



Towards Privacy-Preserving Multi-Cloud Identity Management Using SOLID

Alfredo Cuzzocrea^{1,2}^a and Islam Belmerabet¹^b

¹IDEA Lab, University of Calabria, Rende, Italy

²Department of Computer Science, University of Paris City, Paris, France

Keywords: Digital Identity Management, Privacy-Preserving, Access Control, Identity and Access Management, Identity Management Protocol.


Abstract: Digital identity management services are essential for user authentication in *Cloud Computing infrastructures*. They allow for flexible access control to services based on the characteristics (also called attributes) of the user and the history of interactions. These services ought to safeguard users privacy while enhancing cross-domain interoperability and streamlining identity verification procedures. In this research, we provide a strategy for satisfying these requirements by fusing protocols for *Zero-Knowledge* proofing, semantic matching techniques, and high-level identity verification principles expressed in terms of identity attributes. The paper describes the fundamental strategies we employ as well as the design of a preliminary architecture based on these methods.


1 INTRODUCTION

Cloud Computing and *online services* are evolving *paradigms* for large-scale *infrastructures*. Cloud Computing provides several advantages, including cost savings, flexibility, sustainability, insight, and quality control. In the software business, Cloud Computing technologies such as *Amazon Elastic Computing Cloud (EC2)*, *Simple Storage Service (S3)*, and *Google App Engine* are widely used. However, despite the effect and efficiency of these application services, there are still *Security* and *Privacy* concerns about how these Cloud providers treat user data. The consequences of insecure Cloud Computing platforms may be found in a variety of technical paradigms, including *Web-Based Outsourcing*, *Mobile Cloud Computing*, and *Service-Oriented Architectures (SOA)*. A secure Cloud implementation necessitates an adaptive security mechanism to provide users with a high degree of *confidence* in the Cloud. Without the capacity of such solutions to ensure a significant degree of security and privacy, there will continue to be a major concern

of privacy loss and *sensitive data leakage*, limiting the wide adoption of *Cloud services* (Tari, 2014).

Privacy is a basic right that necessitates the proper use and safeguarding of personal information. Cloud Computing paradigms breach privacy in a variety of ways, including the *theft* of personal information (Deng, 2010), the *unregulated* use of Cloud services, *data propagation*, possibly *unauthorized secondary usage*, *trans-border data flow*, and *dynamic provisioning*. The privacy concerns revolve around *Identity and Access Management (IAM)* challenges, namely *identity provisioning* and *de-provisioning*, maintaining a *single ID* across numerous platforms and organizations, *compliance visibility*, and security when using a third-party or vendor network. Current procedures often prove consensus through a *third-party service* or the general terms and conditions for personal data processing. When providing user permission in an environment with limited or no user interface, security and privacy concerns become more problematic due to unauthorized data usage permission and insufficient processing of personal information, which is frequently overlooked during the design phase.

 <https://orcid.org/0000-0002-7104-6415> – This research has been done in the context of the Excellence Chair in Big Data Managment and Analytics, University of Paris City, Paris, France.

