International Journal of Fuzzy Logic and Intelligent Systems, vol. 11, no. 3, September 2011, pp. 197-203
http://dx.doi.org/10.5391/IJFIS.2011.11.3.197

Standard Implementation for Privacy Framework and Privacy Reference Architecture for Protecting Personally Identifiable Information

Yong Nyuo Shin

Hanyangcyber University, South Korea

Abstract

Personal Identifiable Information (PII) is considered information that identifies or can be used to identify, contact, or locate a person to whom such information pertains or that is or might be linked to a natural person directly or indirectly. In order to recognize such data processed within information and communication technologies such as PII, it should be determined at which stage the information identifies, or can be associated with, an individual. For this, there has been ongoing research for privacy protection mechanism to protect PII, which now becomes one of hot issues in the International Standard as privacy framework and privacy reference architecture. Data processing flow models should be developed as an integral component of privacy risk assessments. Such diagrams are also the basis for categorizing PII. The data processing flow may not only show areas where the PII has a certain level of sensitivity or importance and, as a consequence, requires the implementation of stronger safeguarding measures. This paper propose a standard format for satisfying the ISO/IEC 29100 "Privacy Framework" and shows an implementation example for privacy reference architecture implementing privacy controls for the processing of PII in information and communication technology.

Keywords : Personally Identifiable Information, Privacy Framework, Privacy Reference Architecture, Privacy Protection

1. Introduction

Privacy is a fundamental human right. In the US, the designation “personal information” is limited to information in a particular area, such as medical information, social security number, and bank information. However, other countries, including the European countries, define personal information more broadly as “all personally identifiable information”. Many countries have enacted and implemented various laws and regulations to protect personal information, including Greece, Finland, U.K., France, Canada, Australia, and New Zealand. The efforts of the international community to protect personal information started with “Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data,” which was recommended to member countries by the OECD in 1980. The UN announced the personal information computerization guideline, and the EU published the guideline regarding personal information processing and protection. The OECD’s privacy protection guidelines[1] reflected the consensus of the international community regarding personal information collection and management, and significantly affected the privacy protection laws of many countries. The OECD’s personal information protection guideline defines the scope of application in such a way that the guideline should be applied to sensitive personal information in the private and public sectors, in addition to personal information related to computing processing. Furthermore, the guideline stipulates that it can be deemed as a minimum requirement, and can be supplemented by additional measures. Besides, APEC [2] has proposed privacy violation prevention, limitation on collection, information use, selection of the information subject, notice, updating, safety, accessibility, and privacy principle in line with responsibilities[3].

Since 2005, ISO/IEC JTC1 SC27 Working Group 5 has been performing standardization to protect privacy, the fundamental right of the individual, and concentrating on the standardization of a privacy reference architecture to implement the privacy framework. The privacy protection market is likely to be expanded significantly, as the Privacy Protection Law will be enforced soon, and the personal information of corporate employees, which previously wasn’t a regulatory subject, will be regulated, in addition to paper-type personal information. The scope of regulation will also be expanded to information and communication, education, medical service, and the financial area. For this reason, it is time for the governmental/public agency and private company to make thorough preparations. Protecting personally identifiable information is protecting the basic rights of the public, and is closely related to the concept of protecting personal assets in the knowledge society. As many privacy violation cases have been reported at home and abroad, such as the collection of user location information via smartphones and Google’s street view, the privacy reference architecture has been drawing attention, in order to create a privacy framework, which is the

Manuscript received Jul. 29, 2011; revised Aug. 23, 2011; accepted Aug. 25, 2011.
This work was supported by the Inustrial Strategic Technology Development Program, 10039149, Development of Basic Technology of Human Identification and Retrieval at a Distance for Active Video Surveillance Service with Real-time Awareness of Safety Threats funded by the Ministry of Knowledge Economy (MKE, Korea)

international standard to protect privacy, and implement the framework.

The privacy framework[4] is intended to help an organization to define its privacy control requirements related to personally identifiable information within its information and communication technology environment by: relating all described information privacy aspects to existing security guidelines. The privacy reference architecture[5] provides guidelines on how to develop, implement and operate information and communication technology systems with built-in privacy safeguarding controls; is a resource containing a consistent set of architectural best practices for managing PII in information and communication technology systems; and extends on the privacy framework derived from ISO/IEC 29100.

