International Journal of Fuzzy Logic and Intelligent Systems, vol. 12, no. 1, March 2012, pp. 53-59
http://dx.doi.org/10.5391/ IJFIS.2012.12.1.53 pISSN 1598-2645 eISSN 2093-744X

Implementation Privacy Reference Architecture for Forensic Readiness

Yong-Nyuo Shin*

Department of Computer Engineering, Hanyangcyber University, Seoul 133-791, Korea

Abstract

As the Privacy Act is in force in Korea, the subject of protection responsibility is increased, and continuous efforts are made to protect privacy in overseas countries, as can be seen by standard drafts related to privacy protection. However, the reality is that a formal privacy manual or guidelines are insufficient to help cope with the rapid changes and privacy leak caused by TGIF(Twitter-Google-iPhone-Facebook) these days, and practical effects cannot be expected, even though measures are taken. This paper propose a standard format for satisfying the ISO/IEC 29101 "Privacy Reference Architecture" and shows an implementation example for equipping with forensic readiness capturing indications of the incident rapidly and coming up with an effective counter measure when privacy information is disclosed.

Keywords : Forensic Readiness, Personally Identifiable Information, Privacy Framework, Privacy Reference Architecture

1. Introduction

The “Privacy Act” was enacted in Korea to promote the rights and interests of the people by regulating the collection, disclosure, abuse, and misuse of personal information, and to define regulations regarding personal information processing to realize the dignity and respect of the individual. The law is applicable to all personal information handlers in the private and public sector, and defines the protection criteria by phase (collection, use, and provisioning of the person information) to protect the rights of the information subject, providing the basis for constraining the installation of an image information processing device by increasing the constraints of the unique identification information. It defines the introduction of a personal information impact assessment and the notification and reporting of a privacy information leak. The term “unique identification information” in Article 24.1 in the Privacy Act refers to identification information that is assigned to each individual for identification by law, such as resident registration numbers, passport numbers, and foreigner registration numbers. In particular, the resident registration number is used as a key to link and integrate the data among information systems in Korea. The collection ratio of the resident registration number is 40.7% in the public sector, and 51.5% in the private sector[1].

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.

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. 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[2]. 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.

It is important to improve the efficiency of the domestic PIMS(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 personal information disclosure. Personal information 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

Manuscript received Dec. 10, 2011; revised Dec. 24, 2011; accepted Dec. 30, 2011

* Corresponding author: Yong-Nyuo Shin(ynshin@hycu.ac.kr)

This work was supported by the research fund of Hanyang Cyber University(HYCU-2010-0017)

ⓒ The Korean Institute of Intelligent Systems. All rights reserved.

software and apply it to the actual operating environment, which satisfies the safeguard control proposed by the privacy reference architecture, and introduce and apply the Personal Information Management System. Also this paper provides the security management methodology and tool equipped with forensic readiness capturing indications of the incident rapidly and coming up with an effective counter measure when privacy information is disclosed.

Digital forensics is a paradigm of scientific investigations that analyze digital evidence to prove various assertions, and is accepted as evidence in court[3].

To present a security management methodology equipped with forensic readiness, changes are required from technical limitations to behavioral limitations, from post measures against privacy information disclose to capturing an indication of the incident, and from regular system surveys for privacy protection to constant operation. Most of all, a paradigm change is required from the structure of limiting control to the overall prohibition of privacy information use other than for the original purpose of collection.

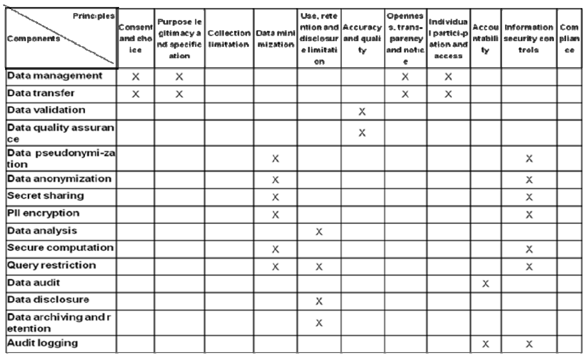
This paper is composed as follows. In chapter 2, the privacy reference architecture of ISO/IEC JTC1 SC27 WG 5 will be reviewed. Chapter 3 describes a proposed forensic readiness skim for personal data protection. Chapter 4 shows the implementation privacy reference architecture for forensic readiness. In Chapter 5, conclusions will be given.

