홍 승 필 교수지도

석사학위 청구논문

개인정보 탐지 및 위험분석 모델

2011

성신여자대학교 대학원

컴퓨터학과

김 유 진

개인정보 탐지 및 위험분석 모델

홍 승 필 교수지도

이 논문을 석사학위논문으로 제출함

2010년 11월

성신여자대학교 대학원

컴퓨터학과

김 유 진

인 준 서

김유진의 석사학위 논문으로 인준함.

심사위원 홍 승 필 인

심사위원 서 동 수 인

심사위원 홍 의 석 인

성신여자대학교 대학원

논 문 개 요

정보화 사회에서 ‘개인정보’는 사회집단통제 및 막대한 부가가치의 원천인 재산권으로 발전하였고, 이에 따른 역기능과 부정적인 사회현상이 급증하고 있다.

최근 대량의 개인정보 유출 사건이 빈번히 발생함에 따라 개인정보 유출에 대한 우려의 목소리는 더욱 높아지고 있으며, 개인정보를 다루는 단체나 기업에서는 개인정보유출 탐지 및 차단 시스템을 도입하거나 대응 방안을 마련하고 있다.

하지만 정보의 실제 주체인 사용자의 환경에서는 이러한 대응 방안 체계가 구체적으로 마련되어 있지 않다. 또한 각 개인정보에 대한 중요도와 취약점에 따른 위험수준을 인지하기 어렵기 때문에 충분한 보호조치가 이루어지지 못하고 있다.

이에 본 논문에서는 개인정보보호의 중요성을 고려하여 다단계 개인정보 보호 정책을 기반으로 실제, 개인정보 오·남용 사용에 대한 체계적인 대응 방안을 제시하는 모델을 제안하였다.

또한 사용자 측면의 PC 환경 내 개인정보 위험도에 따른 개인정보를 안전하게 관리할 수 있는 메커니즘을 개발하였다.

마지막으로 이러한 개인정보 위험분석 및 대응 연구의 활용성을 고취하고자 실제 프로토타이핑을 통하여 가능성을 보여주었다.

목 차

표 목 차

그 림 목 차

I. 개인정보보호 개요

1.1 개인정보 정의

개인정보의 법적 정의

‘정보통신망이용촉진및정보보호등에관한법률’ 제2조 제1항 제6호에서는 '개인정보'를 생존하는 개인에 관한 정보로서 성명. 주민등록번호 등에 의하여 당해 개인을 알아볼 수 있는 부호. 문자. 음성. 음향 및 영상 등의 정보(당해 정보만으로는 특정 개인을 알아볼 수 없는 경우에는 다른 정보와 용이하게 결합하여 알아볼 수 있는 것을 포함한다)라고 정의하고 있다.

또한 공공기관의 개인정보보호에 관한 법률 제 2조 제2호는 생존하는 개인에 관한 정보로서 당해 정보에 포함되어 있는 성명·주민등록번호 및 화상 등의 사항에 의하여 당해 개인을 식별할 수 있는 정보(당해 정보만으로는 특정개인을 식별할 수 없더라도 다른 정보와 용이하게 결합하여 식별할 수 있는 것을 포함) 라고 정의하고 있다. 이 법에서는 공공기관이 보유하고 있는 개인정보 중 컴퓨터로 처리되는 개인정보 만을 대상으로 하기 때문에 민간 부문에서의 개인정보는 제외된다.

그 외에도 전자서명법 등의 개별법에서도 개인정보를 위와 동일하게 정의하고 있으며 국내 뿐 아니라 국외에서도 개인정보에 대해 법·제도적으로 정의하고 있다.

아래 [표 1]은 국외의 개인정보 정의를 정리한 것이다. [26][27]

[표 4] 국외 개인정보 정의

국가	내용
[OECD] 프라이버시보호가이드라인	식별된 또는 식별될 수 있는 개인에 관한 모든 정보
[EU] '95 개인정보보호 지침	식별된 또는 식별가능한 자연인(정보주체)에 관한 정보, 단 식별가능한 사람은 특히 신원증명번호 또는 육체적· 심리적·정신적, 경제적 문화적 또는 사회적 신원 중 하나 이상의 요인을 참고하여 직접적 또는 간접적으로 식별될 수 있는 사람을 말함
[캐나다] 프라이버시보호법	식별할 수 있는 개인에 대한 기록된 정보(인종, 국적, 민족, 종교, 나이 , 성별 등)
[독일] 개인정보보호법	신원이 확인 되었거나 확인가능한 자연인(정보주체)의 인적, 물적, 환경에 관한 일체의 정보
[영국] 개인정보보호법	당해정보로부터 식별할 수 있는 생존하는 개인에 관한 데이터
[오스트레일리아] 프라이버시법	진실하거나 아니거나, 물리적인 형태에 기록되어 있거나 아니건 간에 그의 신원이 명백하거나 합리적으로 판명될 수 있는 개인에 관한 정보 또는 의견
[홍콩] 개인정보법	생존하는 개인에게 직접적 또는 간접적으로 관련되어 있고, 직접 또는 간접적으로 개인의 신원을 확인하기 위하여 이용할 수 있으며, 해당 정보에 대한 접근이나 처리가 이루어질 수 있는 형태의 정보

국내외 개인정보 관련 법·제도에 나타난 개인정보의 개념을 종합해 볼 때 '개인정보'는 '개인을 식별할 수 있는 정보' 라는 사실을 알 수 있다.

개인정보의 의미적 정의

개인정보의 개념은 학자 또는 관련 규범에 따라 다양하게 정의되고 있으며 그 범위 또한 상이하게 해석될 수 있으나 대체적으로, 개인의 건강상태, 신체적 특징, 사상이나 신념과 같은 정신세계, 학력 · 경력 · 재산상태, 사회적 · 경제적 지위 등 개인에 관한 사실 ·판단 · 평가를 나타내는 모든 정보를 개인적으로 파악하는 정도를 규범 질 수 있다.

