Proceedings of the 2014 Federated Conference on Computer Science and Information Systems pp. 1397–1406
DOI: 10.15439/2014F129
ACSIS, Vol. 2

Danger Theory-based Privacy Protection Model for Social Networks

Nai-Wei Lo[1] and Alexander Yohan[2]
Department of Information Management
Nat'l Taiwan Univ. of Sci. & Tech.
Taipei 106, Taiwan
nwlo@cs.ntust.edu.tw [1]
m10109803@mail.ntust.edu.tw [2]

***Abstract*—Privacy protection issues in Social Networking Sites (SNS) usually raise from insufficient user privacy control mechanisms offered by service providers, unauthorized usage of user's data by SNS, and lack of appropriate privacy protection schemes for user's data at the SNS servers. In this paper, we propose a privacy protection model based on danger theory concept to provide automatic detection and blocking of sensitive user information revealed in social communications. By utilizing the dynamic adaptability feature of danger theory, we show how a privacy protection model for SNS users can be built with system effectiveness and reasonable computing cost. A prototype based on the proposed model is constructed and evaluated. Our experiment results show that the proposed model achieves 88.9% detection and blocking rate in average for user-sensitive data revealed by the services of SNS.**

I. Introduction

NOWADAYS, people tend to use Social Networking Sites (SNS) to keep personal connections with others. According to Ho et al. [1], SNS is a website that provides a virtual community for people with similar interests in particular subject, or just to "hangout" together. Based on the definition of SNS, people can use SNS services to share information, thoughts and feelings, chat with other users, play online games with other users, and even promote their own businesses to other users [2]–[5]. The rapid growth and huge amount of service usage of SNS show that SNS has taken a significant role on communication media and culture among people in modern societies. For instance, Facebook Message is a social service provided by Facebook that allows people to chat online or message with each other offline.

During chatting or messaging sessions of SNS services, people may consciously or unconsciously input sensitive user information into their exchanging messages. Facebook provides a set of rules to protect its user's privacy [2], and it allows its users to determine who can access and see information shared by individual user. However, these rules cannot protect sensitive information shown within shared messages. When it comes to user privacy protection, it all depends on individual user itself to carefully avoid inputting user-sensitive data during the usage of SNS services.

This work was supported by Taiwan Information Security Center (TWISC) and National Science Council, Taiwan, under Grant no. NSC 102-2218-E-011-013

In the past few years, several mechanisms had been proposed by researchers to protect individual privacy for different occasions. User anonymization [6]–[10] and data encryption [11]–[14] are two major investigation directions.

User anonymization schemes as described in [6]–[10] protect a user from being identified from a set of data records by replacing user-related data within these records with system-generated strings before outputting these data for people to use. However, Beye et al. [3] explained that this kind of mechanisms still cannot defend against re-identification threat. Moreover, these mechanisms are not suitable to apply directly onto real-time chatting or messaging services since these mechanisms are only suitable for large dataset which contains thousands of different users' information. In addition, users have to identify themselves with a set of unique personal attributes first before they use any service offered by SNS.

Data encryption schemes as described in [11]–[14] protect communication privacy between users by encrypting communicating data or messages. Both communicating parties have to get corresponding keys in advance to encrypt and decrypt messages transmitting between them. Inevitably these mechanisms have key distribution/management problem to be resolved in practice. In SNS services, a real time message is usually shared or broadcast to multiple users/friends. Therefore, it is not easy to do group key management for SNS services provided that data encryption schemes are adopted directly. As the friend list of individual user will dynamically change from time to time, group key management for individual user will become an annoying routine job if the complexity of key management is endurable. In SNS environment, it is desirable to have a dynamic information protection mechanism that meets individual user's need. Since not all information shared by using SNS services is user-sensitive, a user privacy protection mechanism should be able to filter and determine which information is sensitive and needs to be protected.

Danger theory has been investigated in the field of Artificial Immune System [15] for its built-in ability to adapt dynamic changes automatically. Some researches using danger theory to construct application systems such as virus detection, network intrusion detection, and message filtering; have been conducted and shown its effectiveness.

978-83-60810-58-3/$25.00

In this paper we propose a privacy protection model for SNS Messaging System based on danger theory concept. Messages that does not contain sensitive information are defined as healthy cells. In contrary, messages that contains sensitive information are defined as injured cells. Danger signal defines what kind of dangers should be detected by the system. Danger signal in our model is the signal sent out by SNS messaging system when user-sensitive information inside a message is detected. Antigens are defined as the collection of user-related information. Based on our antigen's definition, we define the antibodies as a set of rules that regulate user-related information and determine what information is allowed to be shared to other users. Binary string is adopted to represent antigens and antibodies. Specific semantics of each bit within a binary string format are defined to indicate user-related data items and rules, respectively. A prototype was built based on the proposed model and performance evaluation was conducted accordingly. Based on the experiment results, the average accuracy rate for the proposed privacy protection model to correctly detect and protect user-sensitive information among shared messages is 88.9%.

II. Literature Review

A. Privacy Issues on Social Networking Sites

There are several user-related data that must exist and store in a SNS according to Beye et al. [3] such as profiles, connections, login credentials, messages, multimedia, groups, tags, preferences/rating/interests, and behavioral information. In this paper, we focus on one of the most popular services provided by SNS, i.e., online messaging service. This kind of service allows a user to exchange data/messages online or offline with another user or a group of users in SNS.

Since social networking sites such as Facebook usually collect and store all user-related data as stated by Beye et al., it raises two types of privacy issues. User-related privacy issues are generally caused by limited or lack of privacy control functionality support from service providers of SNS for their users [2], [3]. When SNS services provide a convenient platform for users to freely share information, individual user's sensitive information may be revealed by a user itself and freely accessed by other SNS users or even anonymous attackers. Another privacy issue is inability to hide user-sensitive information from other parties (friends or a specific group of users) [16], [17] because SNS vendors did not provide suitable privacy protection mechanisms. Another issue in this type category is user-sensitive information leakage caused by other users. When someone posted sensitive information related to a user of some SNS, it can harm the privacy of the indicated user.