2. Privacy Reference Architecture

A privacy reference architecture requires a context for its implementation and deployment[4]. The instantiation of architecture relies on a fabric of policy inspired business management functions, processes and procedures that operate in harmony with the architecture to deploy the privacy aware and privacy enabled ICT system. Business processes and the respective PII that is processed need to be examined in order find appropriate technical safeguards, as part of a more comprehensive overall solution that can fulfill the identified privacy safeguarding requirements.

Any ICT system processing PII should have the following basic features for managing local data. This includes data entry, access, update and removal. When needed, the data management system should be able to support a continuous process that provides for the collection of data over a longer period of time. For example, the PII principals may send updates to the PII in the future. The data management features in the ICT system of the PII principal are focussed on collecting PII and transferring it to other actors. The data management component of the PII controller's ICT system should support the management of local data. Additionally, it should be capable of exchanging data with the PII principals' ICT systems (data collection), PII processors's ICT systems (to delegate processing) and optionally, the ICT systems of other controllers or third parties.

Table 1 The relations between privacy principles and the components in the identity management layer.



Principles / Components	Consent and choice	Purpose legitimacy and specification	Collection limitation	Data minimization	Use, retention and disclosure limitation	Accuracy and quality	Openness, trans-parency and notice	Individual partici-pation and access	Accountability	Information security controls	Compliance
Data management	X	X					X	X			
Data transfer	X	X					X	X			
Data validation						X					
Data quality assurance						X					
Data pseudonymi-zation				X						X	
Data anonymization				X						X	
Secret sharing				X						X	
PII encryption				X						X	
Data analysis					X						
Secure computation				X						X	
Query restriction				X	X					X	
Data audit									X		
Data disclosure					X						
Data archiving and retention					X						
Audit logging									X	X	

The data management component of the PII processor's ICT system enables the basic management of received data. Note that the privacy policy, the use of various Privacy Enhancement Technologies and other factors may restrict the data management tools available in the PII processor's ICT system. For example, the PII processor's ICT system may be forbidden from adding or linking PII with other information. The data transfer component should exchange data, including PII, with the ICT systems of other actors. Optionally, identity services and PII encryption may be integrated with data transfer so that the endpoints can authenticate each other and communication is confidential. The PII that is being checked should be validated for correctness of format, as well as accuracy and timeliness of the values. The component should have sufficient information about the data model and the ranges of individual values in order to warn the user about possible errors in data input. Pseudonymization is a technology for transforming data so that direct identifiers are replaced with artificial values. The pseudonymization component in the data layer uses the pseudonymization schemes described in the identity management layer to replace identifiable information in the data with identifiers that do not reveal the respective PII principal's identity on their own. The anonymization process inspects the data set and modifies the values so that global statistical properties remain, but individual values are changed. Anonymized data requires an understanding of the statistical properties of the underlying dataset so this component should be capable of performing statistical analysis.

Secret sharing is a technique for distributing PII values into shares that by themselves reveal no information about the original value. Secret sharing can be used for distributed data gathering to reduce the risk of privacy breach. Secret sharing provides better privacy when performed at the ICT system of the PII principal and used in conjunction with secure multiparty computation. Secret sharing can be used to reduce the risk of insider attacks, as a party with access to a share of a PII value cannot learn the original value from it. This makes insider attacks significantly more complex. For optimal result, secret sharing requires that there is more than once instance of each actor in the system. Each instance should store and process just some shares. The PII encryption component should encrypt PII before it is stored. Depending on the privacy safeguarding requirements, the encryption keys can be shared between ICT systems so each of them can decrypt the data and access it appropriately. If a secure computation technique that is capable of processing encrypted PII is used, the information does not have to be decrypted in order to be processed. If the PII processor's ICT system is tasked with data analysis, it should implement a component performing containing the necessary methods. The goal and construction of this module is heavily dependent on the data processing task being solved.

Secure computation can be used to let PII processors process PII without having access to the raw input values. Instead, secure computation techniques perform computations on PII that has been transformed by PETs such as encryption or secret sharing. Secure computation can reduce the risk of PII leaks from the ICT system of the PII processor, as PII is not provided to the processors in a clear form. The data audit component provides an overview of the PII stored in the ICT system. Information from the PII categorization system can be used to categorize the data values stored in the system. The identity management system can be used to determine the principal related to a particular item of PII. Altogether, the component should be capable of providing the user of the ICT system with at least the following metrics: firstly, the amount of PII in the system (number of records per record type); secondly, the number of PII principals who have provided information.

