

# 개인정보의 패턴 추출과 완전삭제방법에 관한 연구

서미숙\* · 박대우\*

\*호서대학교 벤처전문대학원

## A Study on Pattern Extraction and Complete Elimination Method of Private Information

Mi-Suk Seo\* · Dea-Woo Park

\*Hoseo Graduate School of Venture

E-mail : msseo@smsinfo.co.kr·prof\_pdw@naver.com

### 요 약

개인정보보호법의 시행에도 불구하고 아직 개인정보 유출로 인한 침해사고가 발생하고 있다. 개인정보 유출은 금융피해와 연관되어 진다. 개인정보의 탐색 및 검출과정에서 개인정보의 패턴을 분석하여 추출하는 연구와 불필요한 개인정보의 완전삭제에 관한 연구가 필요하다. 본 연구에서는 개인정보보호에 대한 패턴추출 연구와 완전삭제 방법을 연구하여 개인정보의 패턴추출 및 완전삭제 실험을 하였다.

### ABSTRACT

Enforcement of private information preservation law the infringement accident which is still caused by despite with private information outflow occurs. The private information outflow avoids finance and to become the fire tube. Analyzes the pattern of private information from search of private information and detection process and the research which it extracts and the research is necessary in about perfection elimination of the private information which is unnecessary. From the research which it sees it researched a pattern extraction research and a complete elimination method in about private information protection and it did the pattern extraction and a complete elimination experiment of private information.

### 키워드

개인정보삭제, 개인정보보호법, 고유식별번호, 삭제방법

### I. 서 론

2011년 개인정보보호법의 시행에도 불구하고 개인정보 유출로 인한 침해사고가 계속 발생하고 있다.

정보화 사회의 급속한 발전은 개인정보의 대량 집적·이용에 따른 보호조치가 충분하지 못하며 사회전반의 민간기업 뿐만 아니라 공공기관 등에서도 개인정보 수집·이용이 활발하다. 민간분야에서 수집·이용 되고 있는 개인정보는 비대면 상태에서의 이용자 식별 및 고객관리 등을 통한 서비스 제공에 필수적이며 이는 공공분야 대국민 행

정 서비스의 경우에도 마찬가지이다. 그러나 정보화 사회에서 정보는 개인 뿐 아니라 단체·기업, 정부 등의 조직에 있어서도 중요한 자원으로 인식된다.

그러나 개인정보 유출 및 오남용에 따른 사건 사고 또한 끊임없이 발생되고 있으며 사회 전반의 개인정보보호에 대한 우려가 갈수록 높아지고 있는 실정이다. 개인정보 침해는 그 자체로 개인의 자유로운 삶을 침해하는 것이며, 사생활 침해 뿐만 아니라 금융 피해와도 연관이 되고 있다. 개인정보 침해는 프라이버시 침해를 넘어서 프라이버시의 주체인 개인을 완전히 파괴할 수도 있

다.

개인정보의 탐색과 검출과정 개인정보의 패턴 추출에 관한 연구와 불필요한 개인 정보의 완전 삭제 관한 연구가 필요하다

고유식별정보를 처리하는 경우에는 그 고유식별정보가 분실·도난·유출·변조 또는 훼손되지 아니하도록 대통령령으로 정하는 바에 따라 암호화 등 안전성 확보에 필요한 조치를 하여야 한다<sup>3)</sup>.

## II. 관련연구

### 2.1 개인정보 적용범위의 확대

개인정보보호법은 2011년 3월 29일 개인정보의 수집·유출·오용·남용으로부터 사생활의 비밀 등을 보호함으로써 국민의 권리와 이익을 증진하고, 나아가 개인의 존엄과 가치를 구현하기 위하여 개인정보 처리에 관한 사항을 규정함을 목적으로 하여 시행 되었다. 개인정보의 적용범위가 확대되어 개인정보를 취급하는 ‘개인정보처리자’<sup>1)</sup>를 적용대상으로 규정하고 있으며 개인정보처리자란 업무를 목적으로 개인정보파일을 운영하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리하는 공공기관 법인, 단체 및 개인 등을 말한다. 기관이나 업종의 성격과 무관하게 동호회 종교단체 등 비영리 단체 개인 등 개인정보를 수집·이용하는 모든 법인 및 자연인이 개인정보보호법에 따른 규제 대상이 된다

### 2.2 개인정보 영향 평가 제도 도입

공공기관의 장은 대통령령으로 정하는 기준에 해당하는 개인정보파일의 운용으로 인하여 정보주체의 개인정보 침해가 우려되는 경우에는 그 위험요인의 분석과 개선 사항 도출을 위한 평가(이하 “영향평가”라 한다)를 하고 그 결과를 행정안전부장관에게 제출하여야 한다. 주요 평가 내용으로는 개인정보보호 관련 각종 법규 정책, 지침 등에 대한 준수 여부, 개인정보 수집·저장·이용·제공·파기 등 각 처리 단계별 보호 조치 사항, 정보주체의 권리 보호 및 개인정보보호를 위한 관리체계, 개인정보 처리시스템에 대한 기술적·관리적 보호조치 여부 등이 될 수 있다[1].