 <https://orcid.org/0009-0003-7878-0991>

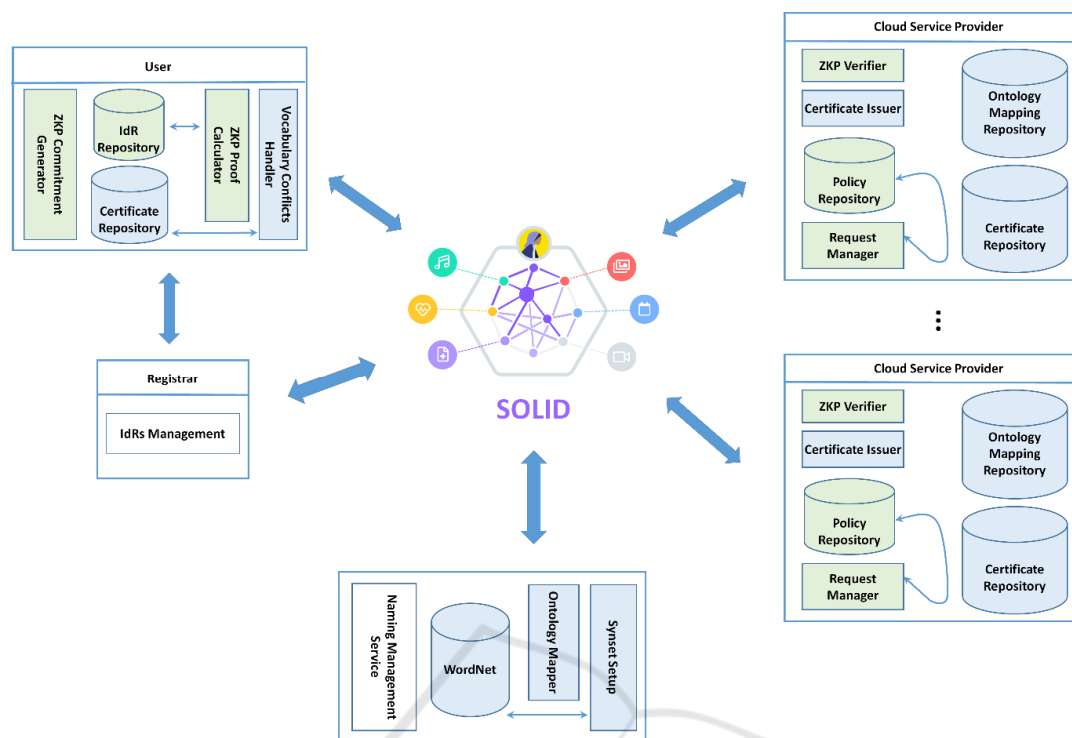


Figure 1: Schema of the Reference Architecture.

The primary issues concerning data security policies for Cloud users in terms of Cloud security implementation are the following ones. First, the *commitments* of *Cloud Service Providers* (CSPs) to ensure information security. Second, there are open and documented data security *policies*. Third, there are the measures set to categorize data access, as well as its justification via third-party *auditing*. As a result, when allowing third-party access, companies must design a data access hierarchy, and good identity management for third-party access should be a priority for any CSP (Kumar *et al.*, 2011). An inside attack can occur without proper identity management by distributing malicious programs on *edge nodes* and exploiting vulnerabilities that affect the *Quality of Service* (QoS). Such hostile behaviors can have a substantial impact on temporarily preserved sensitive data.

In this paper, we present an architecture that aims at improving identity verification management in a privacy-preserving manner by utilizing high-level identity verification policies expressed in terms of identity attributes, *Zero-Knowledge* proof protocols (e.g., (Goldreich & Krawczyk, 1996)), semantic matching techniques, and employing *SOLID Decentralized Secure Data Stores* (SOLID, 2024). The context of application is the *Multi-Cloud environments* (e.g., (Pawar *et al.*, 2015)). To deal with

this, we introduce the adoption of SOLID as one of the main relevant innovations of our proposal, which sees engrafting SOLID as a kind of Cloud service within our reference architecture, and delegating to it the privacy-preservation functionalities mapped on so-called *PODs (Personal Online Datastores)*. *PODs* are decentralized data stores where user data are secured once and used after across *multiple* systems (e.g., *Amazon, Facebook* and *YouTube*). By marrying the *SOLID PODs'* philosophy, we inherit the similar mechanism, initially developed for Web-based systems, to Clouds, with the goal of ensuring privacy-preserving user data management (including identity management) across *multiple* Clouds (e.g., *public Clouds* and *vs private Clouds*). Thanks to *PODs*, we can “transfer” *privacy-preserved user data* across *multiple Clouds*.

2 PRIVACY-PRESERVING IDENTITY MANAGEMENT OVER CLOUDS: AN OVERVIEW

With the growth of Cloud Computing, hundreds of users and many apps are communicating and exchanging sensitive data. As a result, it is critical to

manage identities securely while also protecting data privacy. Several research studies on privacy in Cloud identity management have been proposed to that purpose. Among them, the following ones are relevant:

- (Angin *et al.*, 2010) present *IdM Wallet*, a solution for *entity-centric Identity Management* (IdM) in the Cloud that employs an active package scheme. The active bundle is a container for *metadata*, access control policies, personally identifiable information, and the *virtual machine* (which manages and controls the program code included in a bundle). The zero-knowledge proof is used to authenticate an entity without exposing its *identifier*, resulting in an anonymous identification. With this idea, it is feasible to utilize identity data in unreliable hosts and to reduce sensitive data on the network by providing just the attributes required by each service provider;
- (Weingärtner & Westphall, 2014) combine the use of *encryption*, *policy management*, and notification of service provider *confidence levels*. Their method addresses the lack of control over user identification data when enrolling with *federated Identity Providers* (IdPs). Users can set their attribute distribution policies, choosing which *Personal Identifiers Information* (PIIs) are released. However, user-centric distribution policy management can cause issues because the majority of users lack appropriate expertise about policy generation and management;
- (Spyra *et al.*, 2016) deal with *data storage* in the Cloud and the protection of sensitive data. The proposal adds *eXtensible Access Control Markup Language* (XACML) to the *Office Open XML* (OOXML) document format, defining a sticky policy that ensures the integrity and credibility of information. To enforce the XACML policy, *cryptology-based identity* (IBE) is used as an authentication technique.

3 SOLID-EMPOWERED MULTI-CLOUD IDENTITY MANAGEMENT

The contribution of this research proposal is to improve the different phases of the reference *privacy-preserving management* of *digital identity* attributes in domains with *heterogeneous* name spaces

architecture shown in Figure 1 with particular regards to privacy preserving tasks using SOLID decentralized data stores, which establish a standardized framework for personal data storage and sharing on the Web. This specification enables individuals to exert fine-grained control over their digital identities, foster enhanced privacy and user agency, and epitomizes a paradigm shift by placing data ownership and control squarely in the hands of the user. As mentioned in Section 1, we use the same SOLID philosophy on *different* Web systems applied to *different* Clouds via PODs, so that achieving effective and efficient privacy-preserving identity management over multi-Clouds.

To address the problem of privacy-preserving management of digital identity attributes in domains with heterogeneous name spaces, this privacy-preserving multi-factor identity attribute *verification protocol* can match Cloud service providers and client vocabularies using a matching technique based on look-up tables, dictionaries, and ontology mapping techniques. The protocol uses an *Aggregate Zero Knowledge Proofs of Knowledge* (AgZKPK) cryptographic protocol to allow clients prove knowledge of multiple identity attributes with a single interactive proof without having to provide them in clear (Bertino *et al.*, 2009).

3.1 SOLID

SOLID is a specification that allows for storing data securely in decentralized data stores called PODs, where PODs are like secure personal Web servers for data. The main idea consists in creating, for every user (or user group) *one* POD that contains privacy-preserving user data and access it across multiple Clouds, without the need for re-identification. This approach enforces scalability and self-authentication, thus reducing the risk of cyber-attacks, by also introducing the nice amenity of limiting data entry activities that may increase the possibility of *identity thefts* and *personal-data attacks*.

In this case, SOLID stores the information related to user identity attributes used in this multi-factor identity attribute verification approach which is managed by the Registrar component, namely, *Identity Records* (IdRs) containing identity tuples for each user identity attribute. Each identity tuple consists of a tag, that is, an attribute name, the Pedersen commitment of the attribute value, the signature of the Registrar on the commitment, and two types of assurance, namely validity assurance and ownership assurance, and a set of *nyms* (weak identifiers) along with ontology mappings, set of

synonyms, session data, and mapping certificates provided by the *Heterogeneity Management Service*.

3.2 Identity Attribute Matching Protocol

An Identity Attribute Matching Protocol uses a combination of *look-up tables*, *dictionaries*, and *ontology mapping* in order to address the different variations in identity attribute names as follows:

- *Syntactic Variations*: refer to the use of different character combinations to denote the same term, they can be identified by using look up tables;
- *Terminological Variations*: refer to the use of different terms to denote the same concept, and they can be determined by the use of dictionaries or thesaurus such as *WordNet* (Miller, 1995);
- *Semantic Variations*: are related to the use of two different concepts in different knowledge domains to denote the same term, these can be solved by ontology matching techniques.

There are two important issues related to the identity *matching protocol* which are the following:

- Which party must execute the matching? And it is addressed by performing the matching on the CSP, as performing the matching at the client has the obvious disadvantage of the client lying and asserting that an identity attribute referred to in the CSP policy matches one of its attributes, which is not the case. The usage of ZKPK protocols preserves the user identity attribute privacy by ensuring that the CSP does not learn the values of these attributes – hence, the CSP has no reason to lie about the mapping;
- How to build on previous interactions the client has had with other CSPs? As a result, in order to make interactions between clients and CSPs faster and more convenient for users, the matching protocol relies on the use of *proof-of-identity certificates* – these certificates encode the mapping between (some of) the user identity attributes and the identity attributes referred to in the policies of CSPs with which the user has previously successfully interacted.

