Journal of the Korea Academia-Industrial cooperation Society
Vol. 14, No. 5 pp. 2415-2421, 2013

http://dx.doi.org/10.5762/KAIS.2013.14.5.2415

출력물에서의 개인 정보 제어 및 보안에 관한 연구

백종경[1*], 박재표[2]
[1]숭실대학교 대학원 컴퓨터학과, [2]숭실대학교 정보과학대학원

A Study on Personal Information Control and Security in Printed Matter

Jong-Kyung Baek[1*] and Jea-Pyo Park[2]
[1]**Department of Computing, Graduate School of Soongsil University**
[2]**Information Science Graduate School of Soongsil University**

요 약 개인 정보의 이용이 사회 전반적으로 보편화되면서 이에 대한 중요성이 점차 부각되고, 개인 정보 유출사례가 증가하고 있다. 여러 가지 개인 정보 유출방지 방안이 제안되었으나 프린트 출력 시 개인 정보 유출 및 제어에 있어 기존 방안들은 검출이 되지 않거나 외부로의 유출 시 개인정보가 노출이 되는 단점들이 있다.
본 논문에서는 API-Hook 방법을 사용하여 출력물에 대해 개인정보를 검출하여 제어하고, 출력 된 문서에 대해서는 마스크하여 개인정보의 노출을 보안하는 방법을 제시한다. 또한 실제로 구현하여 개인정보가 포함 된 문서에 대해 보안을 보장 여부를 확인 하였다. 보안을 위해 기밀성만을 중요시하기 보다는 가용성과의 조화가 필요하다.

Abstract Widespread personal data utilization has led personal data protection to its importance at core, and serious data spill has increased constantly as a result. Though various types of protection systems for data spill have been suggested, all these met failures in detection of personal data when printed out or preventing fatal data exposure without any protections when data spill happens.
I propose API-Hook method which detects and controls personal data within printouts, and prevents data leakage through masking on the printed-out data. Also, it is verified if security is guaranteed on the documents containing personal data when implementing. In order to obtain security, it is essential to put more weights on the balance with availability than confidentiality.

Key Words : Personal Information, Print Security, String Mask

1. 서론

정보화 사회가 발전함에 따라 정보의 경제적 가치가 높아지고, 개인 정보의 수집과 이용이 사회 전반적으로 보편화되고 있다. 최근 개인 정보의 유출 및 침해 사례가 꾸준히 증가하고, 기업에서는 기밀정보와 고객의 개인 정보가 외부로 유출되는 사례가 증가하고 있다. 또한, 명의도용, 전화사기 등으로 개인에게 정신적, 금전적 피해를 입히고 있다[1].

기업에서는 이를 방지하기 위해 DRM(Digital Right Management), DLP(Data Loss Prevention), ECM(Enterprise Contents Management)등과 같은 보안 솔루션을 통합 구축하고 있다. 출력물에 대한 보안도 지속적으로 이루어지고 있으며, 그 보안방법에는 출력 제어, 텍스트 로그, 원본 로그 등이 있다. 하지만 이러한 제어 방법들은 개인정보를 식별하여 제어하지 않는다. 텍스트 로그나 원본로그를 생성하는 방안은 출력 시 개인정보를 식별 하지 않고, 출력 후 로그를 통해 개인정보 유·무를 판단할 수 있다.

위의 보안방법들은 개인 정보를 포함한 출력물이 외부

*Corresponding Author : Jong-Kyung Baek(Soongsil Univ.)
Tel: +82-10-4433-9229 email: jkbaek@ssu.ac.kr
Received March 4, 2013 Revised April 2, 2013 Accepted May 9, 2013

에 노출될 경우 정보가 무방비하게 유출 될 가능성이 많다. 정부에서 고시하는 개인정보보호법은 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보를 말한다. 특히 개인 정보를 통해 발급 되는 이력서, 자격증명서, 졸업증명서 등과 회사의 기밀문서가 외부로의 유출 될 경우, 이를 보호하기 위한 대책 마련이 필요하다.