또한 개인정보와 관련해서 반드시 고려되어야 할 개념으로 '프라이버시'를 둘 수 있는데 개인정보와 '프라이버시'는 흔히 동일한 의미로 사용되고 있다. '프라이버시'는 1890년 미국의 Warren과 Brandeis에 의해 "홀로 있을 수 있는 권리(the right to be let alone)"으로 정의한 이후 프라이버시 자체가 그 사회의 정치 ·경제 및 법 · 문화와 같은 각국의 사회문화적 환경에 따라서 그 보호정보를 달리하기 때문에 프라이버시에 관한 정의는 다양한 형태로 존재 되어 진다. 우리나라 헌법 에서는 "모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다." 고 규정하여 프라이버시(사생활의 비밀과 자유)의 보호에 불가침을 선언하고 있다. 이에 따라 모든 국민은 프라이버시를 침해당하지 않을 소극적 권리는 물론 자신의 정보를 통제할 수 있는 권리도 보장받음을 명시 하고 있다. [26]

개인정보의 내용적 분류

현행 개인정보는 [표 2]에서와 같이 주로 개인의 신상 및 개인의 관계성에 기반을 둔 정보가 대부분이며 향후 유비쿼터스 환경에서는 개인정보의 유형이나 그 대상이 정보주체자체 뿐만 아니라, 물품정보 및 위치정보 등을 통한 개인의 라이프스타일까지 개인별 정보 파일로 정리되고 구체화 될 것으로 예상된다. [27]

[표 5] 개인정보의 유형

구분	개인정보유형
일반정보	이름, 주민등록번호, 운전면허번호, 주소, 전화번호, 생년월일, 출생지, 본적지, 성별, 국적
가족정보	가족구성원들의 이름, 출생지, 생년월일, 주민등록번호, 직업, 전화번호
교육 및 훈련정보	학교출석사항, 최종학력, 학교성적, 기술 자격증 및 전문 면허증, 이수한 훈련 프로그램, 동아리 활동, 상벌사항
병역정보	군번 및 계급, 제대유형, 주특기, 근무부대
부동산정보	소유주택, 토지, 자동차, 기타소유차량, 상점 및 건물 등
소득정보	현재 봉급액, 봉급경력, 보너스 및 수수료, 기타소득의 원천, 이자소득, 사업소득
기타수익 정보	보험(건강, 생명 등) 가입현황, 회사의 판공비, 투자프로그램, 퇴직프로그램, 휴가, 병가
신용정보	대부잔액 및 지불상황, 저당, 신용카드, 지불연기 및 미납의 수, 임금압류 통보에 대한 기록
고용정보	현재의 고용주, 회사주소, 상급자의 이름, 직무수행평가기록, 훈련기록, 출석기록, 상벌기록, 성격 테스트결과, 직무태도
법적정보	전과기록, 자동차교통위반기록, 파산 및 담보기록, 구속기록, 이혼기록, 납세기록
의료정보	가족병력기록, 과거의 의료기록, 정신질환기록, 신체장애, 혈액형, IQ, 약물테스트 등 각종 신체테스트 정보
조직정보	노조가입, 종교단체가입, 정당가입, 클럽회원
통신정보	전자우편(e-mail), 전화통화내용, 로그파일(log file), 쿠키(cookies)
위치정보	GPS나 휴대폰에의한 개인의 위치정보
신체정보	지문, 홍채, DNA, 신장, 가슴둘레등
습관 및 취미정보	흡연, 음주량, 선호하는 스포츠 및 오락, 여가활동, 비디오 대여기록, 도박성향

사용자 환경 내 개인정보보호 필요성

PC보유율이 높아지고 인터넷의 발달과 함께 점차 다양한 서비스가 웹을 통해 제공되고 있으며 중요한 자료를 개인 컴퓨터에 저장하는 일이 보편화 되고 있다. 이때 사용자의 개인정보가 오·남용 되면서, 개인정보에 대한 적절한 대안과 보호 방안의 필요성이 나날이 대두되고 있다. 최근에는 기업에서 유출되는 대량의 개인정보 유출 외에도 해킹 및 취약한 인터넷 사용자 PC에 설치된 봇을 이용하여 주민등록번호 및, 계정정보, 신용카드번호 등과 같은 개인정보를 유출시키는 사례가 늘어나고 있다.

일반적으로 개인 PC를 사용하게 되면 인터넷을 통한 서비스 이용이나 사업처리 과정에서 사용자의 인지 없이 개인정보의 기록이 PC에 저장 될 수 있다. 또한 개인적인 프로필을 저장한 문서가 시스템 상에 저장될 수 있으며, 이러한 개인정보들이 기록된 파일을 타인 혹은 불법적인 접근에 의해 획득하거나 열람되어질 경우 사생활 침해 및 금전적인 피해를 유발 할 수 있다.

최첨단 정보화 사회로 변화하고 있는 시점에서 시스템 내에 개인정보를 전혀 저장하지 않고 비즈니스를 수행한다는 것은 불가능한 상황으로서, 개인정보를 안전하게 유지해야하는 책임감을 가져야하며, 그러기 위해서는 사용자 스스로 개인정보를 보호해야하는 것이 가장 우선적인 의무라는 점을 인지하고, 올바를 판단에 따라 자신의 개인정보를 통제·보호하는 것이 필요하다.

따라서 본 논문에서는 개인정보 유출 공격 및 대응 기술 동향 조사를 통해 위협을 알아보고, 사용자에게 개인정보 노출 위험을 인지시키고 개인정보 유출 시 위험수준을 고지할 수 있는 개인정보 탐지 및 위험 분석 모델 제안한다.

1.2 개인정보 침해유형

개인정보가 수집되어 저장되고, 이를 관리하고 필요에 맞게 이용한 후 정보를 파기하기까지의 정보 처리 시스템을 개인정보 생명주기라 한다. 이를 개인정보의 수집, 저장 및 관리, 이용 및 제공, 파기의 4 단계로 나누고, 각 단계별 개인정보의 침해 유형을 살펴보면 아래의 [표 3]과 같다.

[표 3] 개인 정보 침해 유형

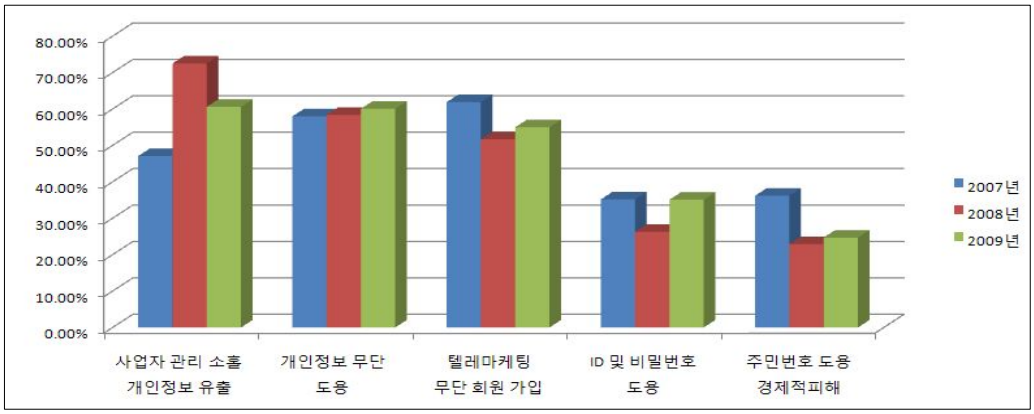
생명주기	침해유형
수집	이용자 동의 없는 개인정보 수집
	개인정보 수집시 고지 또는 명시의무 불이행
	과도한 개인정보 수집
	법정 대리인의 동의없는 아동의 정보 수집
저장 및 관리	개인정보 취급자에 의한 훼손· 침해 누설
	개인정보관리책임자 미지정
	개인정보보호 기술적· 관리적 조치 미비
	동의철회· 열람 또는 정정 요구 등 불응
	동의철회· 열람· 정정을 수집방법보다 쉽게 해야 할 조치 미 이행
이용 및 제공	개인정보처리 위탁 시 고지의무 불이행
	영업의 양수 등의 통지의무 불이행
	고지· 명시한 범위를 초과한 목적 외 이용 또는 제3자 제공
	주민번호 등 타인 정보의 훼손· 침해· 도용
파기	수집 또는 제공받는 남용목적 달성 후 개인정보 미파기
기타	기타 (정보통신망법 규정 외의 침해유형)