Now, it's time to understand the trend of international standardization in the area of privacy, improve the efficiency of the domestic Personal Information Management System and related policy, and set up an international standardization strategy driven by Korea in the privacy area. The Personal Information Management System is a voluntary certificate in Korea that is granted to organizations satisfying the requirements to prevent PII disclosure. PII impact analysis is the procedure that allows the evaluation and improvement of privacy violation factors in advance. Compared with post handle for the privacy infringe, business can be promoted efficiently and the budget can be reduced.

This study will implement the policy-based operating software and apply it to the actual operating environment, which satisfies the safeguard control proposed by the privacy framework and privacy reference architecture, and introduce and apply the Personal Information Management System.

2. Basic Elements of the Privacy Framework

To handle properly and protect PII in designing, implementing, operating, and maintaining information communication systems, we should define the actors which are involved in data processing life cycles as shown in figure 1[5].

2.1 Actor

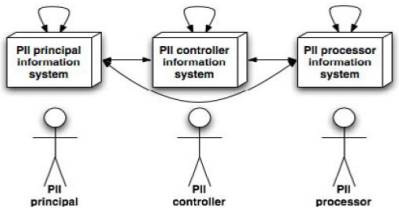

Fig. 1. The actors and their interactions

In PII processing, the data processing life cycle starts with the collection of PII from the PII principals[6], third parties or from PII which for other purposes is already under the control of the PII controller. Therefore, it can be subsumed that when dealing with privacy issues in the context of information and communication technology, there are generally two main actors involved: the PII principal and the PII controller. However, for conceptualizing the processing of PII in this framework several roles of these actors can be differentiated. The roles of PII provider and PII recipient are visualized in Table 1.

Table 1. Possible flows of PII between different actors

	PII provider		PII recipient
Role (a)	PII provider & PII controller	→	PII recipient
Role (b)	PII principal & PII controller	→	PII processor
Role (c)	PII principal	→	PII controller & PII processor

The PII processor has a specific role in these scenarios. He executes the collection, processing, use and transfer of PII on behalf of and under the supervision of the PII controller. The PII processor is bound by legal contract to execute exactly those processing steps the PII controller has stipulated, to subject itself to the control of the PII controller and, possibly, relevant regulating agencies, to observe the stipulated privacy requirements and to implement the corresponding privacy controls. There is a need to distinguish between PII processors and third parties because the legal control over the PII remains with the original PII controller when this PII is turned over to the PII processor, whereas a third party becomes PII controller in its own right once it has received the PII in question. Table 2 describes the case where the PII controller involves a third party to whom it may contract out some of the processing. In many cases the roles of PII controller and PII principal do not coincide.

Table 2. Possible flows of PII between PII controller and a third party

	PII controller		Third party
Role (d)	PII provider	→	PII recipient
Role (e)	PII provider	→	PII processor
Role (f)	PII recipient	←	PII provider

2.2 Recognizing PII

In order to recognize data processed within information and communication technologies as PII, it should be determined at which stage the information identifies, or can be associated with, an individual. Examples of personally identifiable information are given in Table 3. Depending on the context, some items such as gender or age would only constitute PII if associated with other PII such as name or e-mail address. PII also should take into account the context of the individual's privacy preferences, within which the PII is collected, used, transferred, stored, archived, or disposed of by a PII controller or PII processor.

When processing PII within information and communication

technologies, it should be determined at which stage the information identifies an individual. Certain types of PII, shown in Table 3 as unique single identifiers, by themselves can lead to the identification of an individual. The ease of being able to access or search a particular user's PII can also determine the identifiability of the individual. In some instances, a single set of PII as listed in the table above as "Other PII" cannot identify an individual. However, if someone aggregates the PII to build a profile of an individual, he/she can be identifiable by this profile. Organizations should assess the possibility and likelihood of unauthorized entities identifying an individual by combining publicly accessible information with PII being collected and processed and ensure that appropriate control measures are taken.

Table 3. Examples of Personally Identifiable information[4]

	Examples
Unique information explicitly identifiable as PII by itself	National identifiers(e.g. passport number) Customer number Biometric identifier Bank account or credit card number
information not identifiable as PII by itself	Name Gender Date of birth Home address Personal telephone number Personal e-mail address IP address Photograph or video identifiable to an individual Trade-union membership Sexual orientation Criminal convictions or committed offences Financial profile Personal identifiable number(PIN) and Passwords for financial accounts Any information collected during health services Disabilities Racial or ethnic origin Religious or philosophical beliefs Age or special needs of vulnerable individuals Personal or behavioural profile Location derived from telecommunications systems Product and service preferences derived from Customer Relationship Management (CRM) systems Personal interests derived from tracking use of internet web sites Combinations of specific PII

Some PII can become anonymous if it is aggregated and stripped of its connection to an individual. Conversely, anonymous data can become PII when it is correlated with specific PII of the individual. Information in an anonymous form is not considered to be PII. However, even in an anonymous set of data, the smaller the group within the anonymous set, the greater the likelihood of an individual being identifiable.

