A Personal Information Leakage Prevention Method on the Internet

Daeseon Choi, Seunghun Jin and Hyunsoo Yoon

Abstract — *In this paper, we propose a method for preventing personal information leakage on the Internet. The leakage of the personal information might cause severe problems such as privacy violation, impersonation, spam mail, and financial fraud. The main ways of the personal information leakage are the leakage of the personal information registered in web site, the Internet Phishing, and the spyware. The basic idea of our method for preventing these types of personal information leakage is "Do not send the personal information to a hazardous recipient". Every network packet transferred from a user's PC to a server via the Internet is inspected to check if the packet contains the user's personal information. When a packet containing personal information is detected, a decision about safety of the transfer is made. After decision is made, the packet sent to an unsafe destination is dismissed. The decision is made based on the predefined user control policy. The user policy specifies the safeness of a transfer in considering the information such as type of transferred personal information, the application that sends the packet and the trustworthiness of the recipient. The destination's trustworthiness is managed and provided by a trusted third party. In this paper, we present the explanation of information leakage problem and the description of related work. The presentation of our model for controlling personal information transfer and a description of the system architecture for implementing our model is included. And some security analysis of our method that shows the effectiveness of the proposed method is also presented.*

Index Terms — Personal Information, Privacy Protection, Phishing, Spyware.

I. INTRODUCTION

The leakage of the personal information is that the personal information is acquired by the other man who is not intended by the owner of the personal information. The leakage of the personal information on the Internet is a very severe problem in the Internet era. The leaked personal information may violate the person's privacy. Some personal information such as ids and passwords can be used in impersonation. Contact information such as e-mail address and telephone number can causes many spam mail or spam phone call. Credit card numbers, pin code or bank account number may be used in a crime. There are many ways of the personal information leakage. The personal information that was registered by a user on a web site might be drained out by a malicious operator of the web site. Or, it might be happen that a hacker intrudes the web site and steals the personal information. The Internet Phishing, which means that some malicious people make a fake web site that looks like a famous web site and make the user undoubtedly input his personal information, is an another case of the personal information leakage. The spyware, a software that is installed secretly in a user's personal computer steals the personal information stored in the PC or captures user's keyboard input. The spyware sends the information to the malicious person who installed the spyware. And the eavesdropping of a network line between user's PC and application server is another way of the leakage.

In this paper, we propose a method for prevent and minimize the leakage of the personal information. The basic idea of the method is "Do not send the personal information to the hazardous recipient". In our method, all the network traffics transferred from a user's PC to a server via the Internet are monitored, and the packets that include the user's personal information and are being sent to the trustless recipient are filtered out. Filtering decision is carried out based on the types of transferred personal information, the application that sends the packet in addition to the trustworthiness of the recipient. With this method, we take aim at the following types of the personal information leakage.

- The leakage of the personal information registered in web site
- The Internet Phishing
- The personal information draining by the spyware

This paper consists of following sections. In section II, some related researches are presented. In section III, the Personal Information Transfer Control Model is presented. The PITC model is a scheme for decision of filtering packet that includes personal information. Section IV shows the Personal Information Transfer Control System that implements the PITC model. The analysis of how the proposed method solves the aimed problems is presented in section V. The section VI includes conclusion of this paper.

Daeseon Choi and Seunghun Jin are with the Information Security Research Division, Electronics and Telecommunication Research Institute, Daejeon, South Korea (e-mail: sunchoi@etri.re.kr, jinsh@etri.re.kr).

Hyunsoo Yoon is with the Computer Science Division, EECS Department, Korea Advanced Institute of Science and Technology, Daejeon, South Korea (e-mail: hyoon@camars.kaist.ac.kr).

1-4244-0216-6/06/$20.00 ©2006 IEEE

II. PREVIOUS WORKS

In our knowledge, there were few previous researches that had tried to solve the personal information leakage problem. Many cryptographic technologies including SSL[1] had been used to protect transferred information from eavesdropping. But encryption does not solve the types of the personal information leakage problem mentioned above.

Many anti-spyware detector had been developed[2][3]. They search the hard-disk and clean the detected spyware. Anti-spyware can only detect known spyware. Therefore they can not protect personal information leakage caused by the unknown spyware.

Some privacy control techniques has been focusing the control of the sharing of the personal information[4][5]. They are mainly targeting the personal information stored in the organization or internet web server. In those methods, control policy is defined by user and enforced by the server that stores the personal information. It assumes that the server is trustable in observing the user's control. This assumption is contradictory with the current Internet environment where there are many malicious entities that violate personal privacy.