한국인터넷진흥원 개인정보침해신고센터에서 접수된 개인정보 침해건수와 개인정보 침해신고 상담건수를 살펴보면 2010년 9월까지 개인정보분쟁조정위원회에 접수된 개인정보 침해신고 상담건수는 총 35,842건으로 조사되었다. 이는 2000년에 접수된 2,035건과 비교하여 20배가량 증가한 수치로서, 개인정보침해 관련 민원이 꾸준히 증가하였음을 알 수 있다. 이 중 개인정보분쟁조정위원회에 가장 많이 접수된 침해유형은 '주민번호 등 타인 정보의 훼손・침해・도용(6,652건)' 이다. [27]

[표 7] 개인정보 침해발생건수(2000-2010)

침 해 유 형	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
이용자의 동의없는 개인정보 수집	138	388	221	260	564	1,140	2,565	1,166	1,129	1,075	947
과도한 개인정보 수집	6	13	21	38	43	33	61	51	87	115	101
개인정보 취급자에 의한 훼손·침해 또는 누설	28	122	139	172	235	186	206	123	125	158	106
기술적·관리적 조치 미비로 인한 개인정보 누출 등	-	11	37	82	212	390	632	522	1,321	819	1,061
주민등록번호 등 타인 정보의 훼손·침해·도용	956	5,785	8,298	8,058	9,163	9,810	10,835	9,086	10,148	6,303	6,652
개인정보 침해발생건수(소계)	1,128	6,319	8,716	8,610	10,217	11,559	14,299	10,948	12,810	8,470	8,867
합 계	2,035	11,164	17,956	17,777	1,7569	18,206	23,333	25,965	39,811	35,167	35,842

2009년 한국인터넷진흥원이 개인 인터넷 이용자를 대상으로 실시한 설문조사 결과에 의하면, 전체 응답자 중 23.4%가 개인정보 침해로 인한 피해 경험이 있는 것으로 나타났다. 대표적인 침해 유형으로는 사업자 관리 소홀로 개인정보 유출이 60.6%로 가장 많았다. 그 다음으로는 그 다음으로는 동의 없이 개인정보를 본래 목적 이외의 용도로 이용(60.1%), 개인정보 무단 수집 후 텔레마케팅에 이용하거나 무단 회원가입(55.0%), ID 및 비밀번호 도용(35.1%), 주민등록번호 도용(24.7%) 순으로 나타났다. [그림 1]은 이용자의 동의 없이 개인정보를 본래 목적 이외의 용도로 사용한 현황을 년도별로 보여주는 그림이다.



[그림 1] 년도별 본래 목적 이외의 개인정보 사용 현황

1.3 개인정보 유출 유형

개인정보가 유출되는 유형은 크게 내부자에 의한 유출과 외부자에 의한 유출로 나눌 수 있다.

내부자에 의한 개인정보 유출

내부자에 의한 개인정보 유출은 기업이나 단체가 소유하고 있는 개인 정보가 내부자에 의해 외부로 유출되는 경우로 인터넷 환경을 사용한 네트워크를 통한 유출과 내부자가 직접 외부로 가지고 나가는 경우를 살펴 볼 수 있다. 내부자에 의한 개인정보 유출은 내부자가 악의를 품고 행한 행위일 수 있지만 단순히 업무의 연장 차원에서 내부로부터 개인정보를 가지고 나온 것일 수도 있다. 후자의 경우 또한 정보에 대한 중요도를 인식하지 못하는 경우이기 때문에 큰 위험이 될 수 있다. 따라서 내부자에 의한 개인정보 유출은 내부자의 행동을 통제할 수 있어야 할뿐만 아니라 네트워크에 대한 적절한 통제가 동시에 이루어져야 한다. [24]

외부자에 의한 개인정보유출

내부자가 아닌 외부자에 의해서 개인정보가 유출되는 경우는 개인정보 유출을 통해 금전적인 이익을 보려는 시도들이 주를 이루고 있다. 이런한 개인정보가 유출되는 방법은 크게 세가지로 나눌수 있다. 첫째로 웹페이지의 관리 소홀이나 정책의 미비로 인해 정보가 노출되는 경우, 둘째는 특정정보를 획득하기 위해 목표로 하는 시스템을 공격하는 해킹을 통한 정보노출, 마지막으로 불특정 다수를 대상으로 웜·바이러스와 같은 악성코드의 유포를 통한 유출이 있다.

아래의 [표 5]는 개인정보의 유출유형을 내부자에 의한 유출과 외부자에 의한 유출로 나누어 정리한 표이다. [24]

[표 8] 개인정보 유출유형

분류	개인정보유출유형	비고
내부자에 의한 유출	P2P, FTP, 메신저, 메일 등을 통한 개인정보유출	P2P 파일공유 및, FTP 우회접속, 메일의 첨부파일 등을 통한 개인정보 유출
	CD,USB 등의 저장 매체를 이용한 개인정보 유출	저장용량에 크게 제약을 받지않고 손쉽에 이동할 수 있다는 장점이 개인정보 이동 역효과를 가져옴
	PC 또는 노트북 등의 기기 반출을 통한 개인정보 유출	업무의 목적으로 허가받은 디지털 개체(PC, 노트북)를 이용한 개인정보 유출
	개인정보가 출력된 문서 유출	문서에 의한 정보누출은 사실상 기술적으로 제한을 가하기 어려움
외부자에 의한 유출	홈페이지 개인정보 노출	일반 인터넷 사용자가 해킹 등 특별한 방법을 이용하지 않고 정상적으로 인터넷을 이용하면서 타인의 개인정보를 취득 할 수 있음
	해킹을 통한 개인정보 유출	패스워드 공격, 원격제어 프로그램, 운영체제 취약점 등을 이용한 비인가된 방식의 개인정보 취득
	악성코드에 의한 개인정보 유출	불특정 다수의 시스템을 대상으로 웜· 바이러스·스파이웨어 등을 통한 정보 유출