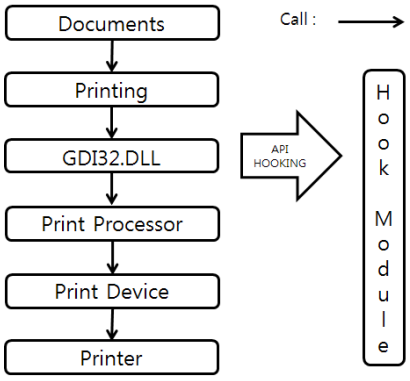
본 논문에서는 위와 같은 문제점들을 해결하기 위한 방법을 도출하여, 이를 구현하여 성능을 알아보고자 기존 방법과 비교를 한다.

2. 관련연구

2.1 출력물 제어

출력을 제어하기 위한 방법으로는 보편적으로 API (Application Programming Interface) Hooking을 이용한 방법, 인쇄처리기를 이용한 방법, 가상 프린트를 이용한 방법이 있다[3].

첫째, API Hooking을 이용한 방법은 Fig. 1과 같다.



[Fig. 1] API Hooking Method

프린트 시 출력을 위해 프로그램에서 사용되는 GDI32.DLL의 API를 Hooking 하여 프린트 작업을 감지하고 프린트 출력 허용 유·무에 따라 프린트 출력을 제어한다.

둘째, 인쇄처리기를 이용하여 제어하는 방법은 프린트 스풀러(Print Spooler)에 스풀 된 프린트 작업을 하드 디스크에서 프린터 장치로 보내는 과정에서 프린트 작업을 감지하여 프린트 출력을 제어한다.

셋째, 가상프린트를 이용한 프린트 제어 방법은 각 사용자의 컴퓨터에 프린트 제어 모듈을 내장 시켜 컴퓨터에 설치되어 있는 모든 어플리케이션에서 인쇄 시 프린트 작업을 감지하여 프린트 출력을 제어하는 것을 말한다.

위 세 가지 방법의 기술을 활용하여 로그, 원본로그, 워터마크를 기능을 제공할 수 있다. 하지만 프린트 출력물 내용에 포함되어 있는 개인정보에 대해 검출 및 보호에 대한 기능은 아직 연구가 되고 있지 않다.

2.1.1 출력물 텍스트 로그

2.1장의 출력물 제어 방법을 통해 출력물에 대한 정보를 추출 후 로그를 생성 한다. 이 로그를 중앙서버로 전송하여 관리 하도록 하여 프린트 출력 정보를 통해 전체적인 프린트 출력 상황을 관리 할 수 있다[3]. 로그의 정보는 출력 문서 명, 일시, 출력 PC등을 알 수 있지만 개인정보가 포함 된 문서인지는 알 수 없다.

2.1.2 출력물 원본 로그

프린트 제어 방법을 통해 사용자가 프린트 출력 작업 시에 출력물 대한 복사 이미지를 생성하여 중앙 관리 서버로 전송하여 관리 하도록 한다. 이 원본 이미지는 프린트 출력물에 대하여 식별 가능 이미지로 사용자가 어떤 프린트 출력 작업을 했는지 확인할 수 있다[3]. 하지만 그림 파일 형태로 작성 되어 개인정보 검출을 할 수 없다.

2.1.3 출력물 워터마크

프린트 출력 작업 시 사용자가 인지 할 수 있는 기업의 CI, 출력 정보, 텍스트 정보(보안 문서임에 대한 경고 문구)를 출력물에 삽입하여 무단 유출 이후에 출력물에 대한 출처 및 소유권을 증명할 수 있는 사후 관리 기능으로 사용할 수 있지만 개인 정보는 노출 된다[3].

3. 출력물에서의 개인 정보 제어 및 보안

본 논문에서는 프린트 출력 시 개인 정보가 포함 된 문서를 차단하고, 허용해야 할 경우 개인 정보 문자열을 마스크를 하여 보호하는 방안을 제시한다.

제안 모듈은 총 5가지의 정책 구조를 가진다.

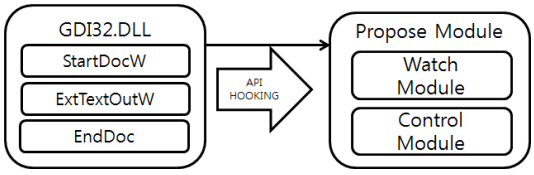
a. 출력 차단/ 허용
b. 개인정보 출력 차단/ 허용
c. 텍스트 로그
d. 원본로그
e. 개인정보 마스크

기존 연구 된 기술인 a., c., d.와 새로 제안된 기술 b., e.으로 구성된다.

b. 정책은 출력 시 개인 정보가 검출이 되면 차단하고, 검출이 되지 않는다면 출력을 허용하고, e. 정책은 허용인 경우 출력물의 개인정보가 외부로 유출이 되기 때문에 개인정보 문자열을 마스크한다.

