User-side Personalization Considering Privacy Preserving in Cloud Systems

Leila Sharifi  Maryam Heidari Beisafar
Department of Computer Engineering & Information Technology
Urmia University of Technology
Urmia, Iran
{l.sharifi, m.heidari}@uut.ac.ir

Abstract—Cloud systems are in the list of the most recent technologies. It is expected that users move to these new systems rapidly; however, privacy issues make users doubtful to migrate to cloud. They need to give a guarantee that their data and processes are protected enough within the distributed cloud infrastructure; on the other hand, to get better and more special services they have to permit the hosts to access their personal data for personalization purposes. To solve this controversy, in this paper we introduce a new personalization framework and architecture conveys privacy issues. Our proposed architecture mostly involves user side processing and no personal data leaves the client. In this way, the users feel more comfortable to use the cloud system while the quality of service is promoted as well. This method embraces personal data processing agent in the client side through personalization techniques and queries are sent to the hosts in an anonymous format. What is more, our personalization system needs only light processing, able to be run on clients with minimum processing capability, one of the most important issues for cloud users. Another advantage of the proposed architecture is its technology independency that matches the modules with various implementation techniques. We explain our framework and architecture in a “location advising service case study.”

Keywords: user-side personalization; privacy preserving; anonymity; service oriented architecture; location advisor.

I. INTRODUCTION

Cloud computing has brought up many benefits for organizations such as efficiency and effectiveness. By using Cloud computing, the traditional corporate data centers have been changed and systems have been moved to outsourced data centers, either in a private, public or hybrid environment.

The diverse range of cloud systems’ potential makes it a really optimal solution for business issues. The cloud computing is able to provide the hardware and software wrapped as services; therefore, business runners can purchase the services on demand. These services which run on cloud give users access to their information anywhere and by lower costs that eventuates to more productivity for IT businesses. According to a study from IDC1, one of the top 5 challenges in the cloud services is personalization. The most important drill is to customize services according to user profiles and to develop new personalization solutions for cloud providers, subsequently.

Collecting and using personal data in order to achieve personalized service is a rudimentary issue in the improvement of result adjusting with user demands and promoting the quality of service (QoS) as a result. However, this matter is in the controversy with privacy and user data protection while it needs to disclose user personal data with the service provider and relative hosts. The issue is more frustrating when we choose cloud platform for the services, because we are not aware of the hosts providing the service for us and the identity of those hosts permitted to access to our data is unknown for us, as a result [1].

Privacy is tightly related to personalization solutions in cloud environment. Several ideas have been published to protect user’s privacy [2]. It is more beneficial that the privacy of data in cloud would be configurable by user. By enabling users to control and manage privacy mechanisms which are applied on their data and customize the privacy level of cloud services, their trust on the cloud and its services will grow and it will bring more benefits on IT businesses.

To this end, we propose an architecture and a framework for the cloud systems that provide user-side personalization service, and promotes privacy preserving in the system as well. In this approach, services are provided in a personalized manner; however, the useful data about user preferences and activities never leave the client machine. Thus, as long as the privacy is not violated, the users trust this service. The main contribution of this paper is an innovative cloud architecture and framework that offers personalized services while protecting user information via user-side personalization and data anonymization to confirm privacy. Our proposed architecture cares about light processing in the client side and conforms the implementation technique independency issue.

The rest of the paper is organized as follows: Section II presents the related work. Section III discusses proposed system architecture. We argued the proposed framework in section IV. We rely on a practical case study on location advisor service in section V and continue with an analytic comprehensive analysis in section VI. Finally, the paper is concluded in section VII.

II. RELATED WORK

Nowadays when users are choosing among different service providers personalization becomes critical. By the rising trend of cloud computing usage, the need for personalized services has increased. Some approaches are proposed in this area [3].

Cloud computing provides many services; One of the services that have growing trend is e-learning. Reference [4] proposes a cloud personalized learning environment. Also
recommending these services to appropriate users is a state of the art research topic. [5] supports a personalized service recommendation model in web log mining that uses cloud principles. [6] explores a context aware recommendation service which utilizes the architectural features of the cloud in order to deliver intelligent personalized services. In [7] a personalized service authentication system is proposed which is used to recognize queries of unauthorized user and limits services in cloud.

