웹 크롤러를 이용한 개인정보보호의 기술적 관리 체계 설계와 해석

[1]박인표, [*2]전상준, [*3]김정호
[1](주)지란지교소프트, parkip.73@gmail.com
[*2 교신저자] 한국지질자원연구원, sjjeon@kigam.re.kr
[*3 교신저자] 한밭대학교, jhkim@hanbat.ac.kr

Design and Analysis of Technical Management System of Personal Information Security using Web Crawer

[1] In-pyo Park , [2] Sang-june Jeon, [3] Jeong-ho Kim
[1] *JiranSoft, parkip.73@gmail.com*
[*2,Corresponding Author] *Korea Institute Of Mineral Resources, sjjeon@kigam.re.kr*
[*3,Corresponding Author] *Hanbat National University, Computer Engineering, jhkim@hanbat.ac.kr*

요 약

개인정보가 포함되어있는 개인정보파일의 경우 개인용 PC 및 스마트 단말기, 개인 저장 장치 등 End-Point 영역에서의 개인정보보호에 대한 의식은 미흡한 실정이다. 본 연구는 웹 크롤러를 통해 생성된 개인정보파일을 안전하게 검색하기 위해 Diffie-Hellman 기법을 이용하여 사용자 키 레벨을 부여하였다. 개인정보파일에 대한 공격을 대비는 SEED와 ARIA를 하이브리드(hybrid)한 슬라이싱(slicing)을 이용하여 설계하였다. 웹 크롤링 방법에 수집된 개인정보파일에 대한 암호화 성능은 키 생성에 따른 암복호화 속도, 사용자 키 레벨에 따른 암복호화 공유를 비교 하였다. 이에 대한 시뮬레이션은 대외기관 전송 프로세스를 대상으로 전달된 개인정보파일에 수행하였다. 그 결과 기존 방법의 성능을 비교하여 기존보다 검출은 4.64배의 향상됨과 동시에 정보보호율은 18.3%가 개선됨을 확인할 수 있었다.

Abstract

In the case of personal information files containing personal information, there is insufficient awareness of personal information protection in end-point areas such as personal computers, smart terminals, and personal storage devices. In this study, we use Diffie-Hellman method to securely retrieve personal information files generated by web crawler. We designed SEED and ARIA using hybrid slicing to protect against attack on personal information file. The encryption performance of the personal information file collected by the Web crawling method is compared with the encryption decryption rate according to the key generation and the encryption decryption sharing according to the user key level. The simulation was performed on the personal information file delivered to the external agency transmission process. As a result, we compared the performance of existing methods and found that the detection rate is improved by 4.64 times and the information protection rate is improved by 18.3%.

***Key words** : Personal Informational Security, Web Crawerl, Diffe-Hellman, SEED, ARIA, Active Attack*

* Corresponding Author
Received: Dec. 10, 2018, Revised: Dec. 23, 2018, Accepted: Dec. 26, 2018

Ⅰ. 서 론

개인정보가 개인 외의 외부로 알려지게 되는 원인으로는 타인에 의한 유출과 다양한 이유로 인해 발생하는 노출이 있다. 외부로 노출된 개인정보는 도용, 사기 등 여러 가지 목적으로 악용될 우려가 크기 때문에 이를 방지하기 위한 기술적, 관리적인 예방 방법을 수립하여야 한다[1].

기존 연구된 “PC 개인정보 보호방법[2]” ,“웹크롤러 기반의 개인정보 침해 점검 시스템에 대한 방법론 연구[3]”는 사용자 키 생성으로 개인정보파일의 보호에 적용하였다. 그러나 개인정보파일이 사용자에 따라 다양화되어 사용자 키 레벨에 따른 추가보호가 요구되고 있다.[2,3]. 따라서 본 연구는 개인정보로 인한 피해를 사전에 방지 또는 최소화하기 위해 개인정보파일의 유·노출에 대한 사용자 키 레벨을 설정하고, 공격에 대한 슬라이싱 보호 방법을 다음과 같이 기술하였다.