As user's sensitive information can also be used to make profits for SNS service providers or other third-party companies that gain user information from SNS, the second type of privacy issues is often involved by SNS vendors and other companies cooperated with SNS vendors. According to Smith et al. [18], SNS users have no control over their published information on social network sites. In addition, users often do not know what SNS companies will do with their published data/messages. This kind of user privacy concern affects trust relationship between users and SNS service providers. Data retention on SNS is an example of the second type issues, in which all information that a user has ever been posted on the site is often impossible to be removed. Another example of privacy issues in this type category is unauthorized access to user data done by employees of SNS. Beye et al. [3], S. Mahmood [17], and D. J. Weitzner [19] indicated that most cases of privacy issue are related to sell user information to another party, such as an advertising company. Since SNS users cannot remove posted messages in SNS, those valuable information related with some users can be sold to other user-hunting companies such as advertising providers or insurance companies.

B. Privacy Protection Techniques

1. Privacy protection at service provider side

To solve privacy issues at service provider side, it is necessary for service providers to develop a privacy protection mechanism. One technique used to protect data privacy is to reduce the possibility that a user is identified based on data collected by a service provider. Several techniques have been proposed by researchers in the past to work on this issue, such as l-diversity [6], (α, k)-anonymization [7], and t-Closeness [8].

According to Machanavajjhala et al. [6], each row of data is composed from three different types of attributes: key attributes, quasi-identifiers, and sensitive attributes. Generally in published dataset, the value of a key attribute is in encrypted form to protect individual user's privacy. Aside from key attributes, quasi-identifiers might also be released in partial-encrypted form or in plain-text form. As for sensitive attributes, it is always released in plain-text form without any modification.

Based on the attribute structure for data records proposed by Machanavajjhala et al., user anonymization techniques developed in [6]–[8] can anonymize user-related information in published dataset. However, these mechanisms applying for huge dataset are not suitable to be applied directly in SNS environment, especially in real-time chatting or messaging services.

2. Privacy protection at the user side

Data encryption schemes were introduced to protect user's privacy in [11]–[14]. Privacy protection techniques based on encryption algorithms can be applied not only by SNS service providers but also by users of SNS. Since SNS users might be aware of privacy protection concern in SNS, they could take the initiative to protect their own privacy. In [11], Koch et al. presented a way to secure information shared among Facebook users by developing a third-party browser plug-in for Mozilla Firefox.

Since the plug-in utilizes cryptographic mechanisms, a user have to provide some information to invoke the plug-in.

The required data are targeted SNS URLs, usernames in the targeted sites, cryptographic algorithms, and their corresponding keys. Once the plug-in is activated, all messages entered by the user within targeted SNS web pages will be encrypted. To display the original messages posted in SNS, a user needs to present the corresponding key to the plug-in.

The mechanism in [11] has key distribution/management problem in practice. Since a real time message in SNS services is usually shared or broadcast to multiple users/friends, therefore group key management in SNS services is not an easy thing to do. Given that individual user's friend list is dynamically changed from time to time, it makes group key management become an annoying routine job. A dynamic user privacy protection should be able to manage this group key distribution/management problem. A good privacy protection mechanism should be able to filter sensitive information and protect it, because not all information shared using SNS services is user-sensitive.

C. Danger Theory

The original danger theory concept proposed by Polly Matzinger [20], [21] is a novel explanation on how the human body's immune system works. It is also an adaptive algorithm in the field of Artificial Immune System that generally used to perform virus detection, network intrusion, and message filtering. According to Lin et al. [15], this theory supersedes traditional self – non-self model, where the traditional model is more focusing on coping with any danger possibilities that may come from individual itself or outside of individual. The danger theory offers two advantages: the ability to prevent dangers in the future and the ability to defend against currently identified dangers.

There are several biology terms used in danger theory:

1. Tissue: a collection of cells in an organized form; multiple tissues can be organized to form an organ.
2. Cell: the basic structural, functional and biological unit of all known living organisms. In danger theory, there are two types of cell: normal cell and injured cell. A cell that does not cause harm to the corresponding organism is known as a normal cell. A cell that harms the corresponding organism is known as an injured cell.
3. Lymphocyte: any one of three types of white blood cell in a vertebrate's immune system. These three types of white blood cell are natural killer cell, T-cell, and B-cell.
4. B-cell: a type of lymphocyte in adaptive immune system. B-cell can be distinguished from other type of lymphocytes because it can bind to a specific antigen.
5. Antigen: any substance or incident that provokes an adaptive immune response in the body of an organism.
6. Antibody: Y-shape protein produced by plasma cells. Antibody is used by the immune system to identify and neutralize foreign objects such as bacteria and viruses.
7. Danger signal: an alarm signal sent out by a cell that is in distress or by an injured cell. The form of danger signal is varied in each immune system. For example, Lu et al. in [22] defined their danger signals based on the spreading and the damaging characteristics of mobile phone virus.

Fig. 1 shows an immune response according to the danger theory concept. A cell that is in distress sends out an alarm signal, known as the danger signal, whereupon antigens in the neighborhood are captured by *Antigen Presenting Cells* (APC), which then travel to local lymph nodes and present these antigens to lymphocytes. Essentially, the injured cell will establish a danger zone around itself. B-cell, one type of lymphocyte, will produce antibodies. Antibodies that can neutralize the antigens within the danger zone will perform clonal expansion process. Those antibodies that cannot neutralize the antigens or are located far away from the injured cell will not be stimulated and performed the clonal process.