3.3 Multi-Factor Authentication

In the Multi-Factor Authentication process, once the client receives Match, the set of matched identity attributes from the CSP retrieves from the Registrar or from its certificate repository local to the client, the commitments satisfying the matches and the

corresponding *signatures*. The client then aggregates the commitments, and according to the ZPK sends the aggregated zero knowledge proof to the CSP.

If the aggregated zero knowledge proof is valid, the CSP accepts it, and if the aggregate signature verification is successful, the CSP issues a proof-of-identity certificate to the client. The certificate proves that the client identity attributes in the Match set are mapped onto CSP ontology concepts and that the client proved its knowledge of those attributes.

The CSP sends the proof-of-identity certificate to the client and stores a copy of the certificate in its local certificate repository. The proof-of-identity certificate can be given to another CSP to allow the client to prove knowledge of an attribute without performing the aggregate ZKP protocol. The CSP that receives the certificate just has to confirm the certificate validity.

3.4 Heterogeneity Management Service

The Heterogeneity Management Service consists of two modules: *Synset SetUp* and *Ontology Manager*. The inclusion of these modules within our digital identity management architecture underscores its critical role in enabling seamless interoperability and coherence across diverse data sources. The Synset SetUp module pivotal function involves querying local thesauri to extract an extensive array of synonyms corresponding to a specified term, thereby enhancing the comprehension and contextualization of digital identity elements. This facet assumes paramount importance in digital identity management systems, where precise understanding and interpretation of identity attributes are fundamental. Complementing this, the Ontology Manager module assumes a strategic role by facilitating ontology mapping functionalities.

Within the domain of digital identity management, this module ability to reconcile and align dissimilar ontological structures becomes indispensable. By transcending discrepancies in schema and semantics, the Ontology Manager module ensures the harmonization of identity-related data elements across varied ontological representations, thus fortifying the coherence and consistency of digital identities.

As an integral component of our digital identity management architecture, the Heterogeneity Management Service stands as a testament to its pivotal role in promoting semantic coherence and facilitating effective data integration within the realm of identity management systems, thus contributing to the overall identity protection goal.

3.5 UML Modelling of the Proposed Architecture

We provide an entity-centric, identity-centric-driven IdM methodology called *anonymous identification*, which is based on the usage of Zero-knowledge proof for entity authentication without revealing its identifier. Figure 2 shows anonymous identification and the IdM service topology, in the context of our reference Cloud architecture enriched by SOLID PODs, which allow us to achieve the multiple Cloud feature.

It is feasible to prove a claim or assertion (authenticate) using Anonymous identity without providing any *credentials*. Consider the following scenario: a customer purchases books from Amazon. To obtain the books via mail, the customer must submit his mailing address. In certain cases, many parties are involved in the same transaction and require distinct information from the user. The shipping firm needs the address. On the contrary, Amazon does not need to know the customer address, but wants to ensure that the user provides a valid address to the delivery business. In this situation, following Anonymous identification, the IdM service generates a token that comprises the address that must be revealed. Aside from the address, this token contains metadata, access control restrictions, and VM. This token is sent to the CSP, which may then distribute it to the mailing firm. The IdM service secures user attributes when transmitted to CSP and allows us to utilize it on untrusted hosts and send tokens.

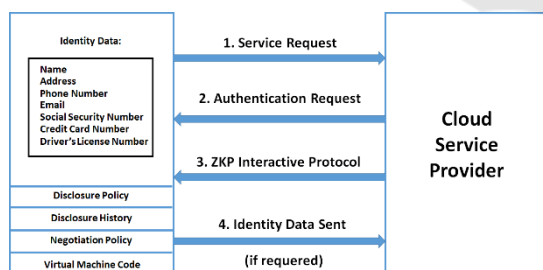


Figure 2: IdM Service Model.

4 CONCLUSIONS AND FUTURE WORK

In conclusion, digital identity management services are critical in Cloud Computing infrastructures for authenticating users and supporting flexible access control to services based on user identity features while maintaining data privacy. To this end, the

proposed methodology aims to improve *interoperability* across multiple domains while also simplifying identity verification management in a privacy-preserving manner by utilizing high-level identity verification policies expressed in terms of identity attributes, zero-knowledge proof protocols, semantic matching techniques, and employing decentralized secure data stores. The critical factor of our proposal is represented by well-understood SOLID PODs.