첫째, 개인정보파일에 대한 기술적 안전성, 컴플라이언스(준거성), 시장성을 고려한 최적화된 인증으로 Diffie-Hellman 기법으로 사용자 키 레벨을 설계한다. 설계에 따른 성능해석은 SEED 와 ARIA 의 암호화를 하이브리드 하여 두 알고리즘의 내성으로 공격에 대한 슬라이싱 보호과정을 수행하였다.

둘째, 기존 구현 및 연구된 파일 내 개인정보파일 보호 방법을 비교하여 해석하였다[2,3] 개인정보파일 시스템은 포털 사이트의 검색 엔진의 도움을 받고, 웹 크롤링의 기법을 활용하여 시뮬레이션으로 개인정보파일에 대해 사용자 키 레벨에 따른 암복호화 속도와 공유에 대한 성능을 평가하였다.

따라서 본 연구는 개인정보파일을 검색한 후 안정성을 위해 Diffie-Hellman 기법을 이용하여 사용자 키 레벨을 부여한 후 SEED 와 ARIA 를 하이브리드한 슬라이싱을 이용하여 암복호화 속도, 공유 영역에 대해 추가적인 보호 방법을 기술하였다.

II. 개인정보파일의 정보보호 해석

2.1 개인정보 파일의 암복호화 과정

개인정보 관심의 변화는 법 제정 이 후 침해 관련 [그림 2.1]에 상담 건수의 증가 현황을 나타내었다. 정보주체의 적극적인 개인정보 자기결정권 행사로 인하여 컴플라이언스를 요구하는 수준 높은 개인정보 보호를 요구하고 있다[4].

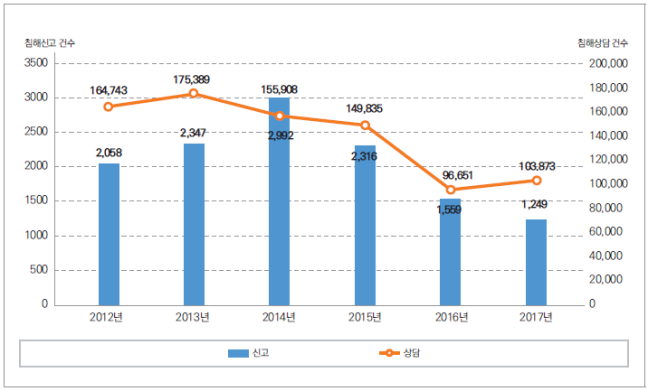

Figure 2.1 Personal Information Protections and Consulting
(개인정보 침해신고 및 상담건수, e-나라지표)

개인정보파일을 보호를 위해 SEED 함수의 초기 G 함수와 ARIA 의 확산계층을 이용하였다. SEED 의 초기 G 함수는 두 개의 8 비트 S-Box 를 이용하여 입력의 각 바이트를 비선형 변환 후

그 출력들을 적절한 permutation 을 수행하게 된다[5]. ARIA 의 확산 계층은 32, 64 비트 단위의 확산 함수를 사용하며, 이의 확산 함수는 함수 A: GF(28)16 → GF(28)16 는 입력을 (x0, x1,, x15)라 하고 출력을 (y0, y1,, y15)라 하면, 16 x 16 행렬의 곱으로 수행하게 된다[6].

2.2 사용자 키 레벨과 배분 설계

단말(클라이언트)와 관리 서버 간의 통신에서 의도적인 공격탐지의 기법으로 공유 비밀 키 분배 방식을 사용하는데 사용자 레벨 키 (Ks) 분배과정을 책임지는 제 3 자 KDC 를 이용하는 방식이다. KDC 는 사용자 정보뿐만 아니라 사용자와 KDC 만 알고 있는 마스터 비밀 키(master secret key)를 미리 공유하고 있다. 이 키를 사용하여 통신 쌍방 간에 필요한 단기간의 사용자 레벨 키(Ks)를 배포할 뿐만 아니라, 이들 간에 상호 인증할 수 있도록 한다. 클라이언트와 서버가 개인정보파일 보호를 위한 사용자 레벨 키를 KDC 로부터 내려 받는 과정은 다음과 같이 설계하였다.

1) 클라이언트 : KDC 에게 [클라이언트 ID, 서버 ID, nonce 값 N1]로 구성된 Request 문을 평문으로 전송한다.