The data disclosure component should prepare data before it leaves the PII controller's ICT system. For each transfer or disclosure event, the component should log data audit results The audit logging component should log each transaction performed on PII, regardless of the action taken. This component should be integrated with every other component so that each component could log specific activities. The logging module requires integration with the authentication module to log the identity of the party that accessed PII. Optionally, secure logging techniques can be used to prevent tampering with the log entries.

3. Forensic Readiness for Personal Data Protection Act

A forensic investigation of digital evidence is commonly employed as a post-event response to a serious information security incident[5]. In fact, there are many circumstances where an organisation may benefit from an ability to gather and preserve digital evidence before an incident occurs. Forensic readiness is defined as the ability of an organisation to maximise its potential to use digital evidence whilst minimising the costs of an investigation. The costs and benefits of such an approach are outlined. Preparation to use digital evidence may involve enhanced system and staff monitoring, technical, physical and procedural means to secure data to evidential standards of admissibility, processes and procedures to ensure that staff recognise the importance and legal sensitivities of evidence, and appropriate legal advice and interfacing with law enforcement. Information security programmes often focus on prevention and detection measures. From a preventative information security perspective there is little need for digital evidence. From a business perspective, however, there are a number of scenarios where collecting appropriate digital evidence would be beneficial. Thus, there is a business requirement for digital evidence to be available even before an incident occurs. What exactly this requirement is, how it is met, and how organisations can exploit digital evidence has not previously been considered in detail. Digital evidence is required whenever it can be used to support a legal process. An organisation, therefore, requires access to the evidence that will be able to support its position in such an event. This is not as easy as it might seem; relevant evidence is unlikely to exist by default. In any computer security incident there will be a tendency to focus on containment and recovery, as these are the foremost business critical issues. However, in stressing these, any evidence that might be required may be damaged, discarded or simply ignored [6]. There is a trade-off to be made between recovery and evidence. A lot of information is also lost or discarded as part of normal business practice. To succeed in a legal process, it is therefore essential that the organisation has actively gathered the evidence it is likely to require[7]. Moreover, it is vital to have the capability to process evidence cost-effectively, and to have suitably trained staff who know how to ensure potential evidence is preserved. In korea, as Personal Data Protection Act which regulates on principle of collecting and utilizing privacy information for the purpose of protecting information privacy effectively and systematically is legislated lately, privacy information is handled only with restrict cause of data subjects' consent or code of law and ordinance. But, not only enough acting PDPA but also forensic readiness effort is require for banking system and government infrastructure to prevent privacy infringement.

4. Implementation Privacy Reference Architecture for Forensic Readiness

4.1 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[8]. 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. 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.

4.1.1 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 1.

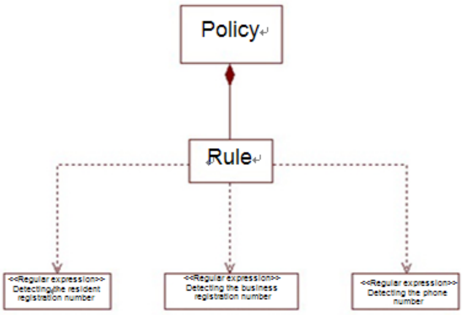

Policy
Rule
<<Regular expression>> Detecting the resident registration number
<<Regular expression>> Detecting the business registration number
<<Regular expression>> Detecting the phone number

Fig. 1: 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.

4.1.2 Personal information detection and protection technique

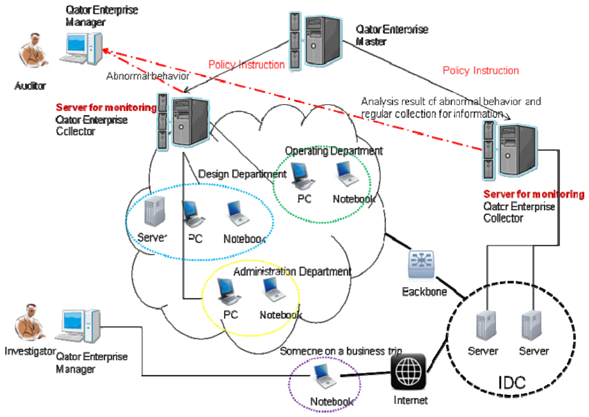


Fig. 2: Composition of the privacy protection system

As shown in Figure 2, 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 personal information in all drives of the agent PC. If any personal information 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 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 personal information 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 personal information that can be described in a regular expression can be searched. Table 2 shows the example of the regular expression-based pattern to extract the searchable personal information.

Table 2 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.