Personalization encounters some drawbacks. As aforementioned, personalization includes the collection of user’s profile and uses it to see user demands and preferences. When it comes to user personal information and preferences, privacy concerns should be mentioned. Data are transmitted to cloud through public networks; therefore, privacy issues should be considered to keeping personal private information safe from cloud provider. An Effective Privacy Protection Scheme (EPPS) is proposed in [8] that protects privacy and decreases performance consumption.

Another way to preserve data privacy is providing privacy as a service. [9] presents a model named PaaS (Privacy as a Service). PaaS provides the secure storage and processing of user data by means of cryptographic coprocessors.

There are many solutions for enhancing privacy; one of them is to leverage it as a tool like what [10] suggests. The new technologies can be applied to protect private data such as anonymity. [11] examines a new approach based on private matching and min-attribute generalization.

III. THE ARCHITECTURE

In this section we introduce our system architecture. The most important property of this architecture is its simultaneous concern on personalization, privacy and security. Here, we do rely on personalized service provisioning with a personalization module in the user middleware that offers more accurate and specific services for an individual user owing to getting feedback from user data instead of a bunch of users. Moreover, by placing the personalization module in the user middleware, privacy is mostly obtained on account of least user information disclosure.

The restricted user information is disclosed with the service provider in an anonymous format that is not traceable to find the user original information. All of the above matters lead to an accurate, user satisfying service that considers privacy as well. Therefore, the quality of service is enhanced in the cloud environment and more users are encouraged to use these environments as a bed for their required services.

Nonetheless, the proposed architecture does not violate the thin-client as a critical issue in the cloud systems [12,13]. We achieve thin-client goal via light computing as well as using cloud storage resources to store the user information in an encrypted format. We encrypt the user data for security reasons that guaranty the least disclosure of user information in the system.

Another striking feature of the proposed architecture is technology independency that makes the framework free of implementation technique requirements. In other words, for each module we are able to choose different implementation techniques regardless of the other modules implementation matters. This design makes our architecture flexible enough to match with different implementation techniques, which is the main goal of cloud systems.

Our proposed architecture takes the advantage of layered Service Oriented Architecture (SOA) [14] and modular design in each layer. It includes five different layers as illustrated in Figure 1.

1. User layer that runs cloud applications for each individual user.
2. User middleware that mainly involves in user information collection, anonymization and management as well as personalization of services for the user.

![Figure 1. The System Architecture](image-url)
3. Service Provider MiddleWare (SPMW) generally supports the user middleware through providing the user with service parameters via Universal Description Discovery and Integration (UDDI) and anonymous data provisioning for user service by anonymity analyzer.

4. Virtualization layer in this cloud architecture is in charge of managing the virtual machines to devote and manage cloud resources for the cloud services.

5. Cloud resource layer embrace all the cloud resources and infrastructures, e.g. data centers and processors, required to support the cloud services in the system.

In this context, we refer to the user or client as the person demanding the service and his/ her connection device; host as the system that serving the service, and the service provider as a broker among user and host. The rest of this section expresses each module’s responsibilities and mechanism in more detail.

A. User Information Collector(UIC)

This module manages all the user personal data in the user middleware layer. The idea behind forming this module is making a centralized data management engine in the user middleware to store and retrieve all the user personal information and provide other modules with appropriate data. With a centralized data management module, we alleviate the burden of data management for other modules in the user middleware and promote the thin-client idea.

UIC stores and retrieves all the user personal information, its interaction history with cloud as well as feedback and query results in the cloud environment in an encrypted format [8,15] to prevent user information disclosure in the cloud environment.

As it is obvious in Figure 1, UIC is a bridge among user applications, personalizer and anonymizer modules. It retrieves and transits all the user data and queries exchanged among these modules and makes a copy of them on the secure storage.

B. Personalizer

In User middleware layer, this module mainly contains three sub-modules as demonstrated in Figure 1.

- Personalization engine is responsible for personalizing the user requests as well as query results using a combination of anonymous user personal information and other users’ anonymous information provided with UIC. This module chooses best results from query responses for user applications, using user personal data.

- System modeler provides the service parameters for each individual user application and queries via connecting to the UDDI module in the SPMW.

- Feedback generator obtains feedback for the personalization engine and system modeler on demand. This module manages the user feedback regarding the received service such as service rate, user satisfaction, quality of service, etc.