2) KDC : 다음 내용들을 마스터키 EKa 로 암호화하여 클라이언트에게 응답한다.

- {사용자 레벨 키 Ks, Request, N1} 사용자 정보
- {EKb 사용자 레벨 키 Ks, 클라이언트 ID} 정보

3) 클라이언트 : 사용자 레벨 키 Ks 를 확보하고 서버에게 개인정보파일을 송신한다.

4) 서버 : 라우팅 선정과 네트워크 슬라이싱 순서정보의 내용을 복호화하여 클라이언트와 서버간의 공유 사용자 레벨 키 Ks 를 확보한다. 또한 서버는 라우팅 선정과 네트워크 슬라이싱 순서정보를 암화화 할 때 사용한 키는 오직 KDC 와 자신 만이 알고 있는 것이므로 라우팅 선정과 네트워크 슬라이싱 순서정보를 KCD 가 생성한 클라이언트임을 믿게 된다. 이어, 클라이언트에게 자신이 새로 생성한 임시값 N2 를 공유키 Ks 로 암호화하여 전송한다.

5) 클라이언트 : N2 에 1 을 더한 값으로 응답한다. 이렇게 함으로써 서버가 클라이언트를 인증할 수 있도록 한다.

개인정보의 유출을 방지하고 암호화 [그림 2.2]에 본 연구에서 선정한 Diffie-Hellman 기법에 의한 키 배정을 나타내었다.

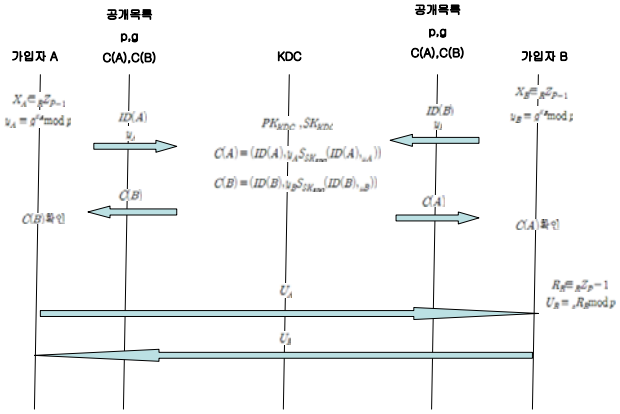

Figure 2.2 Key Arrangement of Diffe-Hellmann Mothod
(Diffie-Hellman 기법에 의한 키 배정)

본 연구에서는 개인정보파일에 대한 의도적인 공격탐지 Diffie-Hellman 방식의 키 사전 분배 방식을 개량하여 중간자 공격을 막을 수 있는 프로토콜이다[12]. 이 프로토콜의 환경은 Diffie-Hellman 키 사전 분배와 같이 소수 p 와 원시근 A 를 가정하며 키 동의에 사용되는 일회성 키들을 서명하여 교환하지 않고 사용자 레벨 키를 계산하는 방법으로, 기존 Diffie-Hellman 방식에서 사용자 A 와 사용자 B 가 항상 동일한 사용자 레벨 키를 가지게 되는 문제를 개선하였다.

III. 개인정보파일의 정보보호 설계

3.1 웹 크롤링의 활용

웹 크롤링은 문서를 수집하여 검색 대상의 색인으로 포함시켜 사용자가 검색 요청을 하면 내부적인 알고리즘에 의해 원하는 검색 결과를 찾아 사용자에게 보여준다. 웹 크롤링은 Seeds 라고 불리는 URL 리스트에서부터 시작하는데 페이지의 모든 하이퍼링크를 인식하여 URL 리스트를 갱신, 갱신된 URL 리스트는 재귀적으로 다시 방문한다. 웹 크롤링을 이용한 개인정보파일 분류를 [그림 3.1]에 나타내었다[3,7]

