Privacy-Preserved and Best-effort Provisions of Cyber-I Information to Personalized Services

Liu, Wenjing

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
11
(開始ページ / Start Page)
1
(終了ページ / End Page)
6
(発行年 / Year)
2016-03-24
(URL)
https://doi.org/10.15002/00012869

Privacy-Preserved and Best-effort Provisions of Cyber-I Information to Personalized Services

Wenjing Liu
Graduate School of Computer and Information Sciences
Hosei University
Tokyo 184-8584, Japan
wenjing.liu.yd@stu.hosei.ac.jp

***Abstract*—User information is needed for personalized services. However, personalized services are often confused either by information inconsistency and incorrectness or the cold start issue while they gather user information. Even some services, such as social network services, provide user information for personalized services to overcome these problems, they are also facing difficulty of the limited diversity of user information. Additionally, they are only capable of providing existing information. While a Cyber-I (short for Cyber Individual) collects any information of a person in its way to gradually approximate to its user. Certainly, a user information needed by personalized services is also included in a corresponding Cyber-I. Therefore, providing Cyber-I information to personalized services could be more prospecting. In order to provide Cyber-I information to personalized services, there are two main problems should be solved. One is the privacy protection problem. Methods should be designed to provide privacy preservation for user. Another one is the best-effort issue which is about how to make full use of existing Cyber-I information to satisfy personalized services as much as possible. Thus, the goal of this paper is providing Cyber-I information to personalized services by best-efforts provisions, simultaneously provide privacy preservation for Cyber-I. To reach that goal, a Cyber-I Information Provision System (CIPS) is proposed.**

Keywords—Cyber-I; Real-I; personalized service; privacy preservation; information provision

I. INTRODUCTION

With rapid development of the Internet, the amount of data is overloading, which makes personalized services be used to address the overloading problem. Personalized services provide customized services by building, managing, and representing information [1]. However, all of personalized services need user information at first to provide services. There are two approaches to get information [2]. One is the explicit approach by explicitly asking help from users. Another is the implicit approach by gathering information without any efforts from users. Additionally, hybrid techniques of the two approaches are also used. However, the explicit approach has the inconsistent and incorrect problem, as it is influenced by the mood and state of users greatly. The implicit approach also has the cold start problem. Because it is short of information when it is just built. Even the cold start issue could be solved by getting information from other services, it still has to solve limited diversity of user information. Because a certain service only has limited types of information. Though it could solve the limited diversity of user information in a degree by gathering information from more than one services, it has to handle heterogeneous information from different services.

A Cyber-Individual, short for Cyber-I, is the counterpart of a Real Individual (which we refer to as a Real-I) in the physical world [3]. The accuracy of the approximation will be continuously improved by collecting and utilizing increasing personal data, which is any information related to a person. Therefore, a Cyber-I holds a large amount of and various kinds of an individual information, which makes personalized services could get lots of benefits and overcome problems existing in traditional approaches if they gather information directly from a Cyber-I. However, in order to provide Cyber-I information to personalized services, there are two problems should be solved. One is the privacy problem. Providing information for other services may incur people worry about their privacy. Another problem is the way to make best effort to provide Cyber-I information for satisfying personalized services. To solve these two problems, this paper proposed a Cyber-I Information Provision System (CIPS) for providing Cyber-I information in a privacy-preserved and best-effort way.

The remainder of this paper is organized as follows. Section II describes related work about information gathering approaches used by personalized services, privacy protection techniques and information provision services. Section III describes system architecture of CIPS. System communication including request format, response format and message transceiver is introduced in Section IV. After that, in Section V, we describe request analysis and response analysis. Privacy preservation and best-effort provisions of Cyber-I information are discussed in detail in Section VI and Section VII, respectively. System prototype and some cases are presented in Section VIII. At the last, Section IX concludes this paper and outline future research.

II. RELATED WORK

The rapid growth of information makes digital explosions [4]. And the information is a valuable resource [1], especially for the personalized services that need personal information to provide user-aware services. Generally, personalized services gather information in two ways: an explicit approach and an implicit approach [2].