And P3P is a technique for negotiation between a web browser and a web site about transferring personal information between the two entities[6]. In P3P, web site's policy about personal information handling is defined. Based on the policy, browser decides to enter the web site or not. Basically, this scheme also assumes that the web site is trustable. So P3P can not solve the personal information leakage problems defined in the section I.

III. PERSONAL INFORMATON TRANSFER CONTROL MODEL

The Personal Information Transfer Control Model is a scheme for decision and enforcement of filtering a packet that includes the personal information. Fig. 1 shows the construction of the PITC model.

A. Components

The components that consist of the PTIC model are as followings.

- Personal Information Transfer Enforcement Point

This entity monitors all the packets sent from the user's pc to the outer entity. When a packet that includes some personal information is sensed, the PITEP queries to the Personal Information Transfer Decision Point whether this packet can be transferred. Depending on the response of the PITDP, it filters the outbound packet. The method of detecting personal information transfer is explained in the section IV.

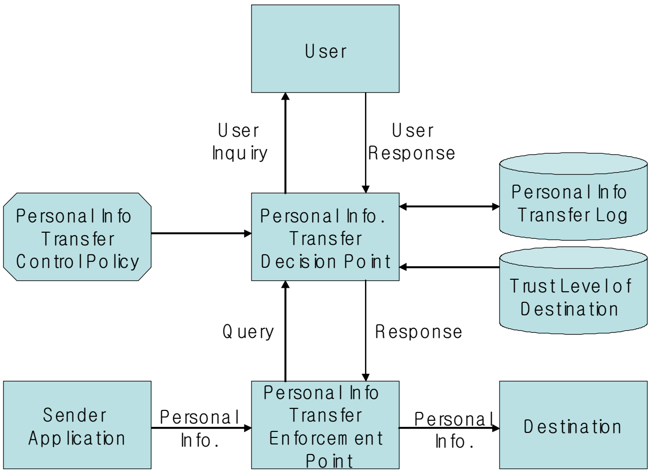
User
User Inquiry
User Response
Personal Info Transfer Log
Personal Info Transfer Control Policy
Personal Info. Transfer Decision Point
Trust Level of Destination
Query
Response
Sender Application
Personal Info.
Personal Info Transfer Enforcement Point
Personal Info.
Destination

Fig. 1. Proposed architecture with normal agent

- Personal Information Transfer Decision Point

The PITDP receives the query from the PITEP. The query from the PITDP consists of the following items.

- Type of transferred personal information
- Destination: IP address or DNS name
- Sending Application

The PITDP decides whether the queried transfer is safe. The decision is made based on the Personal Information Transfer Control Policy using the items in the query, the Personal Information Transfer Log, and the Trust Level of Destination. Details of decision criterion and procedure are presented in the following sections. The decision result is one of the *allow*, *deny*, and *user query*. When the decision is the user query, the PITDP initiates the user inquiry and asks the user to make the final decision. The user response is one of the *allow* and *deny*. After the decision is made, the response including the decision is sent to the PITEP.

- User

The User responds to the inquiry from the PITDP.

- Sender Application

The personal information transfer control process is initiated when the sender application sends a packet including some personal information.

- Destination

The destination is an application server or web site that receives the packet including personal information.

- Personal Information Transfer Log

The Personal Information Transfer Log is the record of previously transferred personal information. The record consists of type of the personal information, the destination, and the sending application. Some type of personal information such as social security number has a characteristic of unique identification. The combination of some types of personal information that doesn't have the characteristic of unique identification alone has the characteristic of unique identification. To decide whether a transfer of some personal information makes the destination be able to identify the user uniquely, it is necessary to check the combination of the currently transferred information and the information that had been transferred to the same destination before.

- Trust Level of Destination

The Trust Level of Destination is information about trustworthiness of a destination. The trustworthiness is evaluated based on the personal information protection procedure of the destination, reputation and privacy invasion incident records. The trust level of destination is assigned by the reliable third party. For an unknown destination for which there is no trust level assigned, the lowest level is assumed.

B. Policy

The Personal Information Transfer Control Policy is a criterion for decision of the personal information transfer. The PITC policy is established and managed by the user. The PITC policy is defined as followings.