PII may be stored in an information and communication technology system in such a way that is not readily visible to other system users. Examples include the author's name stored as metadata in the properties of a document, comments or tracked changes stored as metadata in for example a word processing document, and PII stored in a cookie. The PII principal may not want this information to be distributed publicly. information and communication technology systems should support mechanisms that will make the PII principal aware of such PII and provide the individual with appropriate control over the sharing of that information.

PII may also be stored in an information and communication technology system unsolicited by the PII recipient. An example of unsolicited PII could include PII entered in a Web form by the PII principal that the PII recipient did not request nor seek to collect. The entry of unsolicited PII can be minimized by using fields with predefined entries wherever possible. When a text field is necessary, the User Interface should discourage the PII principal from entering PII in scenarios where the recipient intends to collect only non-PII.

2.3 Establish a Privacy Management System

One of the key success factors for entities trying to manage the importance and complexity of the different facets of privacy in information and communication technology is the establishment of a privacy management system within the organization's information and communication technology systems. An effective privacy management system impacts people, processes and technology, is part of the internal control program and risk mitigation strategy of an organization, and its implementation helps to satisfy compliance with data protection and privacy regulations.

The following are key considerations when establishing a privacy management system:

Table 4. Key considerations for establishing privacy management systems

Key factors for establishing privacy management system	Description
Policies	The PII recipient should set a clear policy concerning the collection, use, transfer, storage, archiving and disposal of PII, aimed at maintaining internal and external requirements, that is binding for every person in charge and every employee handling PII. Every procedure in respect to these tasks is to be evaluated as to its compliance with the defined policy. The PII recipient should also set clear policy and procedures concerning the collection, transfer, usage, storage, archiving

	and disposal of pseudonymous data.
Inventory	Except when the PII is unsolicited, the PII recipient should establish a process to categorize any incoming data as regular data, PII data or sensitive PII data, thus allowing for a separation of PII data at the earliest possible time. The PII recipient should also create and maintain an inventory of all PII processes. By this means the appropriate handling of PII can be ensured from the start. Furthermore, for each data process such as the collection, use, transfer, storage, archiving, and disposal of PII, appropriate data handling procedures should be defined.
Procedures and Controls	Technical and organizational guidelines for privacy should be developed and implemented with regard to the aforementioned privacy principles.
Governance	In order to ensure that privacy principles are adhered to and controls satisfy the specified privacy requirements, a governance and internal control authority should be established. The internal control authority should establish regular privacy assessments as well as key privacy areas to be audited regularly within the organization's internal and external audits to ensure the compliance with privacy rules and regulations and, if included in the overall governance scheme, compliance with this International Standard. The person responsible for all processes involving the handling of PII could either take on the internal control authority him/herself or assign the authority to another function such as the internal audit department or an external auditor.
Compliance	Entities that receive and process PII should develop and maintain privacy assessments to evaluate whether program and service delivery initiatives involving the collection, use, transfer, storage, archiving, or disposal of PII comply with data protection and privacy requirements and to resolve privacy issues that could be of potential public concern.
Documentation	Records are needed for dispute resolution, auditing, and other privacy management purposes. Records should be generated when consent is obtained from PII principals, when PII is received and/or altered, and when PII is passed to another PII recipient. Records should be protected against unauthorized access, alteration, and deletion. Additional records could be needed to meet local legal or auditing requirements.
Training and Awareness	Organizations should ensure their employees, customers and other individuals are aware of their responsibilities when processing PII.

3. Supporting rules and policy to protect privacy, and implementing privacy self-diagnosis tool.

3.1 Standardization on privacy framework and privacy reference architecture

Of the privacy-related standardization works, Working Group 5 of the SC27 Committee under ISO/IEC JTC 1 involves the standardization of identity management and privacy technology[7]. The Working Group is leading standardization for the consistent implementation of personal identity information classification and system installation to protect privacy, such as the privacy framework and the privacy reference architecture[9].