II. 개인정보보호 관련연구

2.1 개인정보보호 법·제도 연구

2.1.1 국제 기구별 개인정보보호 가이드 라인

개인정보보호에 대한 국제 기준을 논하는 경우 OECD Privacy Protection이 가장 빈번하게 언급된다. OECD의 정보보호 규제는 정보시스템, 네트워크, 개인정보 및 기타 영역 4개 분야로 나뉘어 관련 가이드라인이 작성되어 있다. 2004년부터 2006년까지 OECD의 주요 논의는 정보보호가이드라인의 이행, 개인정보 수집 시 고지사항의 국제적 표준화, 개인정보보호 관련 법 이행을 위한 국제협력, 정보보호문화의 확산 추진, 전자인증, 국제여행 보안 강화 관련 정보보호 및 프라이버시에 집중되어 있었다.

이 중에서 개인정보의 수집 및 관리에 대해 국제사회의 합의를 반영한 것이 1980년 개인정보보호 가이드라인으로 생성되었으며, 이는 법적인 구속력을 갖고 있지는 않지만 개인정보보호의 일반 원칙으로 인정되어, 개별국가의 개인정보보호에 관한 각종 법제를 정비하는데 기본적인 지침으로 반영되고 있고 국제적 차원의 개인정보보호 정책 제정과 프라이버시권의 보호에 관한 논의에서 정치적·도덕적 기준으로 활용되고 있는 상황이다.

2007년 이후 주로 논의된 내용은 인터넷 경제의 미래라는 주제와 연관된 프라이버시 보호에 대한 내용이다. 미래 사회는 경제성과 사회복지가 향상되고, 융합의 혜택을 받으며, 글로벌 인터넷 경제 시대가 도래하기 때문에 개방적인 유비쿼터스 시대에 신뢰를 바탕으로 한 프라이버시의 보호가 중요할 것으로 예상되고 이에 따라 전자정부 서비스에 따른 프라이버시와 개인정보보호 문제, RFID의 사용에 따른 프라이버시 보호, 유전자 프라이버시와 관련된 프라이버시 보안 문제, 국가 간 개인정보 유통(TBDF: Transborder Data Flow)에 따른 보안 문제, 전자인증의 국가 간 상호 인증 문제, Digital

Identity의 보안 관리 등을 기본적인 이슈로 주된 논의가 진행 중에 있으며 신뢰 구축 및 유지가 큰 타이틀로 설정되어 있다.

또 다른 국제기구로 개인정보의 표준, 도구, 기술을 연구 및 평가하고 개인정보보호 프레임워크를 정의하기 위해 구성된 전 세계 기업과 기술 공급자들의 연합인 ISTPA(The International Security, Trust, and Privacy Alliance)가 있으며, ISTPA는 계층 분석적 접근을 통해 개인정보보호 요구사항들을 취급하며 개인정보보호 원칙에 없는 정보의 주체에 의해 세워진 개인정보 사용 제약사항을 준수할 수 있는 7가지 개인정보보호 서비스와 3가지의 갖추어야 할 능력을 개인정보보호 프레임워크로서 정의하고 있다.

또한 IPC(Information and Privacy Commissioner)에서는 개인정보보호에 대한 10원칙을 제정하였고, APEC 회원국 간의 전자 상거래 촉진을 위해 제반 활동을 하는 고위관료회의 산하 특별그룹인 ECSG(Electronic Commerce Steering Group)에서 APF(APEC Privacy Framework: Privacy/Data Protection) 개인정보보호 9원칙을 개발하여 채택한 바가 있다.

상기에서 언급한 국제 기구별 개인정보보호 가이드라인 내용을 비교했을 때 표현이나 수준에서 조금 상이한 부분은 있지만 개인정보 수집을 제한하고 해당 개인정보의 무결성을 검증하고 수집목적을 정확하게 명시하고 활용을 제한하고 저장 시 안전성을 확보하는 등 필수요건을 비슷한 정도로 요구하고 있음을 알 수 있다. [17]

2.1.2 국외 개인정보보호 법·규제 내용

①EU (Europe Union)

EU는 개인정보보호를 위해 필요에 따라 부분별 입법화하기 보다는 적극적인 모습으로 전체를 포괄할 수 있는 일반적인 원리들을 규정하고 있다. 유럽연합 산하 단체인 각료회의이사회(European Council)는 유럽협약의 각료회의의 내용을 기초로 개인의 권리와 자유 및 프라이버시권의 보호 원칙을 포함하는 '개인정보처리에 대한 개인의 보호 및 개인정보의 자유로운 이용에 관한 지침'을 제정하였고, 이 규정에 따라 프라이버시의 보호와 유럽연합 국가 내에 정보의 자유로운 흐름을 지향하고 있다.

유럽 국가들의 정보보호 활용 제도는 1995년 제정된 유럽공동체 개인정보관리 지침(EU Directive)을 기본으로 삼아 각국의 사정에 따라 수정 보완된 것이다. 유럽 각국은 전화, 팩스번호, e-메일 주소를 모으고 이를 통해 광고물을 보낼 때 사전 동의 또는 사후 동의를 의무화하고 있다. 이에 따라 정보 수집 업체들은 개인의 전화, 팩스, e-메일 가입 때나 설문조사 등 때 사전 동의를 받거나 광고물을 보낸 뒤 사후 동의를 얻고 있다. 또한, 유럽 대부분의 나라는 광고, 판촉물을 아예 받기를 거부하는 사람들의 이름을 모은 로빈슨 리스트를 만들어 이들에게는 광고물을 보내지 않도록 업계에서 자율규제하고 있다.

독일은 개인정보 리스트를 구입한 회사가 이를 마구잡이로 유통시키는 것을 막기 위해 모든 기업이 정보 리스트를 이용해 광고나 판촉물을 보낼 때는 정부 기관인 레터숍(인쇄, 발송 대행회사)을 통하도록 하고 있다. 또한 유럽 국가들은 최근 개인정보 리스트가 다른 나라의 기업들에 제공되는 사례가 늘어나자 이와 관련된 법, 제도를 만들기 위해 노력하고 있는 중이다. [26][17]

②미국

미국의 개인정보보호는 연방정부기관이 보유하고 있는 개인정보에 관한 보호법규인 1974년의 프라이버시법(Federal Privacy Act 1974)과 각 주단위로 규정된 프라이버시권 관련 법률들이 있다. 미국의 개인정보보호는 공공부문과 민간부문으로 나누어 공공부문에만 법을 적용하고 민간부문에는 원칙적으로 윤리적인 통제만 가능하게 되어 있다.