TI= {All, addr, sex, telephone number... }
CI={Uniquely_Identifiable, Privacy_Sentive, Contact_Point, Financially_Critical}
TI->CI
TI->CI*
TL= {1,2,3,4,5}, max: 5
S={all, program_name}
D={allow(1), inquiry(0), deny(-1) }
R: (S, TI or CI, TL)->D, D for Lower TL <= D for Higher TL
P={Rule}

- *S* : Sender Application
- *TI*: Type of Information
- *CI*: Class of Information

CI is a method for grouping types of information according to the characteristics of the type. Any *TI* may belong to more than one *CI*. Currently, there are four class of information. *Uniquely_Identifiable* can be assigned to the SSN. *Financially_Critical* may contain bank account number, credit card number and etc. Category *Contact_Point* can be assigned to the phone number, and e-mail address. *Privacy_Sensitive* can be taken literally. And a combination of more than two *TI*s can be assigned to a *CI.*

- *TL*: Trust Level of Destination
- *D* : Decision
- *R*: Rule
- *P*: Policy

Set of rules comprises a policy.

C. Procedure

The procedure of the PTIC model processing the query is described below.

1. Query processing

For a Query *Q(S, TI, D)* in which the *D* represents Destination, the *TL* of the *D* is retrieved from Trust Level List and the Query is transformed to a *Q1´(S, TI, TL)*. Based on the policy of mapping *TI* and *CI*, the expanded query *Q2´(S, TI->CI, TL)* is added. Then the *TI*'s, the type of personal information that had been transferred to the same destination are combined with TI and the combination is also used in the query expansion. So *Q3´ (S, TI ·TI' ->CI, TL)* is also added. All the rules that match the queries in the policy are retrieved. If there is not any rule that can be applied, the query is sent to the user and the decision is assigned according to the use's answer.

2. Collision resolution

If there is more than one rules that is applied to the queries, the priority is given according to the following order.

- Rule containing *CI.*
- Rule containing specific type of information rather than all
- Other rule

IV. PERSONAL INFORMATION TRANSFER CONTROL SYSTEM

A. Components

The Personal Information Transfer Control Model described in the section III will be implemented in a Personal Information Transfer Control system as shown in the fig. 2.

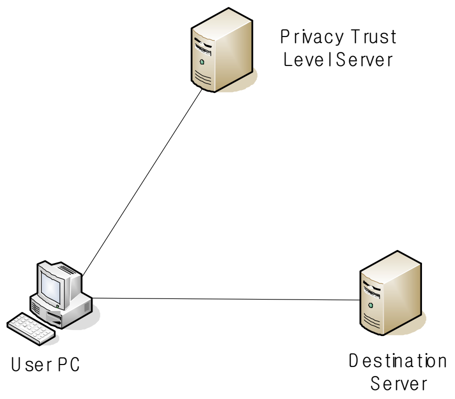

Fig 2. Personal Information Transfer Control System

The Privacy Trust Level Server is operated by a trusted third party and provides the Trust Level of Destination to the user pc. In the user pc, the Personal Information Transfer

Controller software is installed. The PITC software structure is shown in the fig. 3.

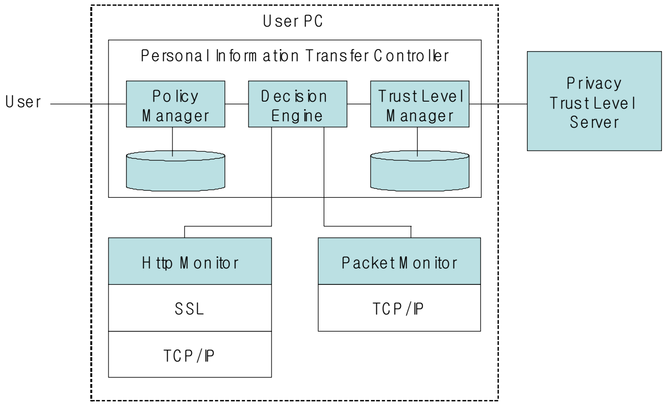

Fig 3. Personal Information Transfer Control Software Structure

The Personal Information Transfer Controller software consists of the Personal Information Transfer Controller installed in a user pc, the Privacy Trust Level Server running in the trusted third party server, and the packet monitors that monitor http packet and other network packet. The followings are the description of the components of the PITC software.

- Decision Engine

The Decision Engine is a core module of the PITC and works as a Personal Information Transfer Decision Point. It responses to the queries that is sent by the Http Monitor or Packet Monitor which is the Personal Information Transfer Enforcement Point.

- Policy Manager

The Policy Manager manages and stores the Personal Information Transfer Control Policy. A User establishes and maintains the PITC policy with the Policy Manager. And the Decision Engine retrieves required policy from the Policy Manager.

- Trust Level Manager

The Trust Level Manager acquires and stores the Trust Level of Destination List. The Trust Level Manager downloads periodically the Trust Level of Destination List from the Privacy Trust Level Server. The Decision Engine inquires the trust level of a specific destination to the Trust Level Manager.