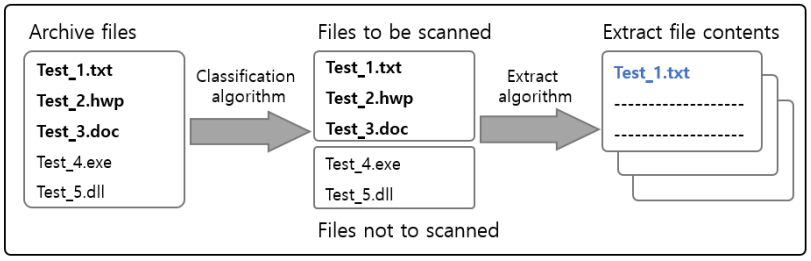

Figure 3.1 Personal Information File Classification (개인정보파일 분류)

3.2 개인정보파일의 키 사전 분배 설계

노출된 개인정보파일의 신속한 검출 및 조치를 위해 개인정보 보호 시스템 구축은 Diffie-Hellman 방식의 키 사전 분배 방식을 개량하여 중간자 공격을 막는 프로토콜은 다음의 단계로 수행한다.

(1) 관리서버에서 검색 스케줄에 대한 개인정보파일을 가지고 있으며, 스케줄에 의한 검색 실행 또는 사용자 필요에 의한 임의 실행으로 키 배정을 알리게 된다.
(2) 관리서버로부터 키 배정 정보를 명령받은 서버는 {사용자 레벨 키 Ks, Request, N1} 사용자 정보, {EKb 사용자 레벨 키 Ks, 클라이언트 ID} 정보 서버에서 KDC 역할을 지정한다.
(3) 서버로 부터 KDC 관련정보를 분배받은 서버는 사용자레벨 K_S를 확보하고 웹크롤링을 수행한다.
(4) 개인정보파일에서 사용자 레벨 K_S 는 검출된 내용이 이력서 등의 주민등록번호 정규화 표현에 적용되면 해당 웹 페이지의 URL 를 DB 서버에 저장한다.
(5)서버에 분배된 {사용자 레벨 키 Ks, Request, N1} 사용자 정보, {EKb 사용자 레벨 키 Ks, 클라이언트 ID} 정보 이력서의 주민등록번호 검색이 전부 종료되면 정보보호 완료 신호 서버로 보낸다.
(6) 서버에서 만들어진 KDC 정보는 개인정보파일 수만큼 분할하여 서버에 분할한다.
(7) 서버로부터 URL 리스트를 분배받은 서버는 할당된 SEED 와 ARIA 알고리즘 접속하여 해당 웹 페이지의 전체 페이지에 대하여 크롤링을 수행하여 암호화 과정을 포함한다.
(8) 검색 서버에서 검출한 개인정보를 검증서버로 전달하여 정오탐 검증 작업과 함께 복호화 과정을 준비한다.
(9) 서버에서 정탐으로 검증된 데이터를 {사용자 레벨 키 Ks, Request, N1} 사용자 정보, {EKb 사용자 레벨 키 Ks, 클라이언트 ID}를 DB 서버로 저장한다.
(10) DB 서버에 개인정보파일의 암복호화 데이터를 웹 서버에서 확인할 수 있도록 구현한다.

3.3 개인정보파일의 슬라이드 공격 대응

Slide 공격은 DC(Differential Cryptanalysis) 공격이나 LC(Linear Cryptanalysis) 공격 등은 암호화 라운드 수를 증가시킴으로써 공격 효율성을 떨어뜨릴 수 있으나 Slide 공격은 라운드 함수 자체의 특성이나 라운드 수 등과 무관하다는 특성을 갖고 있다[12]. 이 공격 방법은 동일한

라운드 함수를 갖고 있는 암호에서 키 스케쥴이 서로 다른 라운드에 대해 동일하거나 혹은 유사한 라운드 키를 출력할 경우 사용할 수 있다. 즉, $Ek = f_k' = f_k \cdot f_k \cdot \ldots \cdot f_k$여기서 f_k 는 비교적 취약한 키 치환 함수이다.