C. Anonymizer

Anonymizer module is in charge of making the information sending to service provider layer anonymous. To preserve the privacy of personal user data, in this architecture all the data that are required to disclose with service provider are anonymous in the user middleware. Here we apply anonymization process to user personal data with the support of anonymity analyzer in the SPMW. The anonymization process in this module is very flexible and is done in different levels as [16,17] and supports diverse implementation techniques such as [16,18,19,20]. The anonymous data are basically generalized in three different levels according to the user request and feedback regarding generalization level.

D. Anonymity Analyser

The main contribution of this module is management of the anonymous data obtained from different users connecting to the service provider in service provider middleware layer as represented in Figure 1. The idea of this kind of anonymization stems from [11]. This module mostly stores the anonymous data received from users for future request anonymization. This means that when a user intends to anonymize its own requests in anonymizer module of user middleware, it asks this module to obtain him/her with a set of anonymous data. Therefore, this module retrieves the most appropriate tuples for that individual request from the anonymous data stored in the cloud and sends them to the user side anonymization module. This process is discussed in more details in section IV.

IV. PROPOSED FRAMEWORK

In this section, we introduce the proposed framework. Our method is explained here in a step by step manner, i.e. in eight steps, according to the Service oriented Architecture (SOA) reference model & Standard [14].

A. Call for Service Info

When a user invokes a service, UIC sends the query to find general information about the service such as input parameters, general definition and related information. This provides more modifiability for services and if any changes occurred in the service definition or implementation, UIC is able to update his service history [21].

B. Call response

The call response message contains Web Service definition language (WSDL) and other related information according to requested service [14].

C. Call for anonymous data

Service provider stores anonymous data of different users for each service such as the users’ activities, rates, etc. as discussed in section III. Anonymous dataset includes the data collected from clients that are anonymized in the user-side with an anonymization algorithm, e.g. [16,17]. Personalizer module requires some useful data about other users’ activities for the personalization purpose. These data are quite anonymous and user machine cannot obtain any identity from them.

D. Extracting Useful parameters

Cloud provider should find appropriate data for the user based on selected services. [22] introduces a new method to solve this
storage for the service. As the next step, some filtering the cloud storage and only the user is able to retrieve them [23]. All the data saved and managed in an encrypted manner within activities, rate of each place, preferences, privacy concerns, etc. the host. In this step, UIC module gathers data regarding user implementation constraints, e.g. client side computing capacity. UIC in this step. The threshold N may be determined based on each record according to (2).

\[
\text{RI} = g_1 \times \text{W_i} \tag{1}
\]

\[
\text{RI} = \sum \text{W_i} \tag{2}
\]

The less the RI is, the lower generalization level for essential parameters’ records would be.

E. Call for data response
Top N tuples with smallest RIs are selected for sending to UIC in this step. The threshold N may be determined based on implementation constraints, e.g. client side computing capacity.

F. User Data Retrieval
As discussed in section III no personal data is revealed for the host. In this step, UIC module gathers data regarding user activities, rate of each place, preferences, privacy concerns, etc. All the data saved and managed in an encrypted manner within the cloud storage and only the user is able to retrieve them [23].

G. Personalization Process
The UIC module receives the anonymous data from previous step, decrypts and extracts them from the cloud storage for the service. As the next step, some filtering techniques are needed to personalize the service, such as collaborative filtering or any other machine learning techniques, to retrieve the personalized results for the user [24]. There is a trade-off between personalization accuracy and privacy. In this approach we try to provide the personalized services with privacy protection. Personalization accuracy is sacrificed with using anonymous data to provide privacy [25].

H. Anonymous feedback
To get better performance, anonymity analyzer collects the user feedback and other data according to personalized service in this step. Afterwards, a copy of these data should be sent to the service provider based on privacy level. To anonymize data different algorithms may be applied [16,18,19,20]. The algorithm blends the anonymous data got in previous steps with other users’ anonymous data available in the service provider side to make the user data tracking impossible.

V. LOCATION ADVISOR: AS A CASE STUDY

We consider a case that location advisor on Google Glass makes an effort to provide the best offer for buying something to its user [26]. The best offer should not only consider the closest place, but also the individual user preferences, previous activities, other users’ rate, etc. However, the main challenge is how to personalize the service for each individual, while protecting privacy. We present UUT (Urmia University of Technology) location adviser system "UUTLA" to solve the challenge as a case study.

In this section, the proposed approach is investigated as a case study in four steps. A) Call for service data and response, B) Call for anonymous data and response, C) Personalization process, and D) Anonymous feedback.