3.1 제안모듈 구성

제안모듈은 프로세스의 출력 데이터를 감시하기 위한 감시모듈과 개인 정보 검출과 제어를 위한 제어모듈로 나뉜다. 제안모듈의 구성은 Fig. 2과 같다.



[Fig. 2] Composition of the Proposed Module

문서 프로세스가 출력 시 GDI32.DLL를 이용하여 DC(Device Context)에 그림 형태로 기록하게 되는데, 이 시점에서 문자열이 그림 형태로 바뀌게 된다. 그래서 이전에 문서 프로세스의 GDI32.DLL를 API 후킹 기술을 사용하여 제안모듈이 호출 될 수 있도록 한다.

제안 모듈은 감시모듈과 제어모듈로 나뉜다. 감시 모듈은 출력의 시작, 종료, 데이터 감시, 데이터 가공을 한다. 제어 모듈은 감시모듈에서 가공 된 데이터를 받아 개인 정보를 검출한다.

3.1 감시 모듈

감시 모듈에서는 개인정보검출에 필요한 문자열을 저장하고, 제어 모듈에게 전송하는 역할을 한다.

3.1.1 출력 시 문자열 추출 방안

문자열을 추출하기 위해서는 출력의 시작, 종료의 정보와 DC에 문자열이 기록되는 시점을 알아야 한다.

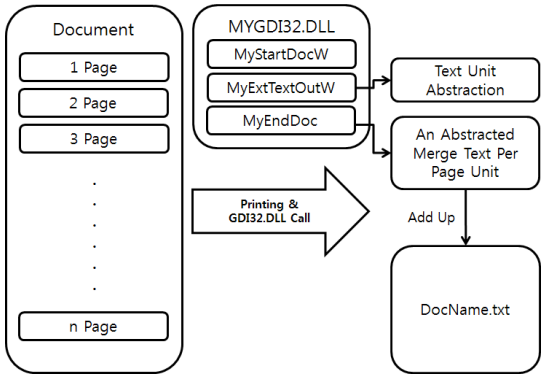
MyStartDoc() 함수는 출력의 시작, MyEndDoc 함수는 출력의 종료, MyExtTextOutW() 함수는 문자열을 추출하는 시점이다.

개인정보를 검출하기 위해서는 문자열이 필요하며. 추출 방안은 Fig. 3와 같다.

문서 프로세스에서 출력 요청이 들어오면 제안 모듈은 문자열을 검출하기 위해 감시모드로 전환한다.

DC에 문자열을 기록할 시 MyExtTextOutW() 함수가 호출이 되고, 3번째 인자로 DC의 Y좌표, 6번째 인자로 문자열 데이터가 들어오게 된다.

Y좌표를 비교하여 문자열의 줄 바꿈을 판단하고, 페이지 단위로 문자열을 조합하여 메모리에 저장한다. 출력이 종료되면, 페이지 단위로 가공 된 데이터를 파일로 저장하여 제어모듈 전달과 동시에 일반모드로 전환한다.



[Fig. 3] Character String Extraction System

3.2 제어 모듈

제어 모듈에서는 감시 모듈에서 데이터를 전송 받아 개인정보를 검출, 판단, 마스크 그리고 출력 제어를 한다.

3.2.1 개인 정보 검출 및 판단

개인 정보 검출 시 패턴과 체크디지트를 이용한다. 패턴은 개인 정보의 자리수와 검출 유형으로 구성한다. 비정규적인 문자열을 추출하게 되면 개인 정보 유형이 변하는 경우가 있기 때문이다.