(1) SEED 의 Slide 공격

SEED 의 키 스케줄링에서 128 비트 입력키는 32 비트씩 4 개 조각으로 분할 후 (A, B, C, D), 1 라운드 키 $K_{1,0}$ 와 $K_{1,1}$ 를 $K_{1,0}$= G(A+C-KC_0), $K_{1,1}$ = G(B-D+KC_0)의 연산을 통하여 생성한다. 이때 연산에 사용되는 KC_0 는 1 라운드 비율에 따른 다음 식과 같이 생성한다[12].

$$KC_0 = int(\frac{\sqrt{5}-1}{2} \times 2^{32})$$

$$KC_i = KC_{i-1}^{\ll 1}, 1 \le i \le 15$$

DC 및 LC 복잡도의 bound 는 F 함수의 DC 및 LC 확률은 최소 2^{-6x4}=2^{-24} 을 갖게 된다. 그러나 실제로 S-Box 의 특성 및 Bit-permutation 의 성질상 이 확률을 갖는 특성이 생길 가능성은 희박하다., 확률이 존재하더라도 덧셈에 의한 Carry propagation 으로 인해 그런 characteristic 을 찾는 것은 불가능하게 된다.

(2) ARIA 의 Slide 공격

ARIA 는 키 스케쥴에서 마스터키를 Feistel 함수에서 π^{-1} 의 유리수 부분의 128 비트 상수 C_1, C_2, C_3 와 마스터 키를 조합하여 부분키를 생성하고, 이 부분키를 이용하여 두 개 혹은 세 개의 부분 키를 비트 회전 및 XOR 연산을 통하여 라운드 키를 생성하므로 각 라운드 키별 유사성이 제거된다. 또한 Involution 암호인 KHAZAD 에 적용된 Twisted Slide 공격 방법 역시 ARIA 에는 효과적이지 않는데, 이 공격 방법에는 2 개 라운드 키가 동일해야 하지만 ARIA 의 라운드 키는 동일한 키가 발생하지 않으므로 Twisted Slide 공격에도 안전성을 확보하게 된다[5,12]

IV. 시뮬레이션 수행과 성능 해석

4.1 시뮬레이션 환경

개인정보 파일의 보호 시스템을 구축하기 위해 필요한 요소는 다음 [표 4.1]과 같다. 또한 각 구축 서버별 기능에 대한 내용을 다음에 기술하였다.

Table 4.1 Simulation Environment (시뮬레이션 환경)

OS	CentOS 6.5
Language	JDK 1.7.1
Storage	MySQL Community Server 5.6.25
Tool	Bones 4.0, Sdk 2.1(trace)

(1) 관리서버 수행

관리서버의 주 기능은 개인정보파일의 보호기능 등을 검색을 시작할 수 있도록 명령을 내리는 서버이다. 또한 CentOS 의 기능인 Crontab 을 이용하여 주기적으로 검색을 수행할 수 있다.

(2) 제어서버 수행

관리서버로부터 검색을 수행하라는 명령이 전달되면 제어서버에서는 엔진 검색에 사용할 검색어 리스트와 웹 크롤링의 대상이 될 URL 리스트를 개인정보파일의 보호 등 각각의

검색서버의 개수에 맞게 분할하여 검색서버로 분배한다. 또한 간접검색 서버로부터 검색이 끝났다는 신호를 전달받으면 직접 검색 서버가 동작을 수행할 수 있도록 명령을 내리게 된다.

```
$crontab -e

# 매일 5,15,25일의 0시 0분에 search_start.sh를 실행한다.
# search_start.sh 실행으로 검색스케줄이 실행된다.

0 0 5,15,25 * 0 /root/bin/serach_start.sh > /dev/null 2>&1
```

Figure 4.1 Management Server crontab Schedule (관리서버 crontab 스케줄)

(3) 검증서버 수행

직접검색 서버로부터 검출된 개인정보 파일의 보호 등의 데이터를 전달받아 정 · 오탐 구분을 위한 동작을 수행한다. 이때, 정규화 표현식을 이용하여 실제 개인정보인지에 대해 판단한다. 본 연구에서는 개인정보파일에서 이력서의 노출에 대한 암호화 알고리즘을 적용하여 정탐으로 판단된 데이터를 검증하여 서버에 저장한다.

4.2 개인정보 전송 프로세서에 적용 사례

본 연구에서의 시뮬레이션은 대외기관 전송 프로세서를 대상으로 [그림 4.2]와 같이 수행한다[8,9].