### 2.3 개인정보의 파기

개인정보처리자는 보유기간의 경과 개인정보의 처리 목적 달성 등 그 개인정보가 불필요하게 되었을 때에는 지체 없이 그 개인정보를 파기하여야 한다<sup>2)</sup>.

개인정보를 파기할 때에는 복구 또는 재생되지 아니하도록 조치하여야 한다<sup>3)</sup>.

### 2.4 고유식별정보의 처리 제한 및 암호화

개인을 고유하게 구별하기 위하여 부여된 식별정보로서 대통령령으로 정하는 정보(이하 “고유식별정보”라 한다)를 처리할 수 없다<sup>4)</sup>.

## III. 개인정보의 패턴 추출연구

### 3.1 개인정보의 패턴분석

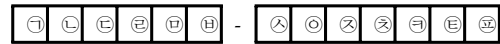
대통령령으로 정하는 정보 고유식별정보<sup>6)</sup>에 대한 패턴은 다음과 같다.

- 1) 주민등록번호
- 2) 여권번호
- 3) 운전면허번호
- 4) 외국인등록번호

고유식별정보의 패턴 검색 방법은 다음과 같다.

- 1) 주민등록 번호, 외국인등록번호 패턴 추출 방법

주민등록번호와 외국인등록번호는 주민등록번호 및 외국인등록번호의 생성규칙에 의거하여 검색



①, ② : 생년 뒷자리 2자리

③, ④ : 생월 01 ~ 12까지의 값만 유효

⑤, ⑥ : 생일 01 ~ 31까지의 값만 유효

⑦ : 성별 1900년대 남 : 1, 여 : 2

2000년대 남 : 3, 여 : 4

1800년대 남 : 9, 여 : 0

1900년대 (남자외국인) : 5

1900년대 (여자외국인) : 6

2000년대 (남자외국인) : 7

2000년대 (여자외국인) : 8

외국인의 경우 재외동포 포함

⑧, ⑨, ⑩, ⑪ : 출생지 동사무소 고유번호

⑫ : 출생신고 순서 1 이상 값만 유효

⑬ : 확인 숫자로 확인숫자 계산 방법은 다음과 같다.

$$[11 - \{(\textcircled{1} * 2 + \textcircled{2} * 3 + \textcircled{3} * 4 + \textcircled{4} * 5 + \textcircled{5} * 6 + \textcircled{6} * 7 + \textcircled{7} * 8 + \textcircled{8} * 9 + \textcircled{9} * 2 + \textcircled{10} * 3 + \textcircled{11} * 4 + \textcircled{12} * 5) \% 11\}] \% 10$$

- 2) 여권번호 패턴 추출 방법

여권번호는 구형여권과 신형 전자여권의 번호가 다음과 같은 형식으로 구성되어 있다. 공관부호는 다음과 같다.

1) 개인정보 보호법안(정부발의) 제2조 5호  
2) 개인정보 보호법안(정부발의) 제21조 1호  
3) 개인정보 보호법안(정부발의) 제21조 2호

4) 개인정보 보호법안(정부발의) 제24조 1호  
5) 개인정보 보호법안(정부발의) 제24조 3호  
6) 개인정보 보호법 시행령(대통령령 제23169호)

구형여권 : 공관부호(2자리) + 여권번호(7자리)  
 전자여권 : 여권종류(1자리) + 여권번호(8자리)

강남구청 KN, 강동구청 GD, 강북구청 GP, 강서구청 GS, 강원도 GW, 강원도 환동해 출장소 GW, 경기도 GK, 경기도 제2청사 GG, 경상남도 GN, 경상북도 GB, 고양시청 KY, 광주광역시 GJ, 광진구청 KJ, 구로구청 KR, 노원구청 NW, 대구광역시 DG, 대전광역시 TJ,

동대문구청 TM, 마포구청 MP, 부산광역시 BS, 부산시 사상구청 SJ, 부산시 해운대구청 HD, 서초구청 SC, 성남시청 SR, 성동구청 SM, 송파구청 SQ, 안양시청 AY, 영등포구청 YP, 용산구청 YS, 울산광역시 UL, 은평구청 EP, 인천광역시 IC, 인천 계양구청 GY, 전남도청 JN, 전라북도 JB, 제주 JJ, 종로구청 JR, 중구청 JG, 중랑구청 JU, 충청남도 CS, 충청북도 CB.