A. Call for service data and response
When a user tries to find the best offer for a place, first of all, he/she should call the related service. In SOA, all the information about the services stores in UDDI by using Web Service Definition Language (WSDL). Client asks for the service to advise him/her to find the appropriate place by sending SOAP (Service Oriented Architecture Protocol) message to UDDI [14]. UDDI responses with sending location adviser service WSDL and other related information such as the related map or other data according to the service.

B. Call for anonymous data and response
When the user machine receives the WSDL of the service, user information collector sends the request for anonymous data to anonymity analyzer according to other users’ data such as rate for each place to invoke the UUTLA service. Anonymity analyzer should find which parameters are useful for location advising based on generalization level and correlation of the parameters. As discussed in section IV, the best anonymous records are selected to send to UIC. In UULA we consider one parameter, rate of the users; the correlation of the parameter for selected service and G is very essential for the personalization of the level of the generalization (anonymization) of that parameter. C_i is a variable with values 0 and 1. C_i=0 denotes parameter i is very essential for the personalization of the service. C_i=1 denotes parameter i less impacts the service personalization. The default situation of the system is C_i=0; however, the service modeler module can change the correlation in step three.

\[
W_i = C_i \times G_i
\]

Cloud provider should compute Record Impact (RI) for each record according to (2).

\[
\text{RI} = \sum W_i
\]

Table rows show the record id # and the columns show the places offered to the user. Each cell is filled by the rates in the

<table>
<thead>
<tr>
<th>Record ID #</th>
<th>Place 1</th>
<th>Place 2</th>
<th>Place 3</th>
<th>Place n</th>
<th>G Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>R102</td>
<td>*</td>
<td>4</td>
<td>-</td>
<td>...</td>
<td>2</td>
</tr>
<tr>
<td>R755</td>
<td>*</td>
<td>-</td>
<td>5</td>
<td>...</td>
<td>2</td>
</tr>
<tr>
<td>R1022</td>
<td>3</td>
<td>5</td>
<td>*</td>
<td>...</td>
<td>3</td>
</tr>
<tr>
<td>R300</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>...</td>
<td>5</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
range of 0 to 5 assigned to the place by the user. “+” indicates that the related user never vote this place and the sign “-” confirms that this value are anonymous by user. As mentioned before, the number of the top selected records depends on implementation constraints.

C. Personalization Process

One of the great advantages of the proposed framework is technology independence. This provides the vast set of personalization techniques can be applied. Collaborative filtering (CF) is a common technique for many personalized services. CF is an automated means for ranking content “based on the premise that people looking for information should be able to make use of what others have already found and evaluated” [27]. The fundamental assumption in CF is that two users rate or act on various items similarly as long as they rank n previous items equally, or have similar behaviors (e.g. buying, watching, listening) [24]. In UUTLA, CF uses a database of rates by user gathered from different users for each place to recommend new place to the user applying the service. Then all the information about the users previous activities and places rates retrieved from the secure storage. These data consist of the rates user assigned to the places previously. To use Item base CF, similarity between the places should be specified. In this case, similarity sim(i, j) between two places i, j is measured by computing Pearson correlation [28]. For item based CF, the set of users u U who rate both i and j, the Pearson correlation is calculated based on (3).

\[
sim(i,j) = \frac{\sum_{u \in U} (R_{ui} - \bar{R}_i)(R_{uj} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{ui} - \bar{R}_i)^2} \times \sqrt{\sum_{u \in U} (R_{uj} - \bar{R}_j)^2}}
\]  

(3)

Where \( R_{ui} \) is the rate of place i assigned by user u, and \( \bar{R} \) is the average rate of the ith place in the user set. The most important step in a collaborative filtering system is to generate the output interface in terms of prediction. Once we isolate the set of most similar items based on the similarity measures, the next step is to look into the target users’ ratings and use a technique to obtain predictions. In this paper, weighted sum method is applied. As the name implies, this method computes the prediction on an item i for a user u by computing the sum of the rates given by the user on the items similar to i. Each ratings is weighted by the corresponding similarity \( s_{ij} \) between items i and j as revealed in (4) [28].

\[
r_{ui} = \frac{\sum_{allSimilarItems} (S_{i,N} \times r_{ui,N})}{\sum_{allSimilarItems} S_{i,N}}
\]  

(4)

First we take all the items similar to our target item, and from those similar items, we pick items which the active user has rated. We weight the user rates or each of these items by the similarity between it and the target item.