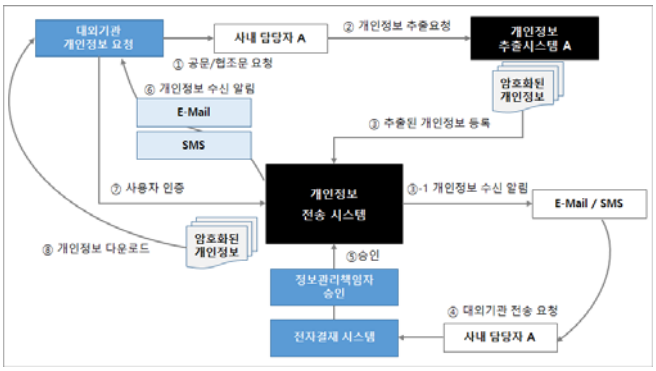

Figure 4.2 External Agency Transmission Process (대외기관 전송 프로세스)

- 1 단계 : 대외기관에서 공문이나 협조문 형태로 개인정보 제공을 요청한다.
- 2 단계 : 사내 담당자는 개인정보 추출시스템을 통해 추출을 요청한다.
- 3 단계 : 개인정보 추출시스템을 통해 추출된 개인정보는 개인정보 전송시스템으로 시스템 간 전송이 수행된다.
- 4 단계 : 추출 후 등록이 완료 된 개인정보에 대해 사내 담당자에게 이메일/SMS 등으로 자료가 등록되었음을 통보하여 준다.
- 5 단계 : 사내 담당자는 대외기관으로 반출될 개인정보에 대해 정보추출책임자에게 전자결재나 결재 등을 통해 인가 받는다
- 6 단계 : 정보관리책임자는 이 개인정보가 대외기관에 전송되어도 된다는 승인 즉 책임을 지겠다는 통보를 하면 전자결재시스템은 개인정보 전송시스템으로 반출에 대해 승인이 이루어졌으며, 이를 통보한다.
- 7 단계 : 개인정보 전송시스템은 대외기관 담당자에게 요청한 개인정보가 수신되었음을 통보하여 준다.
- 8 단계 : 대외기관 담당자는 개인정보 전송시스템에 사용자 인증을 통해 시스템에 접근한다.
- 9 단계 : 자신이 요청한 개인정보를 암호화된 상태로 전송 받아 업무에 활용한다.

이러한 과정을 사례로 하여 Diffie-Hellman 기법을 이용하여 사용자 키 레벨을 부여한 후 SEED 와 ARIA 를 하이브리드 한 Collision_Attack 를 이용하여 개인정보파일의 암복호화 공유 영역(M[i] and M[j])에 대해 해석하였다[14,17].

4.3 개인정보파일 보호의 시뮬레이션 결과

웹 크롤러 방식으로 대외기관 전송 프로세서를 대상으로 개인정보 전송 프로세서로 수집된 개인정보파일을 Diffie-Hellman 의 기반 포털 키워드와 URL 기반의 검색을 수행하였다. 검색후 SEED 와 ARIA 의 하이브리드로 한 슬라이딩 공격으로 Alternate_Collision_Attack 에 대한 기존 검색의 결과(M[k])와 설계한 시스템(M'[k])을 이용해 수집한 결과를 확인하였다. 먼저 구축한 시스템과 기존 시스템의 웹 사이트 당 수집할 수 있는 페이지의 수를 [표 4.2]에 비교하였다[15,17].

```
Alternate_Collision_Attack (M[k], M'[k])
{
        for(i=1 to k)
        {
        D[i] ← h(M[i])
        D'[i] ← h(M'[i])
        if (D[i] = D'[j] return (M[i], M'[j])
        }
  Return failure
}
```

Table 4.2 Suggested System Performance (제안된 시스템 성능 (개인정보파일 수))

(단위:페이지)

개인정보파일	구축 검색 스템(a)	기존 검색 스템(b)	성능비교(a/b)
doc	145	20	7.09
ZIP	90	48	1.86
Txt(한글)	270	35	7.65
Txt(영어)	73	17	4.33
tif	78	22	3.51

기존 검색 시스템에서는 크롤링 하지 않던 동적페이지의 검색을 수행함에 따라 개인정보파일 검색량이 증가했다. 기존의 시스템과 구축한 검색 시스템의 개인정보파일 검색량 차이는 페이지에 따라 다소간에 차이가 있지만 평균적으로 검색 시스템보다 약 4.89 배의 성능이 향상됨을 확인할 수 있다.또한, [표 4.3]에 실제 검출한 개인정보파일의 보호량을 비교 측정한 결과 약 18.3% 해당함을 알 수 있었다.