Fig. 4는 검출 유형에 대한 정의이다.

```
[JUMIN]
count=7
jumin1=######-#######
...
jumin7=# # # # # # - # # # # # # #
[CARD]
...
[ENTERPRISE]
...
[DRIVER_LICENSE]
...
[CORPORATION]
...
[PHONE]
...
[ACCOUNT_NUMBER]
...
[PASSPORT]
count=2
passport1=########
...
```

[Fig. 4] Personal Information Detection Type

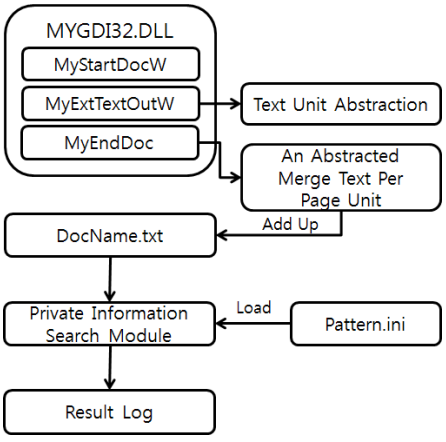
검출 유형은 개수와 개인정보 위치로 구성된다. 예를 들면, [JUMIN] 섹션은 7개의 주민등록번호 유형을 가지고 있고, 6자리 숫자, 7자리 숫자로 구성된다.

######-#######

820917-11111111

문서 마다 주민등록번호 형식이 다를 수 있기 때문에 다양한 유형을 제공한다. 마찬가지로 다른 개인 정보들도 각 형식에 맞는 유형으로 구성된다.

개인 정보를 검출하는 방안은 Fig. 5와 같다.



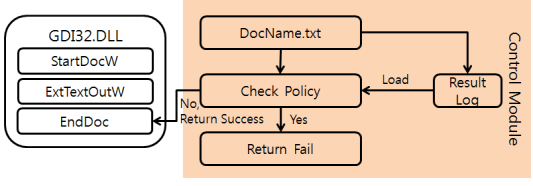
[Fig. 5] Personal Information Detection System

개인 정보 검출모듈은 개인정보 패턴을 로드하고 대기하다가 감시모듈에서 가공된 데이터가 저장 된 파일 경로를 전달 받는다.

패턴에 맞는 개인 정보가 검출이 되면 체크디지트를 체크하여 오류 탐지 및 과다 탐지를 최소화한다. 검출 결과 'ini'파일 형태로 저장한다.

3.2.2 개인 정보가 포함 된 출력물 제어

3.2.1의 과정을 거쳐 검출 결과 파일이 생성되면 Fig. 6와 같이 출력을 제어한다.



[Fig. 6] Print Control System of Personal Information

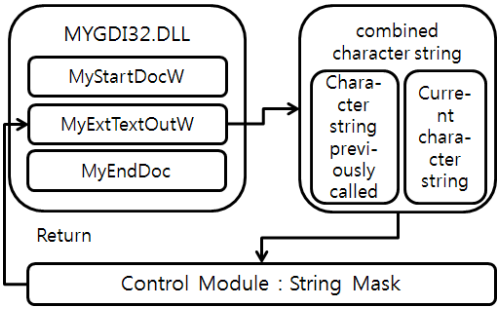
개인 정보검출 모듈에서 생성 된 분석 결과 파일을 로드하여 개인 정보의 종류와 개수를 분석하여 정책에 위반에 되는 경우 출력을 차단한다.

예를 들어 주민등록번호 10건 이상 포함 된 경우만 출력을 차단 할 경우 로그 파일의 주민등록번호의 개수가 10건 이상이면 제어하고 10건 미만일 경우 출력을 허용한다.

제어방안은 MyEndDoc() 함수에서 정상적으로 EndDoc() 함수를 호출하여 성공을 반환하지 않고, 문서 프로세스에게 출력 실패코드(음수)를 반환한다.

3.2.3 개인 정보 문자열 마스크

출력 후 개인 정보를 보안하기 위해 개인정보 문자열에 '*' 문자로 마스크하며 과정은 Fig. 7과 같다.



[Fig. 7] Character String Mask System of Personal Information

최초 MyExtTextOutW() 함수가 호출 될 경우 문자열에 대해 개인 정보가 있는지 확인한다. 개인 정보가 포함되어 있다면 마스크를 하고 이전 문자열 영역에 저장한다.

두 번째 이상 호출부터는 이전 문자열 값과 현재 문자열 값을 조합하여 개인 정보 유/무를 판단한다. 출력 호출 과정에서 개인 정보가 나뉘어 호출 된다면 개인 정보 식별이 되지 않기 때문이다. 마스크가 완료되면 문자열을 반환한다.