- Packet Monitor

The Packet Monitor attaches to the TCP/IP module and monitors all the packets that is sent to the outer server. When it senses that a packet contains the personal information, it makes and sends a query to the Decision Engine. According to the response of the Decision Engine, the Packet Monitor sends the packet to the original destination or dismisses the packet.

- Http Monitor

The reason for existence of the separate Http Monitor in addition to the Packet Monitor is for monitoring the http packets before the packets is encrypted by SSL layer.

B. *Mechanism*

The mechanism for sensing a network packet that contains the personal information is a kind of pattern matching. Three kinds of pattern matching is used.

- Tag matching

It is used to detect a network packet that contains a tag information combined with the personal information. For example, a packet “addr=Seoul%20Korea%20..” contains a tag “addr”.

- Template matching

The template matching uses the form or shape of the personal information. For example, the Citizen Registration Number of Korea is 13 digits of number or is inscribed as of “XXXXXX-XXXXXXX”. If a text stream of six digits and slash and seven digits is found in the network packet, it is most probably the Citizen Registration Number in Korea.

- Value matching

The value of the personal information itself is the most probable matching key for sensing the personal information in a network packet.

V. Security Analysis

In this section, how the proposed method solves the aimed problems is described.

- The leakage of the personal information registered in web site

In our method, the transfer of personal information to the unreliable web site that might be possible to leak the information is blocked or inquired to the user. It minimizes the personal information transfer to the malicious sites or the web sites that have not enough security facility or system for protecting personal information.

- Internet Phishing

A phishing web site imitates a famous web site that has high confidence of users. It makes a user input his personal information without any doubt. It is very difficult for a user to distinguish the original web site and the phishing web site without careful attention. But to a machine, it is very easy job. Even tough the phishing web site fakes the original web site very precisely, the DNS name and ip address are different from those of the original. In the propose method, the phishing site has a lowest trust level. So a personal information transfer to the phishing site is blocked or inquired to the user. It functions as an alert for the user who thinks that the web site is genuine.

- The personal information draining by the spyware

The Personal Information Transfer Control Policy specify which application can or can not send which data to which destination. When a spyware sends some personal information, there is not any matching policy that includes the name of spyware or the name of destination. So the user inquiry is invoked. It invokes user's doubt of the transfer that he does not initiate.

VI. Conclusion

In this paper, we propose a method for preventing or minimizing the leak of personal information in the Internet. The Personal Information Transfer Control Model is a basis for systematic control of transfer. The PITC Model includes components, policy, and procedure. The PITC System architecture is also proposed. The system consists of a program installed in the user pc and a server that provides trustworthiness information about web sites. The propose scheme's effectiveness is analysed. The result shows how the proposed method solves the declared problems. Although the proposed method can not completely prevent all the case of the personal information leakage, it may remarkably minimize the personal information leakage.

Future works is followed. We will sophisticate our model. The type and class of the personal information will be extended and subdivided. The more concrete method for evaluating trustworthiness of a web site in the Privacy Trust Level Server will be also considered.

References

[1] http://www.openssl.org

[2] http://clinic.ahnlab.com/clinic/spyzero.jsp

[3] https://www.networkassociates.com/us/local_content/datasheets/ds_anti_spyware.pdf

[4] Jong-Hyuk Roh, et, al, "Privacy Authorization for Internet Identity Management System", Journal of KICS, Vol.30, No.10, 2005

[5] G. Karjoth, et, al. "Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data", LNCS 2482, 2002

[6] L. Cranor, et, al. "The Platform for Privacy Preferences 1.0 Specification", W3C, 2003

Daeseon Choi is a member of the research group of the "digital identity research team" of ETRI. His major filed of interest lie in the security mechanisms in the distributed environments. He received his diploma in computer science from the POSTECH, Korea in 1997. Since then he has worked on research projects at ETRI and continuously conducted numerous projects on national and international level.

Seunghun Jin is a manager of the research group of the "digital identity research team" of ETRI. His major filed of interest lie in the digital identity management, PKI, PMI, and authentication/authorization. He received his diploma in computer science from the Soongsil University, Korea in 1995. In 2004 he received his PhD in computer science for a national PKI development. He has worked several projects in commerce. Since the 1999 he has joined on research projects at ETRI and continuously conducted numerous projects on national and international level.

Hyunsoo Yoon is a professor of EECS Dept, KAIST. His major filed of interest lie in the mobile adhoc networking, wireless sensor networking, and multimedia virtual network. He received his PhD in computer science from the Ohio State Univ. in 1988. He is member of IEEE Computer Society and Communication Society. He is also member of ACM SIGARCH and SIGCOMM.