Based on [23]–[25], the danger theory model can be

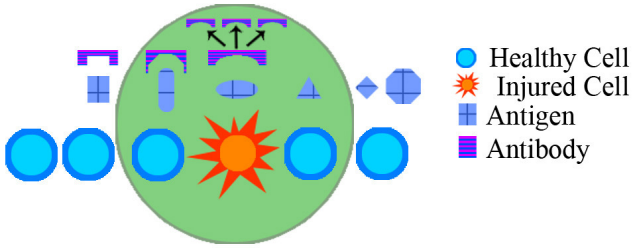

Fig 1. Danger theory model

viewed as an extension of Two-Signal model. In Two-Signal model, the two signals are antigen recognition (signal one) and co-stimulation (signal two). Co-stimulation signal is used to indicate the corresponding antigen is a dangerous one.

According to U. Aickelin and S. Cayzer [25], lymphocyte behaviors in danger theory are determined based on three laws:

- Law 1: A lymphocyte will be activated if the system receives both signal one and signal two altogether. If the system only receives signal one without signal two, then an activated lymphocyte will die. If the system only receives signal two without signal one, then it ignores the signal.
- Law 2: A lymphocyte only accepts signal two from Antigen Presenting Cells. Any cells can issue signal one to the system. Notice that an experienced T-cell or B-cell can act as Antigen Presenting Cells.
- Law 3: After the activation of a lymphocyte, the lymphocyte will revert to the resting state after a short time.

III. Proposed Privacy Protection Model

One of the most important privacy protection issues in SNS is the unintentional leakage of user-sensitive information caused by individual user's own carelessness. The importance of privacy protection on the data/messages shared by SNS users is not only for users' own interest, but also for service providers of SNS by offering privacy-aware SNS

TABLE I.
DEFINITIONS OF DANGER THEORY TERMINOLOGY FOR AISMPP

Danger Theory	AISMPP
Tissue	A message sent by a user to the other user(s).
Healthy cell	A message that does not contain any sensitive information related to individual user.
Injured cell	A message that contains any kind of sensitive information related to individual user.
Antigen (AG)	Collection of sensitive information related to individual user.
Antibody (AB)	A set of rules that regulates sensitive information, related to individual user, to decide which information is allowed to be shared to other(s).
Lymphocyte	The decision center.
Danger signal	A signal sent out by SNS messaging system indicating detected sensitive information inside one message and the user's response according to the given signal.
Signal one (danger signal)	A signal sent out by the SNS messaging system each time it detects user-sensitive information in one message.
Signal two (danger signal)	A signal used to indicate user's response or action toward the notification email sent by the system.
Danger zone	A status (state) indicator for a suspected message which might reveal user-sensitive information.
APCs	Messages received by users with alarm indication.

services to gain more users. To resolve this important user privacy issue in SNS environments, we propose the *Artificial Immune System Model for Privacy Protection* (AISMPP), which is derived from the concept of *Danger Theory* proposed by Polly Matzinger [20], [21]. To investigate detailed design of AISMPP, we select the most popular service among SNS functionalities, i.e., the messaging service, as the objective for AISMPP. The architecture of this system model is shown as Fig. 2.

In AISMPP design, there are six main components: SNS messaging service system, users database, user privacy settings, decision center, general rule repository, and antigen-antibody (AG-AB) database. As seen in Fig. 2, a user utilizes the messaging service offered by SNS to socialize with other users.

Each SNS has its own users and user privacy settings database. The privacy settings database contains a collection of rules and settings. These rules help the user to configure what data item in a user profile can be viewed by others, a black user list for each user, and a user data sharing list for third-party services (e.g. search engine services).

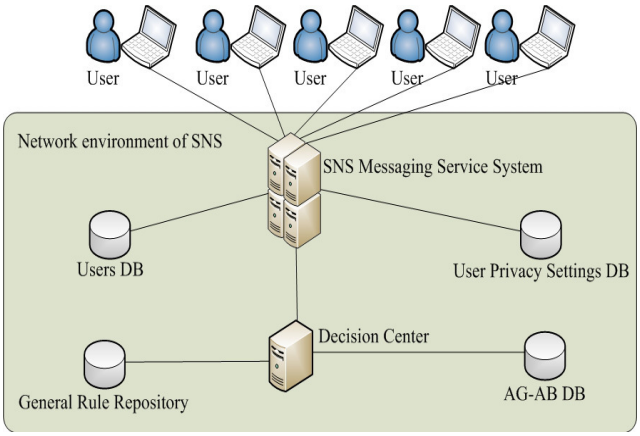

Fig 2. AISMPP architecture design

The decision center is responsible for processing danger signals, building danger zones, receiving antigens, generating and distributing antibodies. The decision center communicates with general rule repository and AG-AB database. The general rule repository describes all default actions for messages containing user-sensitive information. The AG-AB database stores all antigens and antibodies generated by the system.

The corresponding element definitions for AISMPP based on the danger theory are described in Table I.

Based on rules settings in [26] and [27], and our observation on E-Commerce websites and social networking websites; we propose 22 personal data items that are usually collected by SNS services. Thus in our design, an antigen is defined as a 22-bits binary string, where each bit represents one of 22 user's personal data items.

In Fig. 3 the AISMPP data flows are depicted. We describe each one of the flow processes as below:

1. AISMPP data flow processes start when user *A* sends a message *M* to user *B* using SNS's chatting/messaging service. Both user *A* and user *B* are participants in this chatting session.
2. Within the message SNS messaging system searches for user-sensitive information related with user *A*, user *B*, and their SNS friends list. If the system detects user-sensitive information in the message, then it will create an appropriate antigen based on the detected information.
3. After creating the appropriate antigen, the system will check the user privacy settings database for the privacy-affected user (or one of its friends) whether the disclosed user information is allowed to be shared or not. In our scenario, we assume that every user disallows any of its information to be shared by others.

