OpenID
iv) PRIME

3.2. Privacy preservation and privacy policies

3.2.1. In US:

USA has different privacy policies for medical information HIPPA (health information portability and accountability act) which acts as guiding and mandatory document for the software companies who make healthcare software and they have to abide by this privacy act.

Similarly they have privacy policy GLB(Gramme Leach Billey act) for financial institution who stores the customer data and for data related to children is being govern by COPPA(Children’s online privacy act).

3.2.2. In Canada:

In Canada the office of privacy commissioner have a release document that needs to be referred by software companies and the database owner and follow the guidelines with respect to privacy policy of data storage, data retention, purpose of retention, duration of retention. These are strict obligation that any software manufacturer and the database owner must comply with.
3.2.3. In Europe:

In EU was the first to consider the privacy comprehensively and they consistently work in the direction of strict policy framework. They have taken various initiatives in this direction; Prime project was the main project which deals with privacy and identity management is Europe.

3.3. Privacy preservation in context of Identity management system:

Identity management plays a key role for privacy protection, because a digital identity consists of personal information [41].

Privacy is seen as one of the major concern for the people using internet [56].

Europe legislation has alleviated this very concern of privacy preservation through their legislative privacy policy, but have weak support outside EU [62].

As discussed in previous section, different countries are dealing privacy concern through their national privacy act and as such there is no standardization when it crosses nation’s boundary.

Considering this inter dependency of privacy protection and country makes it not only difficult, but impossible to have universal privacy protection abiding identity management system.

There are various identity model. Next section details how privacy is being handled in these models.

3.3.1. Federated Identity management System:

Federated identity management is a setup where identity is shared across domains [54-55]. Within such a federation, additional agreements can be made for further optimisation, e.g. to have a centralized authentication authority. The so called circle of trust (CoT) equals the set of domains that belong to one federation. Note that a domain can belong to several federations and therefore can belong to several circles of trust.

In Federated identity management system one concern is the trust among the partners of federation. They need to have a mechanism to know the level of trust of the partner with whom they are required to share the PII of the user. To address this [43] has proposes TRIMS(Trust reputation in identity management systems) framework which applies a trust and reputation model to guarantee an acceptable level of security when deciding if a different domain might be considered reliable when receiving certain sensitive user’s attributes.

So the partners in federation can take informed decision with whom they will share attribute and with whom they will not share these attributes at all during transaction.

Some of the prevalent federated identity management systems are PRIME, Liberty Alliance, Openid, OAuth, idemix etc.

Features and functioning of some of the prevalent FIMS solution (Openid, Liberty Alliance, Prime, Information card) has been analyzed by [33] in detail with respect to their privacy promises and adoptability in the market. In the similar line [34] has performed comparative analysis of six FIMS solution (Openid, Liberty Alliance, Prime, Shibboleth, Information card, OAuth) on set of privacy requirements and concluded that none of these FIMS solution are ideal on privacy requirement, but still they are alleviating some of the privacy and identity management solution which is why they are being endorsed in the market. In [40] Hommel and Reiser has further researched on the shortcomings to existing FIdM which includes limitation to web service technology.

3.4. Identity management techniques:

Some of the prevalent technologies that are being used in identity management are listed as below.

1. P3P policies
2. Public Cryptography and PKI
3. SAML
4. Openid
5. Oauth.
6. SSO
7. Ping Identity

3.5. Privacy preservation techniques:

3.5.1. Cryptographic techniques:

In Sensitive information are stored and transferred over internet. Few researchers have proposed a solution to keep this sensitive information in encrypted format. The encryption of the data will be done on the basis of the identity attributes. The disadvantage of this approach is that user will not be able to share the data at fine grained level.

[57] proposes attribute based encryption to alleviate this issue of non-sharing of the data at fine grained level. In an attribute based encryption system, a user’s keys and cipher texts are labeled with sets of descriptive attributes and a particular key can decrypt a particular
ciphertext only if there is a match between the attributes of the ciphertext and the user’s key. Further improvement to the attribute based encryption technique has been handled [57] where they proposes a key policy based attribute based encryption technique.