Many personalized services use the explicit approach, such as, using personal information questionnaires. An example of the explicit approach is WIFS [5], which allows users to set the

---

Supervisor: Prof. Jianhua Ma

searching and filtering modalities and input the query retrieved to computer science topics. However, the information gathered by the explicit approach is be inconsistent or incorrect, as it is influenced by the mood and state of users greatly. One typical instance of the implicit approach is tracking search submitted queries and clicked results of users' browsers to gather search history [6]. Even there is no inconsistence and incorrectness, the cold start problem is yet solved. If getting information from other services, it also has limited diversity of information and heterogeneity data issues.

Many researches about privacy protection have been conducted. One way is allowing users to specify their privacy requirements by filling some forms with pre-defined options [7]. An interface is provided for users by PCAP [8] to define their privacy preferences. Even this mechanism offers users privacy protection, their settings are all done by users manually and it would not be flexible enough to appeal to users' real privacy demands, for a person's willingness to share personal information may change frequently.

To overcome above problems, a Cyber-I Information Provision System (CIPS) is designed and implemented to provide information of Cyber-I to personalized services. As Cyber-I information is accumulated from various services including SNS, web searching, electronic devices etc., the inconsistent and incorrect problem can be overcome. Additionally, the large amount of Cyber-I information can solve the cold start problem. Further, the limitation of information diversity is solved by the various information of Cyber-I. Besides, CIPS tries it best to provide Cyber-I information for satisfying personalized services' requests even requested information is not available directly. Considering privacy in view of both system and Real-Is, CIPS also provides privacy preservation which holds system level privacy preservation and Cyber-I level privacy preservation.

III. System Architecture

The ultimate goal of the Cyber-I Information Provision System (CIPS) is to provide precise and complete personal information to personalized services based on a Cyber-I in a privacy protected way. To meet the goal, the architecture of the whole system is introduced in this section.

An overview architecture of CIPS is shown in Fig. 1. Personalized services are usually developed by second-party or third-party, such as a travel recommendation system. The second-party includes developers and partners of Cyber-I Open System Platform (COSP) [9]. Relatively, the third-party means developers or service providers out of COSP. The personal information of a real individual is stored in its Cyber-I DB. Generally, Cyber-I DB divides Cyber-I information into basic data, raw data and model data. Basic data are basic personal information, such as, name, birth date, nationality, etc.; raw data are personal information that collected by COSP, like post of a twitter user; model data are personal information that computed by the COPS. Additionally, settings of Cyber-I DB stores setting information about Cyber-I, such as privacy settings. While System Info DB stores information of personalized services and privacy setting information of CIPS. When a personalized service needs personal data of a Cyber-I, a request with the personalized service own information and demands is sent to the CIPS. In the CIPS, the request is received by message transceiver, which passes the request to the request analysis that figures out the demands of the request and validates the requester's identity. After that, privacy provision provides privacy protection and response provision tries its best to get requested Cyber-I information. Finally, a response is formed by the response formation and sent by message transceiver to the personalized service.

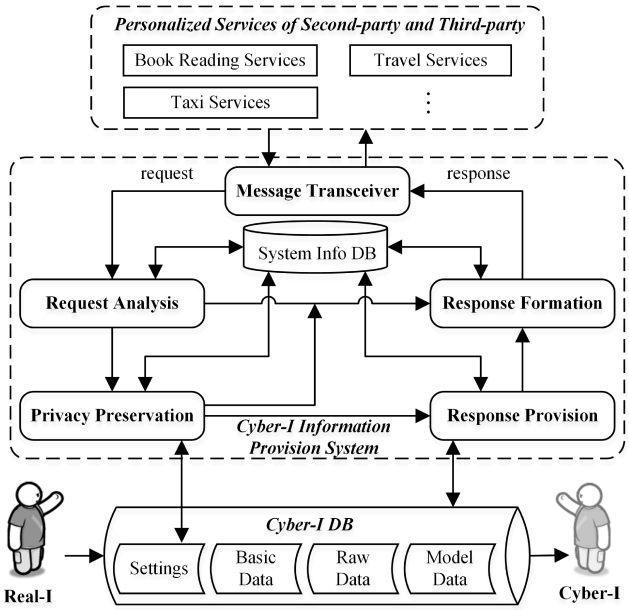

Fig. 1. The system architecture

The system is implemented in a REST style web service. To ease the implementation of the REST style system, Restlet frame is adopted. Furthermore, Tomcat is used as the server and Mongo DB is adopted as the database for the persistence of data since it usually gains higher performance in dealing with the big personal data than the conventional RDBMS.