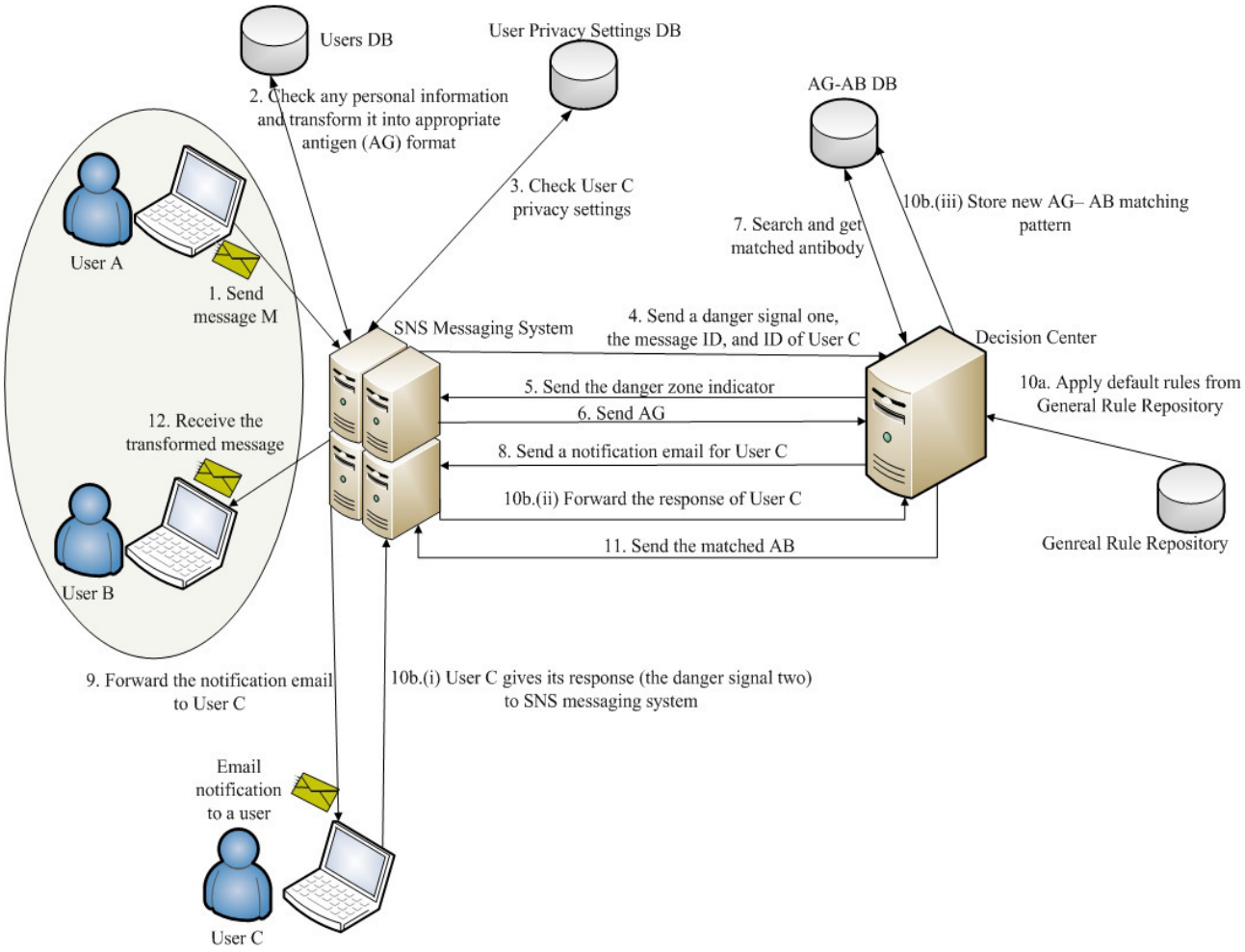

Fig 3. AISMPP data flow diagram

4. Every time the system detects any user-sensitive information in a message, the system will generate and send a danger signal (signal one) to the decision center along with the suspected message ID and the user ID of which user privacy has been breached.
5. The decision center will create a danger zone indicator on the suspected message based on the message ID and send the indicator to SNS messaging system.
6. After receiving the danger zone indicator from the decision center, the SNS messaging system sends the suspected antigen data to the decision center.
7. The decision center then check the antibody database to search the matched antibody for the received antigen.
8. If there is no matched antibody found on the antibody database, the decision center will request the SNS messaging system to forward a notification email to the message-affected user (assume it is user *C*) whose privacy is breached. The email notifies user *C* whether its sensitive information displayed in a message can be shared or not. If the decision center finds a matched antibody, then it will use this antibody and proceed to Step 11.
9. SNS messaging system will forward the notification email to User C.
10. Step 10 is divided into two cases based on how fast user *C* gives its response after the notification email sent by SNS messaging system:
    a. If there is no response from user *C* to SNS messaging system within a given period of time or in case that the decision center does not receive any response of user *C* forwarded by SNS system, then the decision center will apply default conservative rules stored in general rule repository for this suspected message.
    b. (i) If User C gives its response (the danger signal two) in time to SNS messaging system. (ii) SNS messaging system forwards the response of user *C* to the decision center. (iii) Based on the user response, the decision center will create a new antigen-antibody matching pattern and store it into the AG-AB database.
11. The decision center sends the matched antibody to the SNS messaging system. This antibody will be used by SNS messaging system to transform the message sent by user *A*.

12. After transforming the message, SNS messaging system will relay the message to user *B*.

IV. Prototype Design

Based on the data flow and functionality design shown in Section 3, an AISMPP prototype is developed. The AISMPP prototype consists of two stages. In the initialization and training stage, the prototype is initialized with default users list, default antigens, default antibodies, and initial message log based on real Facebook user data. In the system execution stage, the prototype will dynamically identify user-input messages with user-sensitive information and transform these message contents before they are displayed to designated users to preserve user privacy.