전자여권 종류는 다음과 같다.  
 관용 G, 외교관 D, 일반단수 S, 일반복수 M

3) 운전면허번호 패턴 추출 방법  
 운전면허번호는 다음과 같은 형태로 구성되어 있습니다.

지역 NN-NNNNN-NN  
 // 01-234567-89

여기에서 지역은 운전면허증을 발급하는 경찰청의 지역명이다.(16곳)  
 서울, 경기, 인천, 강원, 충북, 충남, 대전, 전북, 전남, 광주, 경북, 경남, 대구, 울산, 부산, 제주  
 0, 1번은 운전면허번호 발급연도이다.  
 2~7번은 일련번호로 알려져 있다.  
 8번은 체크디지트로 알려져 있으나 체크하는 연산방법에 대해서는 알려진 바가 없다  
 9번은 재발급 횟수를 의미한다

3.2 개인정보의 패턴 추출  
 현재 많이 사용되고 있는 20여개의 상용화된 개인정보 검색 솔루션이 있다. 대표적인 제품이 i-Safer, Privacy-i, PC 필터 등이 있다.



그림 1. 개인정보 패턴 선택  
 대부분의 개인정보 검색 솔루션은 먼저 그림 1과 같이 개인정보 패턴을 선택하고 경로를 선택 한 후 검색을 한다.

검색이 시작되면 검색 경로의 검색할 파일 개수를 파악한 뒤 검색이 시작되며 그림 2와 같은 검색결과가 나타난다.

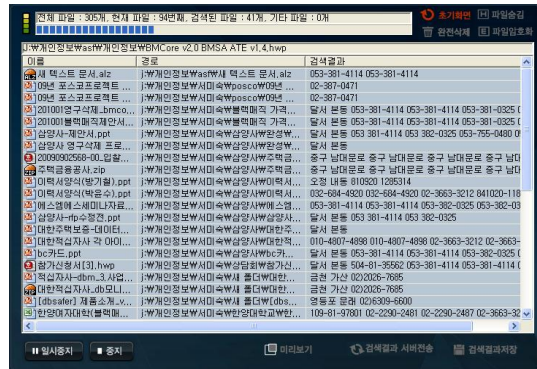


그림 2. 개인정보 검색 결과

#### IV. 개인정보의 완전삭제

##### 4.1 완전삭제 검증

대부분의 파일시스템에서는 파일을 효율적으로 관리하기 위해 그림 3과 같이 메타 영역과 데이터 영역으로 나누어 관리한다. 메타 영역에는 파일 메타 정보인 파일의 이름, 시간 정보, 크기, 삭제 플래그 등의 정보가 저장 되고 데이터 영역에는 실제 파일 데이터가 저장된다[2].



그림 3. 추상화한 파일시스템

표 1. 완전삭제 검증 기준

번호	검증 항목
1	메타 영역에서 파일의 정보를 복구할 수 없는가?
2	데이터 영역에서 파일데이터를 복구할 수 없는가?

운영체제에서 지원하는 파일 삭제 기능을 사용할 경우, 메타 영역의 삭제 플래그 값을 변경하여 파일이 삭제된 것처럼 인식하게 된다. 이 경우 메타 영역의 파일 메타 정보와 데이터 영역의 실제 데이터는 그대로 남아있기 때문에 쉽게 파일 복구가 가능하다. 표 1은 파일시스템의 특징을 기반으로 완전삭제 도구가 갖추어야 할 검증기준을 나타낸 것이다[3].

#### 4.2 완전삭제 방법

완전삭제는 파일 및 폴더를 복구할 수 없도록 완전히 삭제 파일폴더 삭제의 기능을 수행한다. 완전삭제시 사용하는 삭제방법은 다음과 같다.