IV. System Communication

It is necessary to make CIPS and personalized services to achieve a consensus on messages sent between them so as they can understand each other. On account of that, this section will successively introduce request format and response format, and message transceiver.

The request format is shown in Fig. 2. The requester identity is composed of ID field and Key field. The OpenID field is filled by the requested Cyber-I's open identity instead of Cyber-I's identity considering privacy protection. Content field and Description field make up request descriptors to express demands of requesters. To balance effectiveness and flexibility of requests, three categories, i.e., non-open request, semi-open request and full-open request, of requests are provided. The request descriptors of the non-open request are set by the CIPS. Conversely, that of full-open request are filled by requesters themselves according to their own ideas. While the semi-open request allows CIPS set the Content field of the requests descriptors and allow personalized services to fill the Description field freely. In order to balance the demands of personalized services and the reduction of unnecessary waste

resources, a Mode field is necessary in the request format to indicate what kind of efforts that personalized services want the CIPS to make.

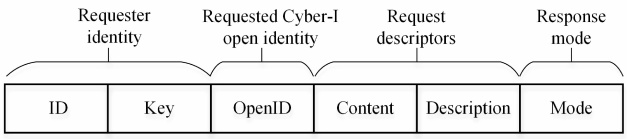

Fig. 2. The request format

The response format is shown in Fig.3. The responder identity and requested Cyber-I's identity are put in the ID field and OpenID field respectively. The response status is described by Code field and Description field. And response data are put in the Data field. Generally, responses are divided into two categories: correct responses and error responses. A correct response is made after the response provision obtains or computes data from Cyber-I DB, while an error response is made because it does not go through the request analysis or privacy preservation.

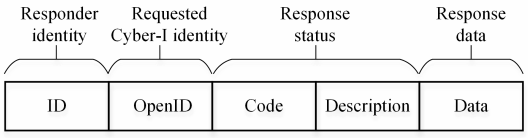

Fig. 3. The response format

Requests are received by the message transceiver and responses are also sent by the message transceiver in the CIPS. For simplicity, message transceiver is implemented by using the connector component of the Restlet which is a REST frame. A connector in the Restlet is a software element that manages network communication typically by implementing a network protocol. HTTP is used as the network protocol to transmit request messages and response messages. All Cyber-I information could be given URIs.

V. Request Analysis and Response Formation

In order to provide personal information of a Cyber-I to a personalized service, the first thing CIPS needs to do is figuring out demands of personalized services. Then, after handing the demands, CIPS should organize the response in the response format and send out. Therefore, request analysis will be described in this section, and so does response formation.

A. Request Analysis

The process of the request analysis is shown in Fig. 4. When getting a request passed by the message transceiver, the field extractor splits the request into fields. Then, the access validator will verify the requester by comparing the ID and Key with the ID and Key which are assigned by the CIPS when they registered successfully and stored in the System Info DB. If the validation fails, it will set the mark variable Tag, equals 0 and directly go to the response formation module. Otherwise, the Cyber-I retrieval firstly maps the OpenID to the corresponding Cyber-I's identity (Cyber-IID) by consulting the configuration of System Info DB, then searches the Cyber-Is DB. If the Cyber-IID exists in the Cyber-Is DB, the privacy preservation will be the next step. Otherwise, the Cyber-I retrieval will set the Tag equal 1 and go to the response formation module.

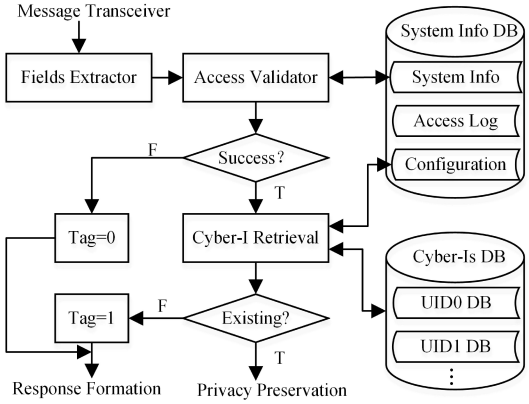

Fig. 4. The request analysis

System Info DB records three kinds of information: configuration, access log and system info. Configuration information is about the CIPS's privacy configuration. History requests and their responses are recorded in the access log. When a personalized service registers to the CIPS, the CIPS will assign an ID and a key to it. They stores in system info with other basic information, such as service name, privacy level, etc. As said in Ma's paper [9], the COPS is a place for holding a possibly large number of Cyber-Is, every Cyber-I exists as an independent entity. Hence, an individual database named as Cyber-I's unique id is assigned to each Cyber-I residing in the Cyber-Is DB.