Table 4.3 Suggested System Performance-protection rate
(제안된 시스템 성능 비교 - 보호량(단위:%))

항 목	구축 검색 시스템	기존 검색 시스템
검 출 량	103	422
노 출 량	83	42
보 호 량	18.3	9.95

4.4. 개인정보파일 보호의 수행 평가

오피스 환경의 전자문서 관리에 보안으로 설계된 암호화를 워드 문서 파일(DOC), 전자파일을 압축하여 특수문자 비율이 월등히 높은 압축파일(ZIP), 전체 문자가 한글로 생성된 텍스트파일(TXT), 이메일 처럼 특수 문자가 포함되지 않은 영문 텍스트파일을 대상으로 수행하였다[9,11]. 시뮬레이션 결과 [표 4.2]에서 5 가지의 개인정보파일의 종류별 암복호화의 성능을 측정하여 [그림 4.3]에 나타내었다. 단, DOC 파일의 복호화 성능이 다른 파일 종류에 비해 더 낮아 보이나 테스트 수행 시점의 환경에 기인한 것으로 판단된다.

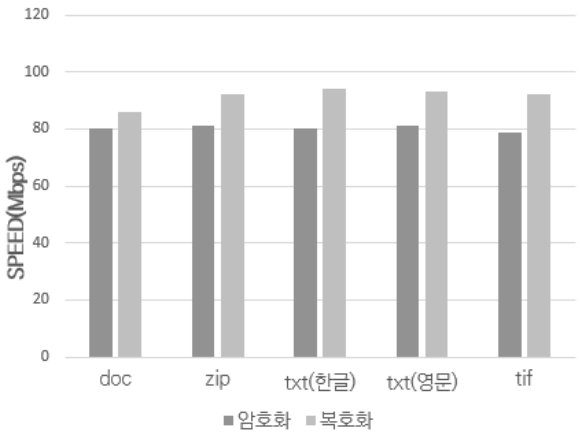

Figure 4.3 Encryption/Decryption Performance of Personal Information Files
(개인정보파일의 종류별 암복호화 성능)

[그림 4.4]는 개인정보파일의 보호에 대한 Diffe-Hellman 키 생성 부분과 라운딩 수를 증가하여 암복호화를 수행하여 각각의 성능을 나타내었다. 알고리즘별 및 각 부분별 측정값은 테스트 환경에서 각 부분의 실행 시간을 나노초 수준까지 취합하여 Mbps 단위로 변환하였다. 측정값 수치가 큰 수가 수행 시간이 더 짧은 것을 나타내며, 이는 수행속도가 빠른 것을 나타내었다.

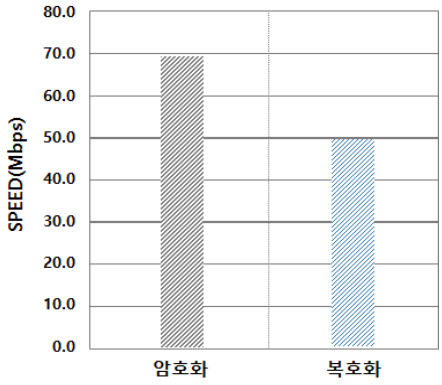

Figure 4.4 Encryption/Decryption Speed of Suggested Algorithm
(제안한 알고리즘에 따른 암복호화 속도)

V. 결론과 향후 연구

개인정보를 보호하기 위하여 관련 법률이 시행되었으며, 암호화, 완전삭제 등 기술적인 조치 또한 많은 발전을 이루었다. 또한, 기존에 구현되고 있는 기술적 조치는 법적으로 요구하는 안전성확보조치 기준의 개정을 따라가지 못하고 있는 실정이다[9]