[56,58,59,61] proposed the concept of anonymous credential system which advocates the principle of minimal disclosure and information disclosure in controlled manner.

[62] present a cryptographic framework that enables data minimization. In this framework, for each transaction, there is a precise specification of what data gets revealed to each participant. They call it “controlled release of data”.

3.5.2. User-centric approaches:

The biggest threat to an online user is the identity theft during communication, networking or sharing contents over internet. Even the network security technologies like SSL/TLS are not effective against identity thefts [22]. Schechter et al [22] reported that the theft or loss of computers accounted for over a 46% of identity theft cases.

The latest data from Identity Theft Resource Center (ITRC) reports the identity thefts at various categories of institutions like Banking/Credit/Financial, Business, Educational etc. In August 2016, ITRC reports 638 breaches in all categories of institutions with over 28,574,795 records thefts [22]. Over 7% of US citizens over 16 years of age were victims of identity thefts in 2014 [23].

The solution for securing the identity of entities must consider privacy, user centrality and generality issues [21]. In [21], authors have proposed a user-centric identity usage monitoring system which detects anomalies in identity usage using the context information of the request. The context information used to detect normal and anomaly behavior is based on timestamp, location, device characteristics [21] etc.

To ensure high privacy and user-control, the authors [21] proposed to deploy the monitoring system (to detect and report identity usage) on users’ devices or on a trusted third party.

Based on the side (users or servers) that have control over personal data, there are two broader categories of Identity management systems: user-centric and server-centric systems. IDM systems were developed according to the context of usage and its characteristics differ widely in terms of its application context.

4. User-centric IDM model

For privacy and user-centrality, user control is an important aspect. There are various systems developed to provide users a control over their credentials. Some of these are:

- CardSpace [28] – users can choose an appropriate identity credential for each transaction
- OpenID [27]
- VeriSign PIP - enables users to select to whom and how much information is disclosed

In [24] authors have proposed a user-centric framework for network identity management which manages users’ identities and protects user interests and business interests from user side.

In [25], authors have proposed to facilitate the users to manage their personal information in federated identity management system. Josang et al. [15] proposed a Federated Single Sign-On (FSSO) systems which has useful properties of the User-Centric Identity Management (UCIM) model. This identity management system called UFed allows the users to control and enforce their privacy requirements while still retaining the convenience of single sign on over a federation of service providers. UFed enables the user-controlled privacy in a federated IDM solution.

CredEx proposed in [29], offers flexibility for credential management in web and grid service environments based on open-source, interoperable standards and implementations. CredEx, is an open-source, standards-based, Web Service that facilitates the secure storage of credentials and enables the dynamic exchange of different credential types using the WS-Trust token exchange protocol.

In [30], authors proposed that multiple parties (including the user) control the disclosure of multiple attributes. The identity information of users is stored by multiple parties and aggregated on demand by a service provider. However, the system can also be used in online systems..

5. Conclusion

In this paper we have explored the journey of identity management system though the lense of privacy preservation and saw its limitation be it from privacy validity in the domain of country. Technical and security aspect which at time sidelines privacy while designing identity management systems, use of different underlying technologies as per the suitability. Usage perspective in terms of business centric, government centric idms to user centric.
Currently there are various federated identity management systems, but they have limited adoptability in their circle of trust. The other major concern in the federated identity management system is the level of trust to other parties of federation, controlled and restricted release of Personally identifiable information (PII) data and the most annoying fact the accountability in case of privacy breach and identity theft. There will always be a tradeoff while designing identity management solutions considering privacy preservation at forefront.

6. Future Scope

In this paper we studied identity management in perspective of privacy preservation. Through this paper researchers will be able to find the quantum of work that has been done in identity management system with respect to the privacy preservation. It will give them the pointer about the gap areas in this direction and further they can take up any of the challenges and extend the contribution in the direction of privacy preserving identity management research.

7. References


Electronic Government and the Information Systems Perspective (pp. 57-70). Springer International Publishing.


[45] Sebastian Clauß DIM’05, November 11, 2005, Fairfax, Virginia, USA.