미국의 개인정보보호제도는 1966년의 정보 공개법(Freedom of Information Act) 제정에 따라 연방정부가 보유하고 있는 정보를 원칙적으로 공개하되 프라이버시법에 의해 정부에 대한 규제를 가하고 민간부문에는 정보의 자유로운 유통을 보장하며 개별 분야에서의 개인정보보호를 목적으로 한 영역별 보호법제를 가지고 있다는 점이 특색이다.

이에 따라 개인정보보호를 위한 많은 개별 법률이 제정되고 있는데 개별법의 장점은 특히 보호가 필요한 개인정보의 취급영역에 한정하여 법적 규제를 행하는 점이라고 하겠으나, 단점으로는 개별 영역별로 법률을 제정하기 때문에 관련업계나 이익단체의 영향을 받을 수 있는 우려가 많다는 점이다.

미국의 경우 프라이버시와 국가보안이 균형 잡힌 IT 보안정책을 전개할 것으로 예상된다. 연방 CTO(Chief Technology Officer: 국가최고기술책임관)를 신설하여 대통령이 직접 IT 보안 방향을 수립하겠다고 주장함으로써 사이버보안에 우선순위를 두겠다는 의지를 표명했다. 국가보안에 대해 강경노선인 사람과 프라이버시 옹호자가 양립하고 있어 다양한 관점이 전략에 반영될 것으로 고려되며 실제 미국 정책을 개인정보 접근규제 및 이용구제를 하는 유럽식 정책 형태로 만드는 한편 프라이버시와 국가보안정책과의 조화를 이룰 것으로 예상된다. [26][17]

③ 일본

일본은 인터넷 시대의 개인정보 대량 유통, 유출사건의 급증 및 프라이버시의 권리 개념의 변모에 의한 필요성에서 2005년에 개인정보보호법을 전면 시행하였다. 일본 개인정보보호법은 주로 EU 개인데이터보호지령 (1995년) 및 OECD 프라이버시 가이드라인 (1980년) 등을 참고하여 작성되었다.

개인정보보호법 제정 이후 긍정적인 면으로는 고객의 개인정보보호를 포함한 정보시스템의 보안에 대한 관심이 높아지고, 투자에 대한 이해도가 상승한 점과 위탁업체에 대한 정보보호 조치를 요구하기 쉬워진 점, 개인정보보호에 관한 보험 상품이 충실해지고 개인정보 관련 인증 취득하려는 기업이 늘어나는 효과를 보고 있다.

하지만, 고객이 개인정보보호에 과민 반응하여 개인정보와 관련된 사고에 대한 일절의 책임을 기업에 넘기는 요구가 많아졌고 개인정보에 관한 민원이 급증하고 종업원에 대한 감시가 강화되어 종업원의 스트레스가 증대되는 한편 정보시스템의 보안에 대한 비용이 상승한 점들이 법 제정 이후 일부 부정적인 면으로 인식되고 있는 점들이다. [26][17]

2.1.3 국내 개인정보보호 법·규제 내용

국내에서 개인정보보호와 연관성을 갖는 법·규제 요건은 굉장히 다양한 종류의 법령 내용 안에 포함되어 있어 공공 및 민간을 포괄하는 개인정보보호 일반법 제정 필요성이 지속적으로 요구되어 왔다. 이에 개인정보보호에 대한 공통적 처리원칙 규정과 국민 권익 구제를 강화한다는 차원에서 OECD 8원칙 등 국제적 기준 반영 및 국내 정책환경을 고려하여 2010년 9월 '개인정보보호법안'이 국회 행정안전위 법안심사소위를 통과했다.

개인정보보호법 제정으로 현행법과 달라지는 내용을 살펴보면 아래 [표 6]과 같다. [22]

[표 9] 개인정보보호법안 제정으로 인한 현행법과 제정안 차이점

구 분	현 행	제 정 (안)
규율대상	o 공공기관, 정보통신사업자, 신용정보 제공·이용자 등 **분야별 개별법**이 있는 경우에 한하여 개인정보보호의무 적용	o **공공·민간 통합 규율로 법적용대상 확대** - 현행법 적용을 받지 않던 오프라인 사업자, 의료기관, 협회·동창회 등 비영리단체, 국회·법원·헌법재판소·중앙선거관리위원회 등으로 확대
보호범위	o 공공기관은 **컴퓨터등에 의해 처리되는 개인정보파일**만을 보호대상으로 함	o 동사무소 민원신청서류 등 **종이문서**에 기록된 개인정보도 보호대상에 **포함**
수집·이용 및 제공기준	o 공공, 정보통신 등 **분야별 개별법**에 따른 처리기준 존재	o **공공·민간을 망라**하는 개인정보 처리원칙과 기준 제시
고유식별정보 처리 제한	o 주민등록번호 등 고유식별정보의 민간사용을 **사전적으로 제한**하는 규정 없음	o **원칙적 처리금지** - 정보주체의 별도 동의, 법령의 근거가 있는 경우 등은 예외 허용

구 분	현 행	제 정 (안)
	o 인터넷상에서 주민등록번호 외의 회원가입방법 제공 의무화 (정보통신서비스제공자 한정)	o 인터넷상 주민등록번호 외의 회원가입방법 제공 **의무화 대상 확대** (정보통신서비스제공자 → 공공기관, 일부 민간분야 개인정보처리자) ※ 대통령령에서 의무화대상 규정
		o 주민등록번호 등 고유식별정보 처리시 암호화 등 안전조치 확보의무 명시
영상정보 처리기기 규제	o **공공기관**이 설치·운영하는 **폐쇄회로텔레비전(CCTV)**에 한하여 규율 - 범죄예방 및 교통단속 등 공익을 위하여 필요한 경우 전문가 및 이해관계인 의견 수렴을 거쳐 설치 - 녹음기능, 임의조작 금지	o 공개된 장소에 설치·운영하는 영상정보처리기기 규제를 **민간까지 확대** - 공개된 장소인 백화점·아파트 등 건물주차장, 상점 내·외부 등에 영상정보처리기기를 설치할 때에는 법령, 범죄예방·수사, 시설안전 및 화재예방, 교통단속 등을 위해서 설치 가능 o 규율대상을 기존 '폐쇄회로텔레비전(CCTV)'에서 **네트워크카메라**도 포함
		o 공중 화장실·목욕탕·탈의실 등 **사생활 침해**우려가 큰 장소는 **설치 금지**
텔레마케팅 등 규제	o「정보통신망법」에 따라 **정보통신서비스제공자**에 한하여 규제 - 마케팅 목적으로 개인정보취급을 위탁하는 경우 정보주체 동의를 받아야 함	o 마케팅을 위해 개인정보처리에 대한 동의를 받을 때에는 다른 개인정보 처리에 대한 **동의와 묶어서 동의를 받지 않도록 명시적으로 규정** - 정보주체가 알기 쉽도록 고지하고 동의를 받아야 함