SC27, which used to manage information security techniques using encryption, and evaluate these techniques, is expanding the scope of its work to the privacy area related to personal information protection, biometric information protection, and identity management, in order to cope with the requirement of standardization of information security techniques, which is caused by the development of the information and communications technology[8]. It is believe that this change will be of great help in resolving social issues related to privacy and personal information protection that are raised in various countries, including Korea.

3.1.1 Relation between privacy rules and policies

The privacy protection software applies the rule to apply the privacy control to the system, in order to satisfy the requirement proposed by the privacy framework and reference architecture. The rule must always be included in the policy. If the rule is not included in the policy, it cannot be transferred to the agent. Applying the rule means including the rule in the particular policy. A regular expression changes the particular set of characters or the string into symbols, and is used to define the expression rule used to describe a set of strings accurately, or to define the grammar of the language, or to designate the string to search.

Rules are managed, such as addition, modification, deletion, change, and application to the policy. The content of the rule is a regular expression or keyword, which is included in the policy and sent to the agent, and is used by the agent to detect a regular expression and keyword designated by the file in the agent PC, based on the regular expression and keyboard in question.

Policies and rules have a 1..n relation, as shown in Figure 2.

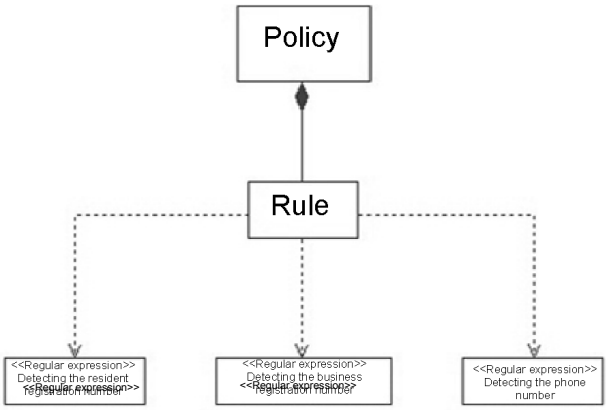

Fig. 2. Relation between rule and policy

When entering a regular expression, the expression that fits into the standard regular expression should be entered. The agent will not execute the expression automatically, if it is not suitable for the regular expression such as * and ?. Furthermore, the personal information will not be detected properly if the expression is incorrect. If the rule has been transferred to the agent already, because the rule in question is included in the policy when modifying, deleting, or applying the rule, the changed rule will be sent to the agent and applied, if the agent in question is online. If the agent is offline, the changed rule is sent to the agent when the agent connects.

3.1.2 Personal information detection and protection technique

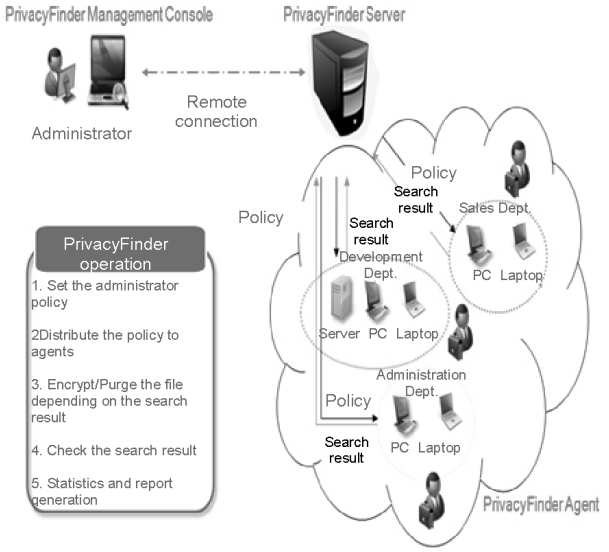

Fig. 3. Composition of the privacy protection system

As shown in Figure 3, the system is implemented in such a way that the agent personal computer is searched immediately by the central console. The search contents and processing result are generated by department or user, depending on the security management policy. The policy applied to the agent PC limits the CPU use of the PC in question according to the importance of the work, to the extent that it is possible to do so, and searches for PII in all drives of the agent PC. If any PII is found, it is sent to the process server. The agent repeats the process when searching all drives is finished, until the new policy is received. The repetition interval of the search work is basically one hour but can vary, depending on the number of files in the agent PC and the system performance. First, the corresponding policy is sent to the process server without condition when connecting to the agent's process server. Second, the policy will be applied immediately, if the agent is online when the rule or policy applied to the agent by the administrator is modified. If the agent is offline, the policy is applied when the agent goes online by connecting to the process server.