Fig. 4 shows the pseudocode of initialization and training stage. In addition to initialize the users list, default antigens are generated in the form of 22-bit binary string for each antigen pattern. If our prototype detects any specific user-sensitive information in a message, then the corresponding

```
PROCEDURE Init_Training_Model()
  INIT USERS ← LOAD (initial users list)
  INIT AG_DB ← LOAD (initial antigen
    data)
  INIT AB_DB ← LOAD (initial antibody
    data)
  INIT L ← LOAD (initial message log
  data)
  FOREACH (person P ∈ USERS)
    FOREACH (message M ∈ L)
      IF (message M contains sensitive
        information)
        new_ag ← create an antigen
          record from the found
          sensitive information
        new_w ← detected sensitive
          words
        new_ab ← create new antibody
          based on new_ag
        IF (new_ag ∈ AG_DB)
          UPDATE the number of
            occurrence of detected
            words based on new_w and
            promote the antigen new_ag
            to have longer life time.
        ELSE
          ADD new_ag to AG_DB
        ENDIF
        IF (new_ab ∈ AB_DB)
          UPDATE the number of usage
            of the antibody new_ab
        ELSE
          ADD new_ab to AB_DB
        ENDIF
      ENDIF
    ENDFOREACH
  ENDFOREACH
ENDPROCEDURE
```

Fig 4. Initialization and training stage of AISMPP

binary bit that represents one specific user-sensitive data item will be set to 1, otherwise 0. A corresponding antibody pattern for each antigen is also generated, in which a 22-bit binary string is used to indicate whether the data item in the corresponding antigen should be encrypted; the corresponding bit of the antibody is set to 1, if the data item in the antigen needs to be encrypted.

The user profile in our prototype contains user's real name, user's email(s), user's birthday date, user's address (city, region, and country), user's phone number(s), and other private information, e.g. religious and political views. According to [26] and [27], we define all items in user profile as user-sensitive information.

The prototype also loads friend information of users from real Facebook user data. The friend information getting from Facebook is limited to the full user name, email addresses, and phone numbers.

Some historical messages in plain-text form are stored in the message log during the initialization and training stage. These historical messages are used as the training data for our prototype. All historical messages will be processed in the initialization and training stage to make our prototype ready for regular operation. Each time the prototype receives an input message, the system will search the message for any user-sensitive information. If there is no user-sensitive information found in the message, then the message will be forwarded directly to the recipient without any transformation. Otherwise, the message-related data will be recorded, i.e., the receivers, the sender, and the message itself.

After the suspected message data has been recorded, a temporary antigen pattern is created and the suspected words in the message are recorded. In case the created antigen exists at antigen database, our prototype will update the number of occurrence of the words at the antigen database, which cause the system to generate the corresponding antigen. Otherwise, the newly created antigen will be added into the antigen database. Furthermore, our prototype will also create new antibody pattern based on the newly created antigen.

In Fig. 5 we show regular operation process of AISMPP at system execution stage. Two additional processes, user feedback processing and user message transformation, are included in this stage in comparison with the process for initialization and training stage.

In Fig. 6 the user feedback process assigns a flag indicator along with the user-affected message based on the user's response. Notice that if the user does not give any response after receiving notification email from SNS messaging system, then the prototype will assume the message possesses user-sensitive information and apply the message transformation process to protect user-sensitive information inside the message.

User message transformation process in Fig. 7 will transform all the detected words, which are user-sensitive information, within a message into encrypted ones. Considering robust security, SHA-256 algorithm is adopted for data encryption operation in our prototype.

```
PROCEDURE Execution_Process()
  LOOP
    M ← a new message inputted by user
    U ← a user ID that generates the
      message M
    FOREACH (person P ∈ friends of U)
      IF (M contains sensitive
        information)
        IF (person P responds to
          system notification)
          SET the flag of applying
            default rules F = FALSE
        ELSE
          SET flag F = TRUE
        ENDIF
        DF = Process_Feedback(M,
          default flag F)
        new_ag ← create an antigen
          record from the found
          sensitive information
        new_w ← collection of detected
          sensitive words
        new_ab ← create new antibody
          based on new_ag
        IF (new_ag ∈ AG_DB)
          UPDATE the number of
            occurrence of detected
            words based on new_w and
            promote the antigen
            new_ag to have longer
            life time.
        ELSE
          ADD new_ag to AG_DB
        ENDIF
        IF (new_ab ∈ AB_DB)
          UPDATE the number of usage
            of the antibody new_ab
        ELSE
          ADD new_ab to AB_DB
        ENDIF
        IF (DF)
          M = Transform_Message(M)
        ENDIF
      ENDIF
    ENDFOREACH
    SEND M to user
  ENDLOOP
ENDPROCEDURE
```

Fig 5. System execution stage of AISMPP

V. Implementation and Experiments

The prototype is developed on Windows 7 platform within a PC hardware of Intel Core i5 3.1 GHz CPU and 4 GB RAM. For prototype initialization, user profiles are generated based on the 274 friend contacts of one real Facebook user account and 100 user contacts generated by the Web service of generatedata.com. The historical message log from the same Facebook user account is loaded for the training stage. In initialization stage, 6 antigen patterns are defined in which 7 user data items including name, email, social security number, phone number, passport number, credit card number and credit card expiration date are used and

```
PROCEDURE Transform_Message(message M)
  M' = M
  FOREACH (get each word string W in M)
    IF (W is user-sensitive data)
      W' = ENCRYPT(W)
    ELSE
      W' = W
    ENDIF
    M' = REPLACE(W,W')
  ENDFOREACH
  RETURN new message M'
ENDPROCEDURE
```

Fig 7. The function to transform a user's message in AISMPP

combined to generate these patterns. In addition, 6 corresponding antibody patterns are generated based on the assumption that any revealed user data item should be blocked (by data transformation). For the system execution stage, extra 200 friend contacts from another Facebook user account and 300 new user contacts generated by the Web service of generatedata.com are loaded into our prototype. In addition, 5000 newly generated messages using the Web service of RandomTextGenerator.com combined with the historical message log from the second Facebook user account are imported to evaluate the effectiveness of our prototype.

Fig. 8 shows the time distribution for processing 5000 messages within 1000 experiments. The processing time includes the time for message scanning process and the time for detecting user-sensitive information among 5000 messages. The higher processing time during the first 30 experiments is caused by adding more antigen and corresponding antibody patterns based on the self-adaptive capability of danger theory model as shown in Fig. 9. Notice that the processing time gradually reaches stable condition around 25 seconds during our experiments.