B. Response Formation

When the response formation is triggered, it will firstly generate response status. Then the response formatter will format responder ID (CIPS's ID), response status and response data into a response message. Lastly, the request & response storage stores the request and its response in the access log of the System Info DB and the response will be handed to the message transceiver.

Response status in the response format is generated according to the Tag by the response status generator. As introduced before, the response status indicates if the response is an error response or a correct response. Currently, the value of the Tag is in {0, 1, 2, 3}. Tag equals 0 means access validation is failed. Therefore, the response status Code is set to "1" and Description is also set. Similarly, when Tag is 1 means CIPS does not find the requested Cyber-I. Thus, "1" is set to the Code While 2 is the value of Tag, it means the privacy provision does not get through, that is to say that privacy protection does not allow the personalized service get the requested data. When Tag equals 3, it means that CIPS uses the response provision after request analysis and privacy preservation, i.e., the response that will be generated is a correct response. At this time, Code is set to 0.

VI. Privacy Preservation

As so much data collected for building a comprehensive Cyber-I, privacy invasions will be greatly concerned about if such data is provided to personalized services of second-party and third-party. Therefore, providing privacy protection is a very important part of the CIPS. Furthermore, it is also necessary to

provide privacy protection according to people's own taste, considering different people have different privacy requirements. Hence, system privacy preservation and Cyber-I privacy preservation are provided by CIPS. Fig. 5 shows the privacy preservation functions of the CIPS, which are discussed below.

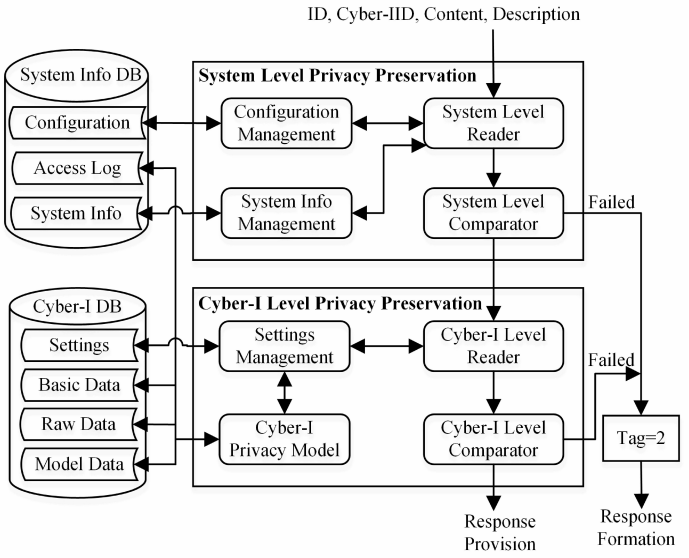

Fig. 5. The privacy preservation

A. System Level Privacy Preservation

As said previously, CIPS privacy preservation consists of system level privacy preservation and Cyber-I level privacy preservation, shown in the Fig. 5. When the request analysis passes ID, Cyber-IID, Content and Description to the privacy preservation, the system level privacy preservation firstly takes action. The system level reader uses the system info management to get the requester's privacy level and uses the configuration management to get the privacy level's corresponding accessible scope of information. After that, the system level comparator would figure out whether the request descriptors (Content and Description) are included in the scope. If the request descriptors are not in the scope, i.e., the system level comparator fails, Tag will be set to 2, and then response formation will be used. Otherwise, the Cyber-I level privacy preservation will come to work.

B. Cyber-I Level Privacy Preservation

Cyber-I level privacy preservation provides the second privacy protection for the CIPS. When the system level privacy preservation passes the privacy protection responsibility to the Cyber-I level privacy preservation, the Cyber-I level reader will firstly use the settings management to get the privacy setting information from the settings in Cyber-I DB. The privacy setting information includes the scope information that the requester could access. Then the Cyber-I level comparator will compare request descriptors with scope to figure out if the request descriptors are included in the scope of the privacy setting information. If the request descriptors are not in the scope, i.e., the Cyber-I level comparator fails, Tag will also be 2 and then go to the response formation. Otherwise, it will go to the response provision module.