In order to implement and test the ISO29101[5] presented previously, the system was implemented that performs various functions as described below. Searching the data inside the deleted file is supported, in order to support e-discovery to enable the enterprise to cope with unauthorized deletion by the individual, using the forensic function of the privacy protection solution. Pattern check is designed to identify the pattern or word from the file, using the regular expression pattern or word. The searchable PII in the user's PC is created as a regular expression by the authorized super user, and applied to the PC in question as a search policy. Therefore, almost all PII that can be described in a regular expression can be searched. Table 5 shows the example of the regular expression-based pattern to extract the searchable PII.

Table 5. Regular expression-based pattern to extract the searchable PII

Regular expression	Description
/^\d\d[0-1]\d[0-3]\d-[1-4]\d{6}$/	Resident registration number search
(^0[1-9]{1,2})-([1-9][0-9]{1,3})-([0-9]{4})	Phone number search
Security|Secret|Destruction	Specific word search

To comply with the ISO29110 international standard, special solutions and technologies must be applied to process and identify multiple languages. Technical concepts for multiple languages can include character encoding, language identification, and tokenization. Character encoding enables the computer to recognize English and other characters, using the code page and Unicode method. The language identification capability is the core element required to prepare the document to review in advance. The document review classification group is composed by language, and saved in the proper cluster. Finally, tokenization is the process of identifying words and sentences. Tokenization requires lexical support because language formats are diverse. For example, Asian languages may not contain a blank between words. Therefore, various language searches should be supported when processing the vocabulary.

In addition, the complete purging function is implemented as

shown in Figure 4, as recommended by the National Security Agency. When this method is used, the file is overwritten with special characters more than three times in order to erase its contents permanently.

Fig. 4. Encryption and permanent deletion

Strictly complying with the regular maintenance schedule for document archiving can be helpful for smooth system operation. In addition, electronic signature should be separated from encrypted signature in the file encryption function. Even when a document is encrypted and signed for the purpose of security, it requires an electronic signature most of the time, rather than a compliance scenario requirement. This responsibility should be imposed on the person who modified the document, together with the time stamp and reason for the change.

One of the major purposes of privacy protection is to check whether the privacy framework standard is being conformed with or not. The system is implemented to execute self-diagnosis using the agent, as shown in Figure 5, in order to make the framework to comply with these standards.

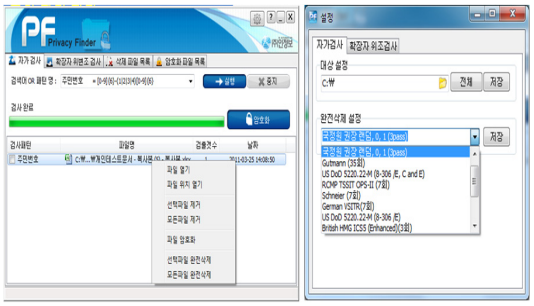

Fig. 5. Self-diagnosis using the agent

4. Implementation of the response according to personal information search

The PC with an agent installed receives the PII search policy from the authorized super administrator, and applies the policy. If the target PII is found, the search details will be sent to the process server and saved in the MS SQL server. Table 6 shows the result of searched personal information.

Table 6. The result of searched PII

Security attributes	Description
Agent ID	Agent user's identifier
Search time	Time at which the personal information is searched
Policy	Identifier of the personal information search policy
Rule	Regular expression identifier used to search the personal information
Search file information	Other information such as the position, size, and format of the file containing the personal information
Search times	Number of times that the personal information is searched
Search content	Paragraph in the file containing the searched personal information
Search result	Successful/Failed research
Response result	Successful/Failed deletion of the searched target file

In addition, various policies can be set, and statistics can be checked by date and user, using the administration console, as shown in Figure 6.