구 분	현 행	제 정 (안)
		o **모든 개인정보처리자**는 마케팅 업무를 위탁시, 정보주체에게 위탁업무 내용 및 수탁자를 고지해야 함 (정보통신서비스제공자 → 모든 개인정보처리자로 규제대상 확대)
개인정보파일 등록·공개 및 영향평가	o 공공기관이 개인정보파일 보유시 행정안전부장관과 **사전협의**	o 공공기관이 개인정보파일 보유시 행정안전부장관에게 **등록**
	o 행안부장관은 사전협의파일 **관보 공고**	o 행안부장관은 등록사항 **공개**
		o 공공기관 대규모 개인정보파일 구축 등 침해위험이 높은 경우에는 **사전영향평가 실시 의무화**(민간은 자율시행)
유출 통지	o 관련 제도 없음	o 개인정보 **유출사실 통지** 의무화
위원회	o 총리 소속 공공기관개인정보보호심의위원회 - **공공부문** 정책 심의	o 총리 소속 개인정보보호위원회 설치 - **공공·민간부문** 개인정보보호정책을 심의하는 기구
	o 개인정보분쟁조정위원회 - **민간**분야 분쟁조정	o 개인정보분쟁조정위원회 기능 확대 - **공공·민간** 분쟁조정

2.2 개인정보보호 관련 기술

개인정보보호를 위한 관련 기술로는 개인정보침해 기술인 PIT(Privacy Invading Technology)와 개인정보보호기술인 PET(Privacy Enhancing Technology), 프라이버시 노출관리 기술이 있다.

2.2.1 PIT(Privacy Invading Technology)

PIT(Privacy Invading Technology)는 컴퓨터 환경 내 개인정보 관련 오·남용 또는 악의의 피해가 발생할 수 있는 분야를 기술적 관점에서 체계적으로 분석하고 대응할 수 있는 구성을 말한다. [26][7]

① TCP/IP 주소

TCP/IP 주소의 분배 및 관리체계 특성 때문에, 인터넷 이용 시 TCP/IP 주소를 추적하여 이용자 신원을 확인하는 것이 용이하다.

② 도메인 네임

E-mail의 출처를 확인하는 것은 매우 간단하며, 누구나 ISP 정보와 e-mail 이용자의 ID를 알 수 있다. ISP는 이용자의 ID를 이용하여 이용자의 계정을 확인하게 된다.

③ Processor Serial Number (PSN)

Intel사는 자사가 개발하는 Pentium III 칩에 고유의 프로세서 일련 번호(serial number)를 부여하여, 인터넷에 접속하는 특정 컴퓨터의 이용자의 신원 정보와 연결시켜 전자상거래에 있어서 인증목적으로 이용한다.

④ IPv6

IPv6의 계획은 인터넷상의 모든 장치에 고정된 주소를 할당 하는 것으로, IPv6의 새로운 주소는 하드웨어 속에 내장될 것이고, 추적 가능한 정보를 포함하게 된다. 이것은 마치 영구적인 쿠키를 심는 것과 동일한 개념이다.

⑤ 쿠키(cookie)

쿠키 파일을 이용하여 인터넷 이용자의 신원을 쉽게 파악 할 수 있다. 두 가지 방식으로, 로그인 정보(이름, 주소, 비밀번호 등)를 불러내는 데에 사용될 수 있다. 또한 쿠키에 담긴 정보와 마케팅 데이터베이스에 있는 이용자의 이름, 주소, 이전의 소비정보등을 상호 비교함으로써 이용자의 신원 확인 가능하다.

⑥ 웹 버그(Web bug)

웹 버그는 온라인 이용자가 모르는 사이에 이용자에 관한 정보를 유출해가거나 이용자의 시스템을 파괴 할 수도 있는 기술이다. 웹 버그는 Web Page에 심어 놓은 매우 작은 그래픽이미지 파일로, 통상 해당 Web Page의 바탕화면과 같은 색을 지니기 때문에 육안으로는 거의 보이지 않는다.

⑦ 스파이웨어(spyware)

무료 또는 유료로 배포되는 소프트웨어에 들어 있는 일종의 프로그램 모듈을 통칭하는 것으로, 해당 소프트웨어를 설치한 컴퓨터 이용자가 인터넷을 서핑 할 때 이용자의 개인정보나 온라인 활동 정보를 스파이웨어를 설치한 회사의 서버에 지속적으로 전송하는 것이 주된 기능이다.

⑧ 고성능 스파이웨어(sophisticated spyware)기술

스파이웨어 기법이 한 단계 진보한 기법으로, 스파이웨어를 탐지하기 위해 사용되는 백신이나 안티 스파이웨어 솔루션 등을 우회하기 위해, 스파이웨어를 통해 수집된 정보를 작은 크기로 나누어 컴퓨터의 파일 시스템 상에 보이지 않는 틈새공간(slack space)에 임시 저장한 다음, 특정 시간대에 내외부의 특정인에게 전송하는 방법을 이용한다. 이러한 기법은 정부의 수사기관에서 범죄자의 감시 및 경쟁사에 대한 정보수집, 국가간첩 정보 수집에 이용된 사례가 있다.

⑨ WLAN 환경

WLAN 사용자가 액세스 포인트에 접속할 때, 해커가 가상의 액세스 포인트를 이용하여 사용자의 중요한 개인정보를 모니터링 한다.

⑩ 웹메일의 첨부 파일 유출

웹메일 첨부파일 유출기법은 기존 e-mail이나 웹메일을 모니터링하여 데이터를 유출하는 방식에서 한 단계 진화하여, 웹메일에 첨부된 파일을 encoding하는 방식으로 주로 기업이 운용하는 메일 모니터링 프로그램을 우회하여 기밀 정보를 유출하는데 사용된다.

⑪ Steganography

이미지 및 오디오 파일에 중요한 파일이나 메시지를 첨부할 수 있는 스테가노그래피 기법이 확산 될 전망이다. 이 기법은 오사마빈라덴이 알카에다 조직원과의 연락을 위해 사용된 것으로 보고되면서 널리 알려졌다.