The setting management is responsible for adding, deleting, updating, finding information in the settings of Cyber-I DB. While information stored in the settings of Cyber-I DB is generated in two ways. One is getting from the Cyber-I's owner, Real-I, who assigns the accessible information scope of a personalized service when the personalized tends to get authorization at the first time. The other is getting from Cyber-I privacy model that generates the privacy information based on the information in the basic data, raw data and model data in the Cyber-I DB and access log in the System Info DB. A representable Cyber-I privacy model is CIPM [10]. This kind of Cyber-I privacy model has two fold advantages: one is reflecting a user's privacy needs to different applications, and is adapting to a user's privacy demand changes. In the CIPS, the Cyber-I privacy model can be enriched in the future.

VII. BEST-EFFORT RESPONSE PROVISIONS

A. Overview of the Best-effort Response Provision

CIPS tries its best to provide Cyber-I information under considering requesters own desire, which is expressed in a request message's Mode field. When CIPS goes into the response provision module, the actions it takes are shown in the Fig. 6.

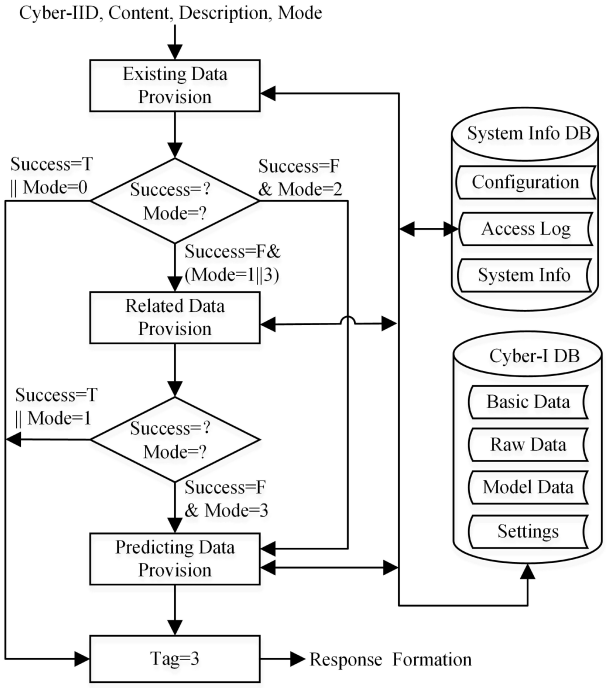

Fig. 6. The best-effort response provisions

The existing data provision firstly tries to find whether requested data has existed in the Cyber-I DB. If the existing data provision successfully finds the requested data, it will set Tag=3 and passes the data to response formation module. However, if it fails, there are three ways to go according the value of the Mode: (1) If Mode equals 0, it still sets the Tag=3 and goes to the response formation module. (2) If Mode is 1 or 3, the related data provision will try to find data which has some kind of association with the requested data. (3) If the value of Mode is 2, the predicting data provision will try to predict the requested data according to the data in the Cyber-I DB. If related data

provision is successful to get the data related to the requested data or the value of Mode is 1, it also will set Tag=3 and go to the response formation. Otherwise, if the value of Mode is 3, the predicting data provision will also be used to compute the requested data. Whether the predicting data provision gets the requested data or not, it still sets the Tag=3 and goes to the response formation.

B. Existing Data Provision

The existing data provision is the first step to find the requested data. One issue is that the request descriptors in semi-open requests and full-open requests filled according to personalized services' own wish may not be understandable for the CIPS. Thus, CIPS should also have methods to solve the problems. It is solved by creating synonyms thesaurus. A part of synonyms thesaurus is shown in the Fig. 7. CIPS stores 'hobby' as a key word to provide hobby information of a Cyber-I. If personalized services use the 'hobby' as request descriptors, it is easy for CIPS to provide service. However, personalized services may use 'hobbies', 'bent', 'favourite' and 'favourites', which are synonyms of 'hobby'. Therefore, by using synonyms thesaurus, the existing data provision first changes the descriptor into its synonyms.