Fig. 9 shows the distribution of total number of antigen patterns within 1000 experiments. After completing the training stage of our prototype, the total number of antigen is around 260. In the first couple of experiments, it can be seen that the total number of antigen increases quickly and reaches beyond 500. The reason is the patterns of newly generated messages for our experiments are different with the message patterns applied in the training stage. As the number of antigen increases, the message processing time is

```
PROCEDURE Process_Feedback(message M,
  flag F)
  IF (!F && user consider the sensitive
    information in message M is alright
    to be shared)
    ASSIGN the safe flag DF=FALSE to
      message M
  ELSE
    ASSIGN the danger flag DF=TRUE to
      message M
  ENDIF
  RETURN DF
ENDPROCEDURE
```

Fig 6. The function to process user feedback in AISMPP

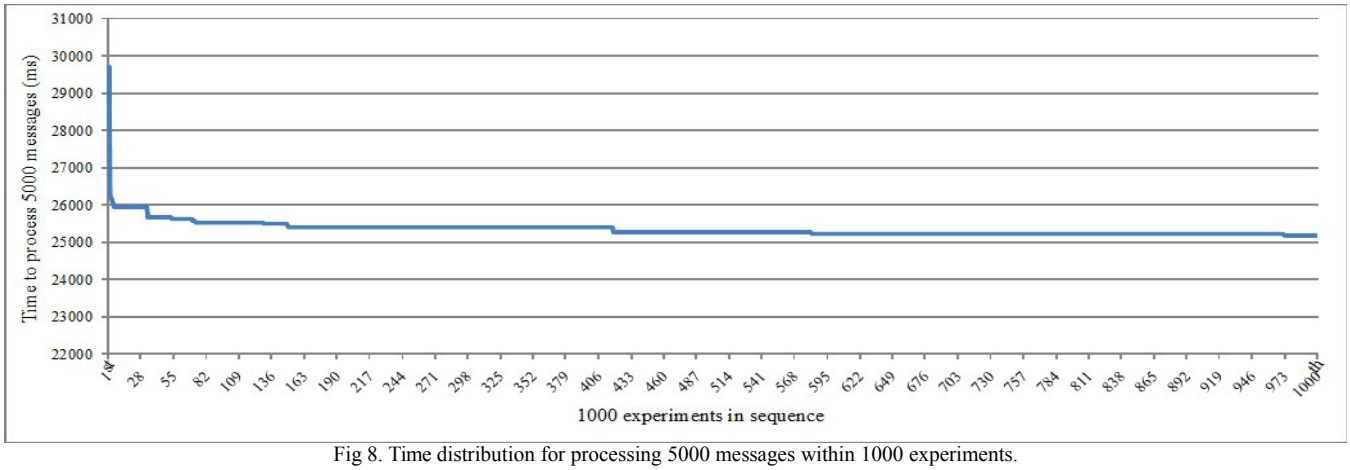

Fig 8. Time distribution for processing 5000 messages within 1000 experiments.

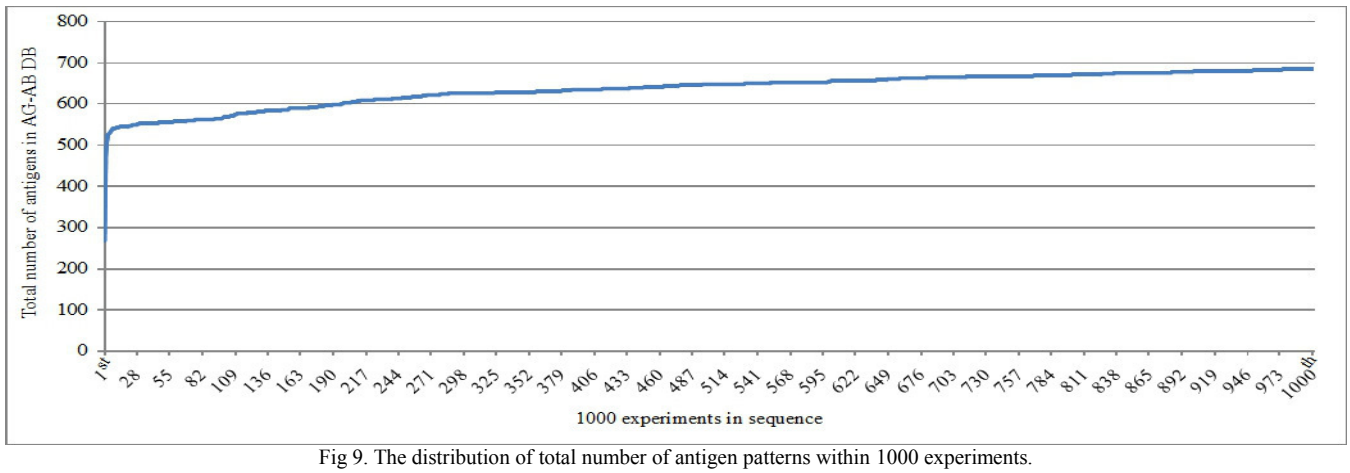

Fig 9. The distribution of total number of antigen patterns within 1000 experiments.

gradually reduced based on the observation of both Fig. 8 and Fig. 9.

Fig. 10 shows the time distribution for processing a message from the first message to the 5000th message at the 50th experiment. In Fig. 10 there are many messages been processed within 10ms. The reason is existing antigens have recognized this message pattern. Therefore, it does not require much time for the prototype to process these messages. For messages with new pattern, the prototype requires more time to process and generate new antigens and antibodies if necessary.

Fig. 11 shows the average time for our prototype to process one message in the first 100 experiments. For example, the average time to process one message in the 51st experiment is 5.21ms as shown in Fig. 11. Based on Fig. 11, we can know that the average time to process one message is ranged between 5ms and 10ms. We think the processing time is acceptable for a SNS messaging service.