4. 구현결과 및 성능평가

4.1 시험환경 및 방법

시험 환경은 제안모듈은 서버와 클라이언트로 구분된다. 서버에서 정책 값 설정 및 클라이언트 개인 정보 현황을 제공한다.

클라이언트에서는 서버에서 개인 정보 제어 정책을 받아 반영한다. 정책은 개인 정보 검출 형식, 개수, 클라이언트 환경설정 값으로 이루어지며, 시험환경은 Table 1과 같다.

[Table 1] Test Environment

Section	O/S	Language	Note
Server	Windows 2008 R2	JSP	MS-SQL 2008
Client	Windows 2000 NT ~ Windows7 (x86/x64)	C/C++/ ASEM	
Print			SIndoricoh Aficio 2022
Print			Cannon MX308
Print			Samsung CLX-6220FXK

서버는 웹 기반 통신을 위해 Windows 2008 R2로 설정하고, 개인 정보 저장 현황 및 정책을 저장하기 위해 MS-SQL 2008을 사용한다.

클라이언트는 윈도우 운영체제 기반 32비트, 64비트 환경 모두에서 동작 할 수 있도록 구성하고, 여러 종류의 프린트 언어 및 드라이버를 시험하기 위해 각 제품별로 3대를 시험하였다.

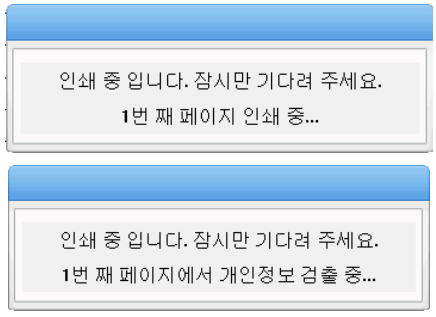
제안 모듈 정책 설정을 위해서는 서버에서 검출 조건을 설정하여 클라이언트로 내려준다. Fig. 8은 서버 웹 페이지에서 개인 정보 검출 조건 설정 화면이다.

[Fig. 8] Condition Setting of Personal Information Detection

각 개인 정보 별로 검출 조건 개수를 정할 수 있다. Fig. 7의 개수의 의미는 해당 개인 정보가 10건 이상 있는 경우 제어된다는 의미이다.

정책 수립이 완료 된 경우 클라이언트를 설치한다. 출력물 개인 정보 제어 및 마스크에 대한 시험 시나리오는 아래와 같다.

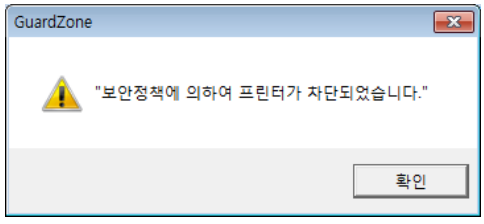
① 시험 대상인 프린트 드라이버를 설치한 후 테스트 PC와 연결한다.
② 서버에서 정책을 전송 받은 후 개인 정보가 담긴 문서를 열고 출력한다.
③ 출력 시 출력 중이라는 화면과 개인 정보 검출 중이라는 화면이 출력되며, Fig. 9와 같다.



[Fig. 9] Personal Information Detection Guide

출력 시 실시간으로 개인 정보를 검출하고 제어한다.

④ 개인 정보가 있는 경우 차단되었다는 알림창이 화면에 출력되며, Fig. 10과 같다.



[Fig. 10] Printer Block Guide

출력이 되지 않을 경우 사용자가 문제가 어디서 발생되는지 모르기 때문에 사용자 편의성을 위해 제공한다.

⑤ 출력물을 확인하여 개인 정보가 있는 문자열이 마스크 되어 있는지 확인한다. 클라이언트 모듈에서도 검출 된 정보가 확인이 가능하며, Fig. 11과 같다.

[Fig. 11] Person Information Detection Display

마스크 된 개인 정보 내용과 개수가 표시 되고, 출력 시 텍스트로 추출 된 문자열이 내용요약에 표시 된다.

4.2 구현결과 및 성능평가

개인 정보가 있는 문서를 출력 시 제어 및 개인 정보 문자열에 대해 마스크가 되어 있는지 확인하였다.