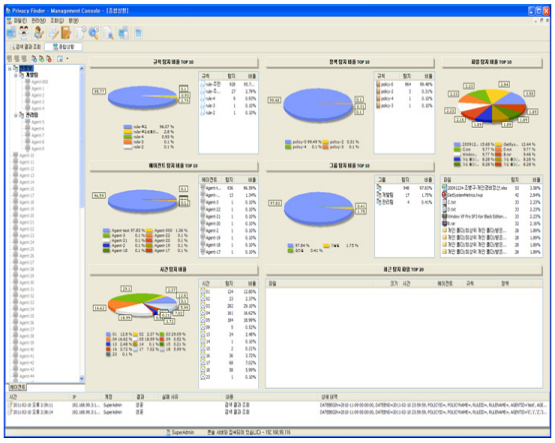

Fig. 6. Administration console for checking by date and user

As described above, the process described in this paper can be used to determine whether the entire process of efficient policy establishment regarding the management of the important data in the enterprise is running normally as described in Figure 6, based on the implemented system. An efficient policy will prevent the exposure of the personal and confidential information of the enterprise, encrypt the personal information file or delete it permanently, and manage the status of the personal information and confidential information. Through self-diagnosis of the user, the user's awareness about protecting the important information saved in the business PC can be enhanced, and the privacy protection obligation can be carried out, which is required by the Privacy Act.

5. Conclusions

Protecting personally identifiable information is protecting the basic rights of the public, and is closely related to the concept of protecting personal assets in the knowledge society. As many privacy violation cases have been reported at home and abroad, such as the collection of user location information via smartphones and Google's street view, the privacy reference architecture has been drawing attention, in order to create a privacy framework, which is the international standard to protect privacy, and implement the framework.

The privacy framework is intended to help an organization to define its privacy control requirements related to personally identifiable information within its information and communication technology environment by: relating all described information privacy aspects to existing security guidelines. The privacy reference architecture provides guidelines on how to develop, implement and operate information and communication technology systems with built-in privacy safeguarding controls; is a resource containing a consistent set of architectural best practices for managing PII in information and communication technology systems; and extends on the privacy framework derived from ISO/IEC 29100.

Now, it's time to understand the trend of international standardization in the area of privacy, improve the efficiency of the domestic Personal Information Management System and related policy, and set up an international standardization strategy driven by Korea in the privacy area.

The Personal Information Management System[12] is a voluntary certificate in Korea that is granted to organizations satisfying the requirements to prevent PII disclosure. PII impact analysis is the procedure that allows the evaluation and improvement of privacy violation factors in advance. Compared with post response, business can be promoted efficiently and the budget can be reduced.

In this paper, we proposed a policy-based operating tool and apply it to the actual operating environment, which satisfies the safeguard control proposed by the privacy framework and privacy reference architecture.

Close cooperation and active participation in standardization activities are required to perform the important tasks that can create a new market in the privacy protection industry. Therefore, priority should be given by domestic enterprises, government-supported research centers and academia to proactive and continuous interest and participation in standardization activities related to privacy protection.

References

[1] http://www.oecd.org/document/18/0,3746,en_2649_34255_1815186_1_1_1_1,00&&en-USS_01DBC.html, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data", 1980.

[2] Asia Pacific Economic Cooperation, *"APEC Privacy Framework"*, 2005.

[3] http://www.law,cornell.edu/rules/frcp/index.html# chapter_v, "Federal Rule of Civil Procedure".

[4] ISO/IEC JTC1 SC27 "*Privacy Framework*", *SC27 N9226*, 2011.

[5] ISO/IEC JTC1 SC27 "*Privacy Reference Architecture*", *SC27 N9228*, 2011.

[6] ISO/IEC JTC1 SC27 WG5 "*Study Period for a harmonized SC 27/WG 5 Vocabulary"*, *SC27 N9401*, 2011.

[7] ISO/IEC JTC1 SC27 WG5 *"WG 5 SD1-WG 5 Roadmap*", *SC27 N9237*, 2011.

[8] ISO/IEC JTC1 SC27 "*Business plan for JTC1 SC27 Security Technique*", SC27 *N9463*, 2010.

[9] ISO/IEC JTC1 SC27 "*Resolutions of the 11th meeting of ISO/IEC JTC 1/SC 27/WG 5 in Singapore, April 11-15, 2011", SC27 N9920,* 2011.

[10] HomelandSequrity Whitepaper, "*Computer Network Security & Privacy Protection*", 2011.

[11] http://www.cs.ucdavis.edu/~hchen/paper/passat09. pdf, "Noise Injection for Search Privacy Protection", 2011.

[12] http://isms.kisa.or.kr/kor/main.jsp, "Personal Information Management System", 2011.

She received the PhD degree in computer science from Korea University in 2008, Republic of Korea. Currently, she is a professor at Department of Computer Science, Hangyang Cyber University. Also, she is an editor for efforts and continued support in progressing the many standardizations such as ITU-T SG17, ISO/IEC JTC1 SC27 and SC37. Her current research interests are telebiometrics, authentication technologies and privacy.