We define a true positive (TP) value as our prototype correctly predicting one user-sensitive data shown in a message. A true negative (TN) value is defined as our prototype correctly predicting no user-sensitive data shown in a message. A false positive (FP) value is defined as our prototype wrongly predicting one user-sensitive data shown in a message. A false negative (FN) value is defined as our prototype wrongly predicting no user-sensitive data shown in a message. The precision rate is defined as TP / (TP + FP). The recall rate is defined as TP / (TP + FN). The accuracy rate is defined as (TP + TN) / (TP + TN + FP + FN). The average precision rate of our prototype is around 91.7% and the average recall rate is around 96.7%. The average accuracy rate of the prototype to correctly detect user-sensitive information in messages is around 88.9%.

VI. Conclusion

Lacking of automatic user privacy control mechanisms and privacy protection schemes is one of the most concerning issues in Social Networking Sites (SNS). In this paper, a privacy protection model based on danger theory is proposed. Several danger theory components are re-defined to fit into SNS environment and binary string format is adopted to represent antigens and antibodies used in our privacy protection model. Specific semantics of each bit within a binary string format are defined to indicate user-related data items and rules, respectively. Based on these components, an automatic adaptive immune system for user-sensitive information protection during online chatting/messaging is designed.

A prototype of our design was built based on the proposed model and performance evaluation was conducted accordingly. Based on the experiment results, the average accuracy rate for the proposed privacy protection model to correctly

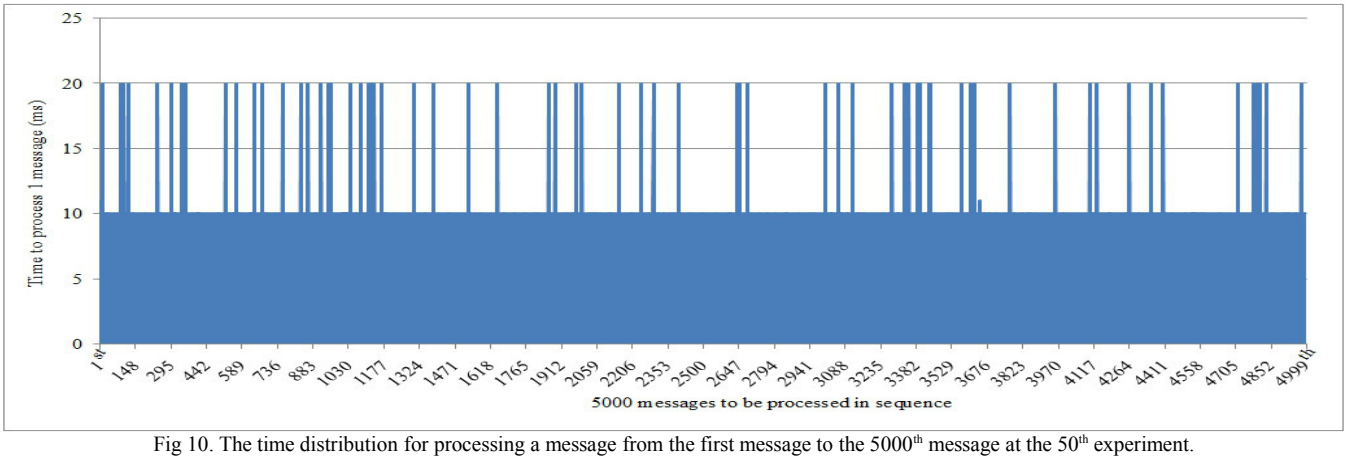

Fig 10. The time distribution for processing a message from the first message to the 5000th message at the 50th experiment.

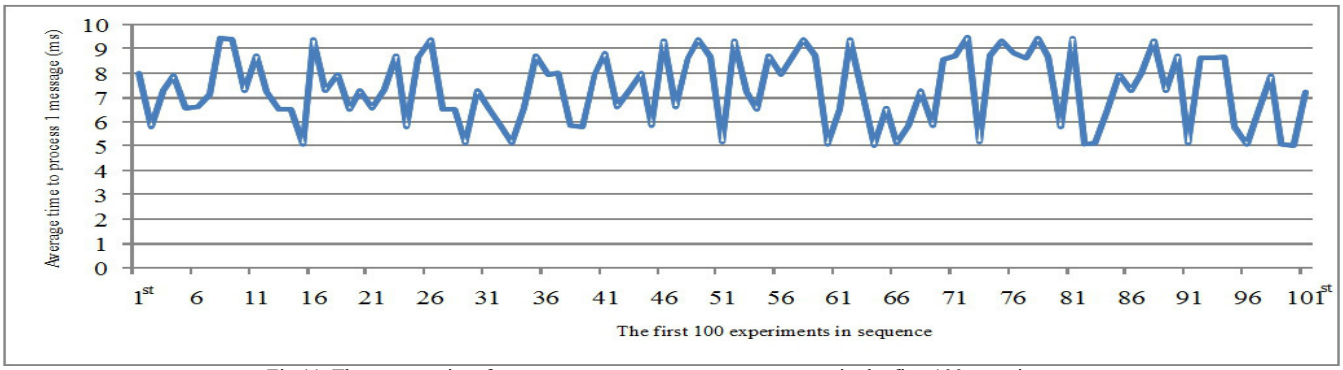

Fig 11. The average time for our prototype to process one message in the first 100 experiments.

detect and protect user-sensitive information among shared/broadcast messages is 88.9%. In addition, we found that the system processing time for each shared message is reduced with the increase of the number of recognized anti-gen patterns.

References

[1] A. Ho, A. Maiga, and E. Aimeur, "Privacy protection issues in social networking sites," presented at the IEEE/ACS International Conference on Computer Systems and Applications, 2009, pp. 271–278, http://dx.doi.org/10.1109/AICCSA.2009.5069336

[2] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, "Privometer: Privacy protection in social networks," presented at the IEEE 26th International Conference on Data Engineering Workshops (ICDEW), 2010, pp. 266–269, http://dx.doi.org/10.1109/ICDEW.2010.5452715

[3] M. Beye, A. J. P. Jeckmans, Z. Erkin, P. Hartel, R. Lagendijk, and Q. Tang, "Privacy in Online Social Networks," in *Computational Social Networks*, A. Abraham, Ed. Springer London, 2012, pp. 87–113.