⑫ 접속세탁(connection laundering)

접속세탁기법은 해커들이나 해커 그룹간 공간 창조를 통해, 해커 역추적 경로 파악을 어렵게 만드는 것으로, 해커가 여러 국가를 경유하여 해킹을 할 경우, 중간 단계에 해커 그룹이 운용하는 가명경로(anonymizer)를 거쳐 해커에 대한 역추적이 불가능하게 하는 방법이다.

⑬ 위치측정 정보 침해

GPS 또는 휴대전화기의 위치 측정 내용을 인터넷을 통해 확인 할 수 있게되어, 개인의 위치 정보가 유출되어 개인의 신변에 위협이 될 수 있다.

2.2.2 PET(Privacy Enhancing Technology)

개인정보보호기술은 이미 다양한 솔루션이나 기술이 상당수 개발되어 있고 또한 진행 중에 있다. 대표적인 기술로는 익명화 기술, W3C(the World Wide Web Consortium)에서 개발한 P3P, OECD에서 개발한 프라이버시정책생성기(Privacy Policy Statements Generator), 사용자들이 쿠키 수용여부를 결정하며 저장된 정보가 공개될 수 있는지를 판단하는 쿠키 관리 통제(Cookie Manager or Blockers) 기술, 암호화를 통한 전자메일 메시지, 저장된 파일, 온라인에서 커뮤니케이션을 보호할 수 있게 하는 기능을 제공하는 암호화 소프트웨어(Encryption Software) 등이 존재한다. 이렇게 대표적인 개인정보보호기술은 프라이버시보호를 위한 효율적인 방법 중 하나로 구분되어 개발하고 발전되고 있다. 또한 최근에는 개인정보침해기술에 대응하여 크게 6개 영역(에이전트기반기술, 웹 기반 익명성 제공기술, 네트워크 기반기술, 암호화 기술, 정책협상 기술, 내부정보보안기술)에 걸쳐 세부 개인정보보호기술을 분류될 수 있다. [26][27]

[표 10] 개인정보보호 기술

분 야	방 법
웹 기반의 익명성 제공 기술	정보의 노출 자체와는 무관하게 정보와 소유자 간의 관계나 송수신자 간의 관계를 비밀로 하여 사용자의 개인정보보호를 제공하는 기술로 사용자들 간의 비연결성을 통하여 익명성을 제공하는 기술
에이전트 기술	개인정보보호를 위한 에이전트(agent)는 사용자가 파악하기 쉽지 않은 인터넷상에서의 정보 유출에 대해 사용자를 대신하여 통제해 주는 역할
네트워크 기반 기술	현실적으로 가장 빈번하게 일어나는 개인정보 침해 사고들은 네트워크 환경에서 정보를 전달할 때 중간에 가로채거나 수정하거나 또는 단순히 그 데이터를 보기만 하는 행동들에 의해 발생하며 이를 예방하는 기술

분 야	방 법
정책협상기술 (P3P)	웹사이트 운영자에게 이용자 자신의 정보를 관리할 수 있는 권한을 넘겨주는 것이며 이용자 정보가 잘못된 방법으로 사용되지 않도록 보호하기 위해 만들어짐. 따라서 웹 브라우저나 다른 사용자 도구로 하여금 자동적으로 해당 웹사이트의 프라이버시에 관한 정보를 읽고 사용자가 미리 설정해 놓은 정보공개 수준과 비교하여 정보를 선별적으로 제공
암호화 기술	암호화를 통해 자신의 전자 메일 메시지, 저장된 파일, 그리고 온라인에서의 커뮤니케이션을 보호할 수 있게 하는 기능을 제공함. 한번 암호화가 이루어지면 오직 그 당사자만 암호화된 정보에 대한 디지털 키를 가지고 그 정보를 열람하며, 디지털 키는 브라우저, 생체인증, 스마트카드 등과 결합하여 생성됨
내부 정보보안기술	주요 기술정보, 개인정보, 국가기밀 등 이권에 관계된 정보가 유출됨을 보호하는 기술. 대표적으로 정보유출 주체에 정보접근권한자를 배제한 내부자로 한정한 기술과 내부 통신 내용을 모니터링하거나, 시스템 내부에서 일어나는 기술적인 침입을 탐지/방어하는 기술을 탑재.

2.2.3 프라이버시 노출 관리 기술

프라이버시 노출관리 기술은 외부 송신 또는 웹 서비스 등록 시 개인정보 또는 프라이버시 침해정보 노출을 방지하기 위해 전체 정보를 저장하거나 사전차단, 사후 점검하는 기술을 모두 포함하며, 프라이버시 보호를 위한 스팸 필터링 기술도 여기에 포함된다. [11]

① 프라이버시 아카이빙 기술

외부에 유출 또는 시스템에 등록되는 프라이버시 정보를 저장하고, 검색하는 기술을 의미한다. 이 기술은 정보유출 방지를 위한 아카이빙 기술을 적용하고 있으나, 현재의 아카이빙 기술은 저장된 정보 중 필요로 하는 특정 정보를 검색하는 데 걸리는 시간이 가장 큰 기술적 한계이므로, 이를 해결하기 위해 프라이버시 보호 정책에 따라 특정 범위 또는 패턴을 따르는 콘텐츠에 대해서만 아카이빙을 수행하도록 함으로써 개인정보 또는 프라이버시 침해정보에 대한 검색이 용이하도록 하는 아카이빙 기술이나 이에 따라 검색속도를 증가시키는 기술에 대해 지속적으로 연구가 이루어지고 있다.

② 프라이버시 필터링 기술

외부에 유출 또는 시스템에 등록되는 정보를 검사하여 특정 패턴에 대해 차단하는 기술을 의미한다. 일반적으로 정보유출 방지를 위한 필터링 기술을 적용하지만, 최근 기술은 기존의 키워드 검사에 따라 필터링을 하는 것이 아니라 프라이버시 보호와 운영 합리성을 위해 정책과 여러 조합별 규칙을 이용하여 필터링을 수행하는 기술이 개발되어 수행되고 있다. 또한 이 기술에는 프라이버시 보호를 위한 스팸 필터링기술도 포함되는데,스팸 필터링의 정확도를 향상시키기 위해 다양한 인공지능 기법을 적용한 연구가 최근까지 지속적으로 수행되고 있다. 개인정보 필터링은 보다 상세하게는 개인정보 필터링기술, 프라이버시 침해정보 필터링기술, 스팸 필터링기술 등 3가지로 나눌 수 있으며, 이 세 가지 기술 중 개인정보

필터링기술과 프라이버시 침해정보 필터링기술은 특정 패턴을 비교하는 시그니처 기반의 기술만이 개발되어 있고, 스팸필터링기술은 시그니처, 휴리스틱, 블랙리스팅 등의 비학습 기반 필터링 기술에서 베이지언, KNN (K-Nearest Neighbor), SVM(Support Vector Machine), 신경망 등 다양한 학습 기반 알고리즘을 적용한 필터링 기술이 개발되고 현재도 지속적으로 연구되고 있다.