본 연구는 웹 크롤러를 통해 생성된 개인정보파일를 안전하게 검색하기 위해 Diffie-Hellman 기법을 이용하여 사용자 키 레벨을 부여한 후 SEED 와 ARIA 를 하이브리드한 슬라이싱을 이용하여 개인정보파일의 안전성을 목표로 설계하였다. 웹 크롤링 방법에 대한 시뮬레이션으로 암호화 알고리즘 성능을 키 생성에 따른 암복호화 속도, 사용자 키 레벨에 따른 암복호화 공유를 비교 하였다. 시뮬레이션은 대외기관 전송 프로세스를 대상으로 기존의 시스템과 성능을 비교하여 기존보다 검출은 4.64 배의 향상됨과 동시에 정보 보호율은 18.3%가 개선됨을 확인할 수 있었다. 기존의 시스템[2,3]과는 다르게 사전 검증 관계를 통한 암복호화의 속도 향상과 동적콘텐츠 검색을 통한 검출 페이지의 양, 개인정보 검출량 등이 향상 되었다. 이는 실질적인 개인정보 취급자를 대상으로 한 개인정보보호법의 시행과 기술적 조치에 개선을 가져올 것이다.

향후 연구에서는 이미지 파일에 대한 검출율 향상 방안에 대한 연구와 검출 결과의 신뢰성을 좌우하는 오탐 처리 방법과 비정형데이터(상담기록, 성향, 지문 데이터 등)에 대한 검출 및 보호 방법에 대한 연구가 필요하다.

VI. 참고문헌

[1] H.Y.Kwon, “Significance and Classification of Privacy,” Privacy Security Enhancement Forum, Aug.2014.

[2] G.M.Shim, “A Study on the Exposure Type and Classification System of the Personal Information on the Internet Inspected by the Countermeasure System against the Web Invasion of Personal Information,” Yonsei University, 2009.

[3] S.T.Kim, Dissertation(M.S),“A Methodology for Privacy Incident Inspecting System based on Web Crawler”, 2016

[4] Ministry of the Interior and Safety, Korea Internet & Security Agency, 2013~2017 Checking the Status of Personal Information and the Case of Administrative Disposition, Apr.2018.

[5] J.I.Lee, Dissertation(M.S),“Study on Comparison of SEED and ARIA”, Sogang University, Feb, 2010.

[6] " ARIA Algorithm Specification ", National Security Research Institute, 2014.

[7] J.M.Yang, Patent,“Method and Device for Diagnosing Personal Information of Server”, G06F 17/00, 2010.

[8] Ministry of the Interior and Safety, “Standards and Instructions for Ensuring the Safety of Personal Information”, 2015.

[9] Ministry of Science, ICT and Futrue Planning, Korea Internet & Security Agency, “Enabling Password Activation Password Technology Implementation Guide”, 2013.

[10] M.S.Han, Dissertation(Doctor), “A Legal Study on The Protection & Use of Personal Information”, DDC 343.0858 22, 2015.

[11] E. Biham, O. Dunkelman, and N. Keller, "Related-Key Boomerang and Rectangle Attacks", EUROCRYPT 2005, LNCS 3494, pp.507-525, Springer-Verlag, 2005.

[12] A. Biryukov and D. Wagner, "Advanced Slide Attacks", EUROCRYPT 2013, LNCS 1807, pp. 589-606, 2013.

[13] A. Biryukov, C. D. Canniere, J. Lano, S. B. Ors and B. Preneel, "Security and Performance Analysis of AIRIA Ver.1.2", Katholieke Universiteit Leuven, Belgium, 2003.

[14] H. Yanami and T. Shimoyama, "Differential Cryptanalysis of a Reduced-Round SEED", SCN 2002, LNCS 2576, pp. 186-198, Springer-Verlag, 2003

[15] T. Jakobsen and L. Knudsen, "The Interpolation Attack on Block Cipher", LNCS 1267, FSE, pp.28-40, 1997

[16] A. Biryukov, C. D. Canniere, J. Lano, S. B. Ors and B. Preneel, "Security and Performance Analysis of AIRIA Ver.1.2", Katholieke Universiteit Leuven, Belgium, 2008.

[17] D. Wagner, "The Boomerang Attack", FSE'99, LNCS 1636, pp.156-170, Springer-Verlag, 2003.