4.2 Privacy Protection System equipped with forensic readiness

Reviewing the function of components in figure 3, Qator Master performs collector management, researcher registration and management, research authentication, DB storage and query processing, and manager supervision. Qator Manager investigates the target PC, checks and reports various collection data, and performs an interface with the forensic tool. Qator Master instructs registration and business processing to Collector. Agent is installed on each individual PC for job processing.

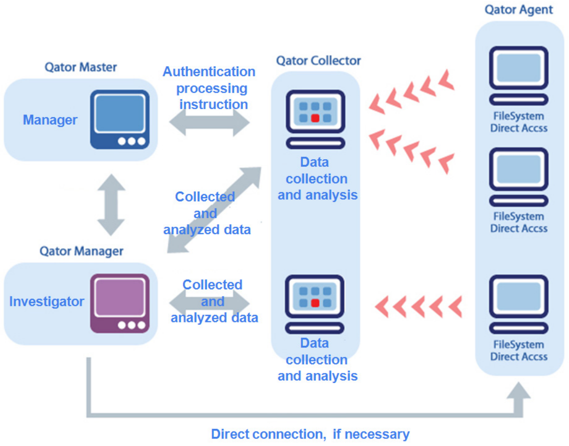

Qator Agent
Qator Master
Authentication processing instruction
Qator Collector
Manager
Data collection and analysis
FileSystem Direct Accss
FileSystem Direct Accss
Collected and analyzed data
Qator Manager
Investigator
Collected and analyzed data
Data collection and analysis
FileSystem Direct Accss
Direct connection, if necessary

Fig. 3: Function of components the Qator Enterprise

To present a security management methodology equipped with forensic readiness, diverse functions are required for providing the forensic readiness capturing indications of the incident rapidly and coming up with an effective counter measure when privacy information is disclosed, as shown in table 3.

Table 3 Diverse functions of privacy protection system

Function	Description
Case management	Cases are managed remotely, while researchers share the case when conducting research. Reports can be managed at one place.
constant data collection	The necessary data can be collected constantly, depending on the policy, even though researchers haven't performed research.
Remote system investigation	Remote client investigation using an agent
File system analysis	File system analysis through direct access to the disk
Recovery of Removing client files	Removed client files can be studied after recovery
File Distribution	Distribution check of the particular file by an individual client
Analysis of file time zone	Distribution charge by file creation date, changed date, and accessed date
Search	File content search by a particular keyword or pattern
Remote disk imaging	Imaging that supports EnCase analysis of the client storage media for in-depth data research
Collection of user environment	Collecting current user system status and volatile data as a means of continuous research
Automatic imaging work for the USB storage device	Automatic imaging work for the USB storage device, and occasional monitoring of the USB media use details
Snapshot for file system	Saving and analyzing the file system status at certain points
Bookmark	Saving the necessary work details during the work
Export of the report and data	Export functions are available for all data that can be checked, and the report integrating all contents can be created separately
View	Supporting the original format, text, and hex view of the file
simultaneous research	Supporting simultaneous research on several clients

The file containing the intended keyword and pattern can be searched from various files saved in the storage device remotely, as shown figure 4.

Various events created in the research target PC can be investigated, and investigations of the storage media in the target PC can be performed on the Disk Check tab.

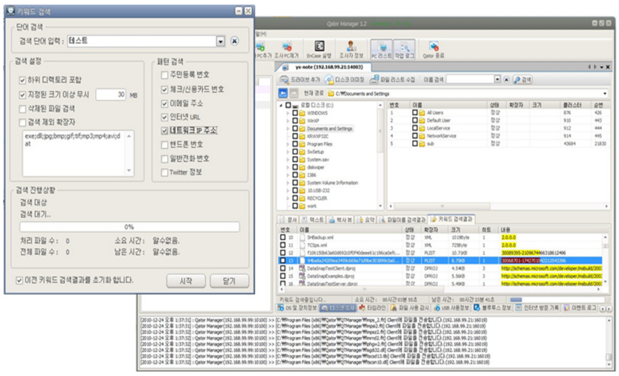


Fig. 4: Remote keyword and pattern search

In addition, various policies can be set, and statistics can be checked by date and user, using the administration console, as shown in figure 5.