	_id	descriptor	synonyms
1	ObjectId("5...	time	time
2	ObjectId("5...	period	time
3	ObjectId("5...	date	time
4	ObjectId("5...	hobby	hobby
5	ObjectId("5...	hobbies	hobby
6	ObjectId("5...	bent	hobby
7	ObjectId("5...	favourite	hobby
8	ObjectId("5...	favorites	hobby

Fig. 7. A part of synonyms thesaurus

C. Related Data Provision

When there is no requested data in the Cyber-I DB, according to personalized services (Mode=1 or 3), related data provision comes into effect. In general, there are various kinds of data in the Cyber-I DB. Therefore, sometimes, when a request comes into the CIPS, the requested data maybe not exist in the Cyber-I DB. But there may exist relevant data of the requested data. The relevant data may also be useful for personalized services. In the CIPS, we use data on information relations to provide related data. At the beginning the data relation is simply intuitive. An item is shown in Fig. 8. As shown, the related data of 'interest book' include 'education' and 'profession'.

```
/* 1 */
{
    "_id" : ObjectId("558bab102013faf52ededa36"),
    "original descriptor" : "interest book",
    "related data" : [
        "education",
        "profession"
    ]
}
```

Fig. 8. A item of data relation

D. Predicting Data Provision

When Cyber-I DB has no requested data, there are sometimes requesters are may not be able to use the related data to compute the requested data or they just do not want to use related data but want the data that computed by CIPS. In other words, requesters only need requested data no matter how the data comes. Therefore, CIPS could compute requested data according the Cyber-I DB or Cyber-Is DB. We name this effort as predicting data provision.

In the CIPS, currently, we use some kind of association rules. More concretely, we use dimensions as a rule to select Cyber-Is who have requested data and randomly select one Cyber-I from the result. Then its information will be used as the requested Cyber-I's predicting data. For example, to predicting interest books of a Cyber-I, we could use major information of the Cyber-I as a predicting dimension. Other Cyber-Is whose majors are same with the requested Cyber-I may have interest book information. So a simplest way is randomly selecting a Cyber-I who has the interest book information and has same major with the requested Cyber-I. Then use its interest book information as the requested Cyber-I's predicting data. Another way is using all interest book information of Cyber-Is whose majors are same with the requested Cyber-I. For the further selection, we use profession as a second dimension to choose a Cyber-I whose profession is same as the requested Cyber-I from the same major set. Then it randomly selects one Cyber-I and use its interest book information. The detail of an item is shown in Fig. 9.

```
/* 1 */
{
    "_id" : ObjectId("55912ed32c2e66015495521a"),
    "descriptor" : "interest book",
    "dimension" : [
        "major",
        "profession"
    ]
}
```

Fig. 9. A item of predicting data demention

VIII. System Prototype and Cases Study

The CIPS is composed of five parts, i.e., message transceiver, request analysis, privacy preservation, response provision and response formation. The interfaces include webpages about how to become a register, how to compose a request and send it, how to handle a response and so on. Further, based on CIPS, we developed a simple requester which sends requests to the CIPS for case study.

The first case of existing data provision is shown in Fig. 10. There are two requests in Fig. 10 (a). As shown in Fig. 7, the synonyms thesaurus of hobby include favorite. Hence, we use Request 2 whose description is 'favourite' as an example of synonyms thesaurus. The response of request 2 is similar to that of request 1, shown in Fig. 10 (b), except replacing 'hobby' field with 'favourtie'.

Related data provision uses the data relations described in Section VII to provide related data of requested data for requesters. As shown in Fig. 8, the related data of 'interest book' is 'education' and 'profession'. Therefore, when a request whose mode is "1", showed in Fig. 11 (a), comes, the related data provision will come into effect. Hence, the response for the request is formed, shown in Fig. 11 (b).

Request 1	http://localhost:8080/CIIPS/services?appid=appid00000&accesstoken=token00000&openid=Cyber_I00000&content=model&description=hobby&mode=0
Request 2	http://localhost:8080/CIIPS/services?appid=appid00000&accesstoken=token00000&openid=Cyber_I00000&content=model&description=favourite&mode=0

(a) Requests


```
{
  "id":"CIIPS",
  "openid":"Cyber_I00000",
  "code":"0",
  "description":",",
  "hobby" : [   "travel", "basketball" ]
}
```


(b) Response

Fig. 10. A case of existing data provision


```
http://localhost:8080/CIIPS/services?appid=appid00000&accesstoken=token11111&openid=Cyber_I00000&content=interest&description=book&mode=1
```