Future work is mainly oriented towards enriching our framework with innovative features of privacy-preserving identity management over multi-Clouds (e.g., (Chaudhary & Kalra, 2019; Cui *et al.*, 2019; Raja *et al.*, 2021)), and improving the integration with *big data methodologies*, which span even-heterogenous domains (e.g., (Langone *et al.*, 2020; Morris *et al.*, 2018)).

ACKNOWLEDGEMENTS

This work was partially supported by project SERICS (PE0000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

REFERENCES

- Angin, P., Bhargava, B., Ranchal, R., Singh, N., Linderman, M., Othmane, L.B., & Lilien, L. (2010). An Entity-centric Approach for Privacy and Identity Management in Cloud Computing. In: *SRDS 2023, 29th IEEE Symposium on Reliable Distributed Systems*.
- Bertino, E., Paci, F., Ferrini, R., & Shang, N. (2009). Privacy-Preserving Digital Identity Management for Cloud Computing. *IEEE Data Engineering Bulletin* 32(1), pp. 21–27.
- Chaudhary, T., & Kalra, S. (2019). Interoperable identity management protocol for multi-cloud platform. *International Journal of Big Data Intelligence* 6(2), pp. 69–85.
- Cui, J., Zhang, X., Zhong, H., Zhang, J., & Liu, L. (2019). Extensible conditional privacy protection authentication scheme for secure vehicular networks in a multi-cloud environment. *IEEE Transactions on Information Forensics and Security* 15, pp. 1654–1667.
- Deng, M. (2010). Privacy Preserving Content Protection (Privacy Behoud Content Protection). *Faculty of Engineering Katholieke Universiteit Leuven, Leuven, Belgium*.
- Goldreich, O., & Krawczyk, H. (1996). On the composition of zero-knowledge proof systems. *SIAM Journal on Computing* 25(1), pp. 169–192.

- Kumar, P., Sehgal, V.K., Chauhan, D.S., Gupta, P., & Diwakar, M. (2011). Effective Ways of Secure, Private and Trusted Cloud Computing. *arXiv preprint arXiv:1111.3165*.
- Langone, R., Cuzzocrea, A., & Skantzos, N. (2020). Interpretable Anomaly Prediction: Predicting Anomalous Behavior in Industry 4.0 Settings via Regularized Logistic Regression Tools. *Data & Knowledge Engineering 130*, art. 101850.
- Miller, G.A. (1995). WordNet: a lexical database for English. *Communications of the ACM 38(11)*, pp. 39–41.
- Morris, K.J., Egan, S.D., Linsangan, J.L., Leung, C.K., Cuzzocrea, A., & Hoi, C.S.H. (2018). Token-Based Adaptive Time-Series Prediction by Ensembling Linear and Non-Linear Estimators: A Machine Learning Approach for Predictive Analytics on big Stock Data. In: *ICMLA 2018, 17th IEEE International Conference on Machine Learning and Applications*, pp. 1486–1491
- Pawar, P.S., Sajjad, A., Dimitrakos, T., & Chadwick, D.W. (2015). Security-as-a-service in multi-cloud and federated cloud environments. In: *IFIPTM 2015, Trust Management IX, 9th IFIP WG 11.11 International Conference*, pp. 251–261.
- Raja, S.K.S., Sathya, A., Karthikeyan, S., & Janane, T. (2021). Multi cloud-based secure privacy preservation of hospital data in cloud computing. *International Journal of Cloud Computing 10(1-2)*, pp. 101–111.
- SOLID (2024). <https://solidproject.org/>
- Spyra, G., Buchanan, W.J., & Ekonomou, E. (2016). Sticky Policy Enabled Authenticated OOXML. In: *SAI 2016, SAI Computing Conference*.
- Tari, Z. (2014). Security and Privacy in Cloud Computing. *IEEE Cloud Computing. 1(1)*, pp. 54–57.
- Weingärtner, R., & Westphall, C.M. (2014). Enhancing Privacy on Identity Providers. In: *SECURWARE 2014, 8th International Conference on Emerging Security Information Systems and Technologies*.