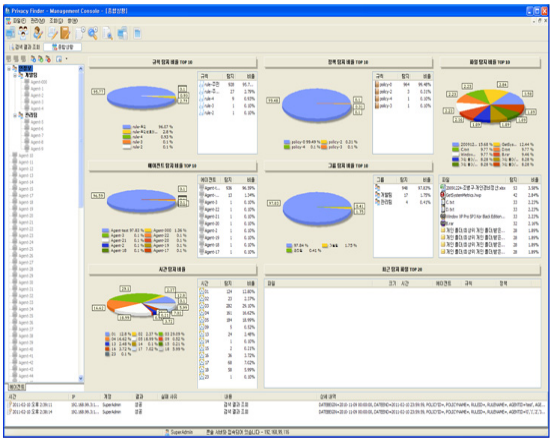


Fig. 5: Administration console for checking by date and user

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 5, 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. Also, the event display function by time zone is provided for the researcher to observe the data effectively. The researcher can understand what happened at a particular time zone, as shown figure 6.

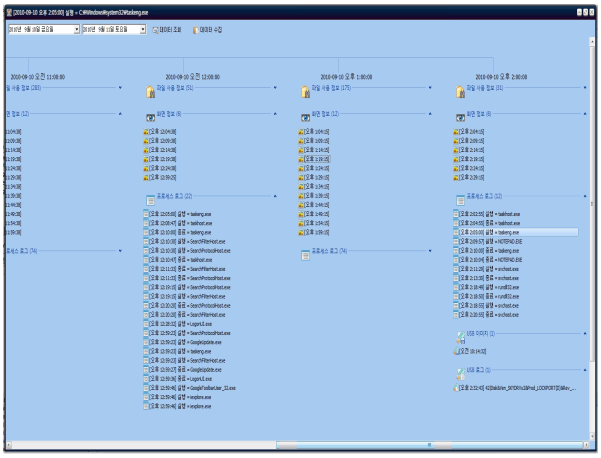


Fig. 6: Event display function by time zone

5. Conclusion

This paper proposed a policy-based operating tool with forensic readiness and applied it to the actual operating environment, which satisfies the safeguard control proposed by the privacy reference architecture.

A formal privacy manual or guidelines are insufficient to help cope with the privacy leak caused these days, and practical effects cannot be expected, even though measures are taken.

Existing privacy tools have strong aspects of post handle for the privacy infringement. Korea investigative agencies began adopting digital evidence as circumstantial in court, applying digital forensic investigation techniques. Such investigations have become a worldwide trend and their utilization will likely increase in the foreseeable future as acting the personal data protection act.

To overcome these shortcomings, the forensic readiness skim was adopted. The proposed method and tool produced a better result than previous methods in terms of forensic readiness. A massive forensic service will be further studied.

References

[1] M. Lee, "A Study on Consistency of Dealing with Residnet Registration Number PDPA(Personal Data Protection Act)," Journal of Korea Institute of Electronic Communication Sciences, vol, 2, no. 1, pp. 90-104, 2011.

[2] Y. Shin, "Standard Implementation for Privacy Framework and Privacy Reference Architecture for Protecting Personally Identifiable Information," International Journal of Fuzzy Logic and Intellignet Systems, vol, 11, no. 3, pp. 197-203, 2011.

[3] G. Lee, "A Study on Influence of Korea-EU FTA Ratification upon Legal Service and Forensic Investigation," Internet and Information Security, vol, 6, no. 5, pp. 683-688, 2011.

[4] ISO/IEC JTC1 SC27 “Privacy Framework,” SC27 N9226, May 2011.

[5] Y. Shin, S. Shin, “An Empirical Study on Massive Forensic Services,” Journal of Korea Institute of Electronic Communication Sciences, vol, 1, no. 2, pp. 83-100, 2010.

[6] Robert Rowlingson, “A Ten Step Process for Forensic Readiness,” International Journal of Digital Evidence, vol, 2, Issue 3, 2004.

[7] Tan, J. Forensic Readiness, July 2001, Electronic version retrieved 14th. 2003.

[8] ISO/IEC JTC1 SC27 “Privacy Reference Architecture,” SC27 N9228, May 2011.

[9] ISO/IEC JTC1 SC27 “Business plan for JTC1 SC27 Security Technique," SC27 N9463, Jun. 2010.

[10] ISO/IEC JTC1 SC27 WG5 “StudyPeriod Vocabulary,” SC27 N9401, May 2011.

[11] ISO/IEC JTC1 SC27 WG5 “Recommendation,” SC27 N9237, May 2011.

[12] ISO/IEC JTC1 SC27 “WG5 Resolution,” SC27 N9920, May 2011.

[13] Homeland Security Whitepaper, "Computer Network Security & Privacy Protection," 2011.

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

Yong-Nyuo Shin
Professor of the Hanyangcyber University
Research Area: telebiometric, mobile programming
E-mail : ynshin@hycu.ac.kr