(a) Request


```
{
  "id":"CIIPS",
  "openid":"Cyber_I00000",
  "code":",",
  "description":"related data",
  "interest book" : [
          "education" :   {
                    "school_name" : "Dalian University of Technology",
                    "qulification" : "Master Degreee",
                    "graduate_time" : "2015-07",
                   "enter_time" : "2013-09"
                },
         "profession" :
                        {
                   "job_duty" : "clean tables",
                   "end_time" : "2014-11-09",
                   "job_title" : "cleaner",
                   "employer" : "Restruant",
                   "enter_time" : "2013-12-15"
              }
        ]
}
```


(b) Response

Fig. 11. A case of related data provision


```
http://localhost:8080/CIIPS/services?appid=appid00000&accesstoken=token11111&openid=Cyber_I00000&content=interest&description=book&mode=3
```


(a) Request


```
{
  "id":"CIIPS",
  "openid":"Cyber_I00000",
  "code":"0",
  "description":"predicting data",
   "interest book":[
     "computer networks",
     "software defined networks",
     "Operating System",
   ]
}
```


(b) Response

Fig. 12. A case of predicting data provision

As said in Section VII, currently, we only use the simple associated rules to predict data in CIPS. To show the function of predicting data provision, we use a request shown in Fig. 12 (a) to request 'interest book' information. The mode value of it is "3". The 'interest book' has two dimensional data. The first one is 'major' and the second one is 'profession'. Therefore, the response is predicted according this two information. The response for the request is shown in Fig. 12(b).

IX. CONCLUSION AND FUTURE WORK

In this paper, we prosecute our research on privacy-preserved and best-effort provisions of Cyber-I information to personalized services by constructing a CIPS system prototype. This system is capable of providing Cyber-I information by using (1) two level privacy protection, i.e., the system level and the Cyber-I level privacy preservations; (2) three best-effort provisions, i.e., existing data provision, related data provision and predicting data preservation. Message formats including request and response was defined to support communications between systems. And cases study was carried out to verify functions of the CIPS.

This research is only the first step of privacy-preserved and best-effort provisions of Cyber-I information to personalized services and it should be continued in several dimensions. First, system level privacy preservation should be more complete and more Cyber-I privacy models should be designed. Next, the best-effort provisions should be enriched. Data relations in related data provision should be more abundant and it is worthy to develop a generation method to automatically generate data relation. The predicting data dimension should also be enriched and other predicting data methods should be designed.

REFERENCES

[1] S. Gauch, M. Speretta, A. Chandramouli, et al, "User profiles for personalized information access," the adaptive web, Springer Berlin Heidelberg, pp. 54-89, 2007.

[2] M. R Ghorab., D. Zhou, A. O'Connor, et al. "Personalised Information Retrieval: survey and classification," User Modeling and User-Adapted Interaction,vol. 23, No. 4, pp. 381-443, 2013.

[3] J. Ma, J. Wen, R. Huang and B. Huang, "Cyber-Individual meets brain informatics," IEEE Intelligent Systems, vol. 26, No.5, pp. 30-37, September/October 2011.

[4] J. Ma, "Digital explosions and digital clones in cyber world," the IEEE Int'l Conf. on Intelligent Computing and Integrated Systems, Keynote Speech, 2011.

[5] A. Micarelli, F. Sciarrone, "Anatomy and empirical evaluation of an adaptive web-based information filtering system," User Modeling and User-Adapted Interaction, vol. 14, No. 2-3, pp. 159-200, 2004.

[6] S. Stamou, A. Ntoulas, "Search personalization through query and page topical analysis," User Modeling and User-Adapted Interaction, vol. 19, No. 1-2, pp. 5-33, 2009.

[7] O.R. Kurkovsky, J. Bhalodi, "Classification of privacy management techniques in pervasive computing," International Journal of u-and e-Service. Science and Technology, pp. 55-71, 2007.

[8] H. Chen, T. Finin, and A. Joshi, "An ontology for context-aware pervasive computing environments," The Knowledge Engineering Review, pp. 197-207, 2003.

[9] J. Ma and R. Huang, "Digital explosions and digital colones", in Proc. of the IEEE International Conference on Internet of People (IoP-2015), Beijing, 2015.

[10] L. Tang, J. Ma, R. Huang, et al, "Awareness and control of personal data Based on the Cyber-I privacy model", in Proc. of the 11th IEEE International Conference on Autonomic and Trusted Computing (ATC-2014), Bali, Indonesia, December 9-12, 2014.