기존의 출력물 보안 방안들은 출력물에 추적 할 수 있도록 이력을 남기거나 워터마크를 삽입하였다. 또한 출력물을 외부로 유출을 차단하기 위해 제어, 일반 텍스트 로그, 원본 로그를 생성하여 관리하였다.

Table 2에서 기존 보안 방법과 제안 모델을 비교하였다.

[Table 2] Comparison between the Previous Security System and the Proposed Model

Section	Previous Model	Proposed Model
Print control range	Block/Accept print	Block/Accept print, Block/Accept Personal data
Security system after printing out	Insult trace words, Watermark	Characterized string for personal data, Mask
Personal Data Detection	None	Detected
Availability	Low	High, Possible to control while accepting print + Block only for personal data document
Confidentiality	Low Personal data exposed after printing out	High Personal data security after printing out

기존 모델은 출력 제어의 범위를 전체 출력 문서에 대해 차단, 허용밖에 되지 않아 일반 문서인 경우에도 출력을 못하거나, 관리자에게 승인을 받아 출력해야 하는 불편함이 있다. 하지만 제안모듈은 이를 보완하기 위해 개인 정보가 있는 경우에만 차단하고, 일반문서는 허용한다.

출력 후 출력물에 대해서도 기존에는 워터마크나 추적 관련 문구를 삽입하지만 개인 정보는 그대로 노출이 된다. 하지만 제안 모델은 개인 정보 문자열을 마스크를 하여 보안성을 높인다.

5. 결론

개인정보보호법 발효에 따라 개인정보가 중요해지고 있으며, 기업 내에서도 개인정보 보안 솔루션을 보급하고 있다. 하지만 출력물에 대해서는 아직 보호가 되지 않고 있다.

본 논문에서는 출력물 보호를 위해 제안 모델을 구현하여 개인 정보를 검출하여 제어하고, 외부 유출 시 출력물의 개인정보 마스크를 하였다. 또한 일반문서와 개인정보 문서를 식별하여 일반문서인 경우 허용하여 기존 모델에 보다 가용성을 높였다.

정보가 중요시 되고 가치가 높아지면서 기밀성이 강조되고 있지만, 필요이상의 높은 기밀성에 의해 업무에 불편함을 느낀다. 기밀성과 가용성을 조율하여 보안 솔루션을 구축해야 하겠다.

References

[1] Y. C. Baek, et, al., “The Internet and personal information protection act”, Korean Studies Information, 2012.

[2] Feng Yuan, “Windows Graphics Programming : Win32 GDI and DirectDraw“, Hewlett-Packard Professional Books, 2001.

[3] S. J. Kim, “The Status and Perspect of Personal Information Protection Technology”, The Korean Institute of Information Scientists and Engineers, vol. 27 no. 12 pp.10-18, 2009

[4] Y. S. Seo, et al., “A Study on Digital Fingerprinting Technology for the Copyright Protection of the Image Contents Printout”, The Korea Contents Association, vol. 4 no. 2 pp. 242-245, 2006.

[5] A. R. Kim, et al., “A Study in Notary System for Web Posting Digital Evidences”, Korea Institute of Information Security, Vol.21 no.3 pp. 155-163, 2011.

[6] Y. S. Seo, et al., “A Study on Digital Forensic Marking against Print-and-Capture”, Korea Information and Communications Society, vol. 33 no. 12 pp. 418-426, 2008.

박 재 표(Jae-Pyo Park) **[정회원]**

- 1998년 8월 : 숭실대학교 대학원 컴퓨터학과 (공학석사)
- 2004년 8월 : 숭실대학교 대학원 컴퓨터학과 (공학박사)
- 2008년 9월 ~ 2009년 8월 : 숭실대학교 정보미디어기술연구소 전임연구원
- 2010년 3월 ~ 현재 : 숭실대학교 정보과학대학원 교수

<관심분야>
컴퓨터통신, 정보보안, 디지털포렌식, 암호학

백 종 경(Jong-Kyung Baek) **[정회원]**

- 2010 2월 : 숭실대학교 정보과학대학원 정보보안학과 (공학석사)
- 2011년 3월 ~ 현재 : 숭실대학교 일반대학원 컴퓨터학과 (박사과정)

<관심분야>
정보보안, 데이터 유출 방지, 개인정보, 클라우드