In (4) the summations are over all the items \( n \) N rated by user u so far; \( S_{i,N} \) is the weight between items i and n, and \( r_{ui,N} \) is the rate from user u on item n. Basically, this approach tries to capture how the active user rates the similar items. The weighted sum is scaled by the sum of the similarity terms to make sure the prediction is within the predefined range. Personalizer selects top \( N \) places (in this case \( N=3 \)) with the highest \( P_{ui} \).

D. Anonymous feedback

As discussed before, anonymity analyzer should be updated after each personalization process. All the data should be stored on secure storage non anonymization but encrypted after personalization. Conversely, these data are stored in the anonymity analyzer unencrypted but anonymous at the same time.

VI. COMPREHENSIVE ANALYSIS AND COMPARISON

In this section we compare the proposed architecture with the pioneer state of the art ones in a various range of critical features required for an inclusive cloud system that is able to obtain diverse range of user satisfaction and provides a wide-range of services with elevating quality. These features include privacy, personalization, security, storing method, client or cloud side processing and thin-client. Here personalization means providing services according to user preferences and requirements. We consider privacy as protecting user personal information against disclosure and security is defined as preserving the exchanging data from the leakage. In storing method we mostly study if the framework proposes any encryption method, anonymization, or uses an engine for storing objective. Thin-client is one of the most important issues in the cloud systems, aims at alleviating the burden of computing and storing in the client side. Increasingly, in this line, we study if the main processes and storages are in the client side or the cloud side. As it is obvious in TABLE II, most of the proposed frameworks which consider personalization do not preserve privacy, but the framework in [3]. The personalization framework in [5,4] are cloud side and consider thin-client issue, but do not care about the privacy and security of user information. Although, frameworks [6,7] consider data security, they do not have any privacy preserving mechanism. In [3] all the user personal information is stored in the user-side to preserve privacy. This method violates the thin-client issue while many user devices are not capable of storing this amount of data due to the storage limits. Proposed framework uses secure cloud storages to store user data. What is more, the data exchanged with the service provider are anonymous and occasionally traceable to get user original data.

<table>
<thead>
<tr>
<th>Framework</th>
<th>Privacy</th>
<th>Personalization</th>
<th>Storing Method</th>
<th>Thin-client</th>
<th>Client/Cloud side</th>
<th>Security</th>
</tr>
</thead>
<tbody>
<tr>
<td>hang Guo et al. [3]</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>Client</td>
<td></td>
</tr>
<tr>
<td>J.chang bin [5]</td>
<td>√</td>
<td></td>
<td></td>
<td>√</td>
<td>Cloud</td>
<td></td>
</tr>
<tr>
<td>Soo hyun kin, et al. [7]</td>
<td>√</td>
<td></td>
<td>Encrypted</td>
<td>√</td>
<td>Cloud</td>
<td></td>
</tr>
<tr>
<td>Our Proposed Architecture</td>
<td>√</td>
<td></td>
<td>Anonymized</td>
<td>√</td>
<td>Client</td>
<td>√</td>
</tr>
</tbody>
</table>

TABLE II. COMPARISON OF THE STATE OF THE ART FRAMEWORKS
VII. CONCLUSION

Personalization and privacy are two significant issues in the cloud systems, since they lead to QoS promotion in these systems. Up to now, cookies, which are installed on the web browsers, collect the data for servers. They process the incoming data and provide the personalized services. Although this method works properly for personalization purposes, it may violate privacy issues. This may be much worse when the service bed is a cloud system and the host is not visible for end user. Consequently, it is very difficult for a user to trust this unknown host and provide it with his personal data. The previous works in this area are not accurate enough and do not result in an effective applicable solution. To surmount this problem, in this paper, we propose a personalization method that makes almost all the processes in the client side; i.e. no personal data are revealed for the host and privacy is preserved as a result.

Our studies reveal that the proposed architecture and framework outperforms the state of the art methods, since it provides personalized services and preserves the privacy simultaneously. Proposed framework and architecture not only promotes the QoS, but also considers the thin-client issues. Increasingly, the proposed architecture is technology independent and supports various combinations of implementation techniques for each module. As a future work in this line, we intend to evaluate our framework and architecture via simulation and an analytic model.

REFERENCES