[4] M. S. Choi, H. W. Kim, Y. H. Kim, K. H. Chung, and K. S. Ahn, "Private Information Protection System with Web-Crawler," presented at the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2008, pp. 672–677, http://dx.doi.org/10.1109/WiMob.2008.63

[5] Z. Chi, S. Jinyuan, Z. Xiaoyan, and F. Yuguang, "Privacy and security for online social networks: challenges and opportunities," *IEEE Network*, vol. 24, pp. 13–18, 2010, http://dx.doi.org/10.1109/MNET.2010.5510913

[6] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," in *Proceedings of the 22nd International Conference on Data Engineering*, 2006, p. 24, http://dx.doi.org/10.1109/ICDE.2006.1

[7] R. C. W. Wong, J. Li, A. W. C. Fu, and K. Wang, "(α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Philadelphia, PA, USA, 2006, pp. 754–759, http://dx.doi.org/10.1145/1150402.1150499

[8] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," presented at the IEEE 23rd International Conference on Data Engineering, 2007, pp. 106–115, http://dx.doi.org/10.1109/ICDE.2007.367856

[9] P. Jurczyk and L. Xiong, "Privacy-preserving data publishing for horizontally partitioned databases," in *Proceedings of the 17th ACM conference on Information and knowledge management*, Napa Valley, California, USA, 2008, pp. 1321–1322, http://dx.doi.org/10.1145/1458082.1458257

[10] X. Jin, N. Zhang, and G. Das, "Algorithm-safe privacy-preserving data publishing," in *Proceedings of the 13th International Conference on Extending Database Technology*, Lausanne, Switzerland, 2010, pp. 633–644, http://dx.doi.org/10.1145/1739041.1739116

[11] R. Koch, D. Holzapfel, and G. D. Rodosek, "Data control in social networks," presented at the 5th International Conference on Network and System Security (NSS), 2011, pp. 274–279, http://dx.doi.org/10.1109/ICNSS.2011.6060014

[12] A. P. A. G. Deshmukh and R. Qureshi, "Transparent Data Encryption - Solution for Security of Database Contents," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 3, pp. 25–28, Mar. 2011. Available: http://arxiv.org/abs/1303.0418

[13] Xukai Zou, Peng Liu, and J. Y. Chen, "Personal genome privacy protection with feature-based hierarchical dual-stage encryption," presented at the IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS), 2011, pp. 178–181, http://dx.doi.org/10.1109/GENSiPS.2011.6169474

[14] S. Jahid, P. Mittal, and N. Borisov, "EASiER: Encryption-based Access Control in Social Networks with Efficient Revocation," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, New York, NY, USA, 2011, pp. 411–415, http://dx.doi.org/10.1145/1966913.1966970

[15] L. Lin, L. Yiwen, Y. He, and Y. Chao, "Danger Theory: A new approach in big data analysis," presented at the International Conference on Automatic Control and Artificial Intelligence (ACAI), 2012, pp. 739–742, http://dx.doi.org/10.1049/cp.2012.1083

[16] A. Johnston and S. Wilson, "Privacy Compliance Risks for Facebook," *IEEE Technology and Society Magazine*, vol. 31, pp. 59–64, 2012, http://dx.doi.org/10.1109/MTS.2012.2185731
[17] S. Mahmood, "New Privacy Threats for Facebook and Twitter Users," presented at the 7th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2012, pp. 164–169, http://dx.doi.org/10.1109/3PGCIC.2012.46
[18] M. Smith, C. Szongott, B. Henne, and G. von Voigt, "Big data privacy issues in public social media," presented at the 6th IEEE International Conference on Digital Ecosystems Technologies (DEST), 2012, pp. 1–6, http://dx.doi.org/10.1109/DEST.2012.6227909
[19] D. J. Weitzner, "Google, Profiling, and Privacy," *IEEE Internet Computing*, vol. 11, pp. 95–96, c3, 2007, http://dx.doi.org/10.1109/MIC.2007.129
[20] P. Matzinger, "Tolerance, Danger, and the Extended Family," *Annual Review of Immunology*, vol. 12, no. 1, pp. 991–1045, Apr. 1994, http://dx.doi.org/10.1146/annurev.iy.12.040194.005015
[21] P. Matzinger, "The Danger Model: A Renewed Sense of Self," *Science*, vol. 296, no. 5566, pp. 301–305, Apr. 2002. Available: http://www.sciencemag.org/content/296/5566/301.abstract
[22] T. Lu, K. Zheng, R. Fu, Y. Liu, B. Wu, and S. Guo, "A Danger Theory Based Mobile Virus Detection Model and Its Application in Inhibiting Virus," *Journal of Networks*, vol. 7, pp. 1227–1232, 2012, http://dx.doi.org/10.4304/jnw.7.8.1227-1232
[23] M. Read, P. S. Andrews, and J. Timmis, "An Introduction to Artificial Immune Systems," in *Handbook of Natural Computing*, G. Rozenberg, T. Bäck, and J. N. Kok, Eds. Springer Berlin Heidelberg, 2012, pp. 1575–1597.
[24] E. Hart and J. Timmis, "Application areas of AIS: The past, the present and the future," *Applied Soft Computing*, vol. 8, pp. 191–201, 2008, http://dx.doi.org/10.1016/j.asoc.2006.12.004
[25] U. Aickelin and S. Cayzer, "The Danger Theory and Its Application to Artificial Immune Systems," in *1st International Conference on Artificial Immune Systems (ICARIS)*, Canterbury, UK., 2002, pp. 141–148.
[26] E. McCallister, T. Grance, Scarfone, and NIST U. S. Department of Commerce, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. 2010.
[27] M. E. Callahan and U. S. Department of Homeland Security, *Handbook for Safeguarding Sensitive Personally Identifiable Information*. 2012.