- Kr Random(국정원 지침1) : 1차례 덮어쓰기
  - ① 난수값을 생성하여 1회 덮어쓴다.
- Kr Random - 00 - FF(국정원 지침2) : 3차례 덮어쓰기
  - ① 난수값을 생성하여 1회 덮어쓰고 ② 0x00(모든 비트가 0)값으로 1회, ③ 0xFF(모든 비트가 1)값으로 1회 덮어쓴다.
- Quick Erase : 1차례 덮어쓰기
  - ① 0xFF값으로 1회 덮어쓴다.
- DoD Short(미국방성 표준방식1) : 3차례 덮어쓰기
  - ① rand() 함수로 임의의 단일문자값을 생성하여 1회 덮어쓰고 ② 앞에서 생성한 단일문자값의 보수값으로 1회 덮어쓰고 ③ 난수값을 생성하여 1회 덮어쓴다.
- DoD 5220.22-M(미국방성 표준방식2) : 7차례 덮어쓰기
  - ①~③ DoD Short 방식으로 덮어쓰고 ④ rand() 함수로 임의의 단일문자값을 생성하여 1회 덮어쓰고 ⑤~⑦ DoD Short 방식으로 덮어쓴다.
- Gutmann Wipe(Peter Gutmann 삭제방식) : 35차례 덮어쓰기
 

다음의 순서대로 덮어쓴다

  - ①~④회 난수값
  - ⑤ 0x55 0x55 0x55(비트값이 01010101 01010101 01010101)
  - ⑥ 0xAA 0xAA 0xAA(비트값이 10101010 10101010 10101010)
  - ⑦ 0x92 0x49 0x94(비트값이 10010010 01001001 00100100)
  - ⑧ 0x49 0x24 0x92(비트값이 01001001 00100100 10010010)
  - ⑨ 0x24 0x92 0x49(비트값이 00100100 10010010 01001001)
  - ⑩ 0x00(모든 비트가 0)
  - ⑪ 0x11(비트값이 00010001)
  - ⑫ 0x22(비트값이 00100010)
  - ⑬ 0x33(비트값이 00110011)
  - ⑭ 0x44(비트값이 01000100)
  - ⑮ 0x55(비트값이 01010101)
  - ⑯ 0x66(비트값이 01100110)
  - ⑰ 0x77(비트값이 01110111)
  - ⑱ 0x88(비트값이 10001000)
  - ⑲ 0x99(비트값이 10011001)
  - ⑳ 0xAA(비트값이 10101010)
  - ㉑ 0xBB(비트값이 10111011)
  - ㉒ 0xCC(비트값이 11001100)
  - ㉓ 0xDD(비트값이 11011101)
  - ㉔ 0xEE(비트값이 11101110)
  - ㉕ 0xFF(모든 비트가 1)

- ㉖ 0x92 0x49 0x24(비트값이 10010010 01001001 00100100)
- ㉗ 0x49 0x24 0x92(비트값이 01001001 00100100 10010010)
- ㉘ 0x24 0x92 0x49(비트값이 00100100 10010010 01001001)
- ㉙ 0x6D 0xB6 0xDB(비트값이 01101101 10110110 11011011)
- ㉚ 0xB6 0xDB 0x6D(비트값이 10110110 11011011 01101101)
- ㉛ 0xDB 0x6D 0xB6(비트값이 11011011 01101101 10110110)
- ㉜ 난수값
- ㉝ 난수값
- ㉞ 난수값
- ㉟ 난수값

## V. 결 론

개인정보보호법의 시행은 개인정보보호 의무를 개인정보를 취급하는 개인정보처리자 전반으로 확대하고, 보다 강화된 개인정보보호 범규의무를 부과하고 있다. 그럼에도 불구하고 개인정보 유출로 인한 침해사고는 계속일어 나고 있는 실정이다. 개인정보보호법 제정으로 온라인 회원등록이나 회원가입 및 본인확인과정에 주민등록번호를 사용하지 않는 방안이 필요한 사항이며 개인을 식별할 수 있는 고유식별정보도 사용제한을 하고 있다.

개인정보 유출로 인한 피해를 최소화 하고 정보를 보호하기 위하여 PC내의 개인정보를 검색하여 관리하여야 하며 또한 개인정보의 처리 목적 달성 등 그 개인정보가 불필요하게 되었을 때에는 복구 또는 재생되지 아니하도록 본 논문에서 처럼 완전삭제를 하여 조치하여야 한다.

향후연구에서는 PC의 개인정보 유출탐지 및 차단에 대한 연구가 이루어져야 하겠다

## 참고문헌

- [1] 한국인터넷진흥원 “개인정보보호법 시행으로 인한 개인정보보호 규제 환경 변화 대응전략 3.
- [2] Brian Carrier, “File System Forensic Analysis,” Addison-Wesley, 2005.
- [3] 김연수, 방제완, 김진국, 이상진 “완전삭제 도구 사용 흔적에 관한 연구 정보처리학회논문지 C 제17-C권 제2호 2010. 4.