③ 프라이버시 스캐닝기술

현재 운영 중인 PC, 내부 시스템, 내외부 웹사이트 등에 포함된 개인정보 또는 프라이버시 침해정보를 검색 하는 기술을 의미한다. 이 기술은 시스템 검색 및 웹사이트 검색 기술을 주로 활용하고 있으며, 특히 웹사이트를 대상으로 한 프라이버시 스캐닝 기술의 경우, 인터넷의 특성 상 이 웹사이트를 통해 노출된 개인정보가 포함된 웹페이지가 외부 웹사이트에 링크될 수 있으므로, 외부 링크를 검사하는 기술까지 포함된다. 또한 최근에는 스크립트 형 웹사이트와 같이 웹사이트마다 다르게 개발되어 스캐닝이 어려운 웹사이트에 대해서도 스캐닝을 수행하는 가상화 기술까지 개발이 진행되고 있다.

[표 8]은 위에서 살펴본 프라이버시 노출관리 기술에 대한 요약이다.

[표 11] 프라이버시 노출관리 기술 요약

세부기술	특징
프라이버시 아카이빙 기술	In-bound 또는 Out-bound를 통해 유출입되는 데이터 중 특정 영역 또는 프라이버시 징후의 정보만을 저장하여, 추후 검색하는 기술
프라이버시 필터링 기술	In-bound 또는 Out-bound를 통해 유출입되는 데이터 중 개인정보 또는 프라이버시 침해정보 패턴을 검사하여 유출입을 차단하는 기술
프라이버시 스캐닝 기술	특정 시스템에 대해서 개인정보 또는 프라이버시 침해정보 포함여부를 검사하고 관리하는 기술

III. 개인정보 탐지 및 위험분석 모델

3.1. 개요 및 구성도

PDRA(Privacy Detection and Risk Analysis) 모델은 PC에 기록된 개인정보 항목을 탐지하고 그에 따른 개인정보 중요도를 평가하여 사용자에게 개인정보 유출 시 발생할 수 있는 위험수준을 고지하고, 사용자의 선택에 따라 파일 삭제 혹은 암·복호화를 통해 개인정보를 안전하게 보호할 수 있도록 한다.

PDRA는 총 4가지 구성으로 이루어져 있으며 이에 대해 간략히 살펴보면 다음과 같다.

① PIDM(Privacy Information Detection Mechanism)

사용자가 설정한 경로에서 개인정보 노출 및 오·남용 위협이 있는 일반 text, binary 파일 및 복합 문서 등을 추출한다.

② PIPM(Privacy Information Patten Analysis & Parsing Mechanism)

PIDM에서 Gathering 된 파일에서 사용자의 사회적 정보, 금융정보 및 일반정보에 대한 항목을 시스템에 설정한 패턴을 통해 탐지한다.

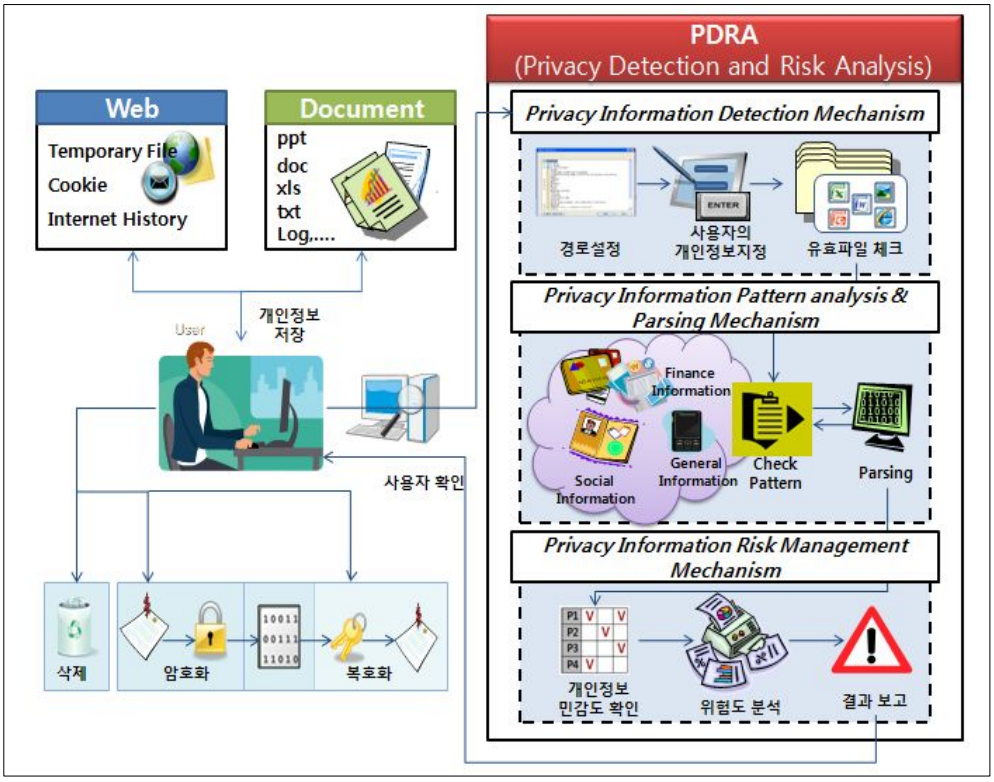
③ PIRM(Privacy Information Risk Management Mechanism)

각 파일에서 발견된 개인정보에 대해 개인정보 민감도와 개인정보의 조합을 토대로 위험수준을 측정하여 사용자에게 고지한다.

④ CEM(Crypto Enhanced Mechanism)

사용자의 선택에 따라 개인정보 노출 위협이 있는 파일을 삭제하거나 암호화함으로써, 해킹 및 불법적인 접근을 통한 개인정보 유출을 방지하며, 필요시 RDCM에서 암호화한 파일을 다시 복호화 하여 파일에 대한 정상적인 접근이 가능하도록 한다.

아래의 [그림 2]는 PDRA모델의 전체 구성도이다.



[그림 2] PDRA 모델 구성도

3.2 PDRA(Privacy Detection and Risk Analysis) 모델

PDRA는 개인정보의 탐지 및 위험분석을 위해 아래와 같은 4가지 주요 메커니즘으로 구성되었다.

3.2.1 PIDM(Privacy Information Detection Mechanism)

사용자 PC에서 개인정보의 노출 위협이 있는 파일을 Gathering 하는 역할을 담당하며 사용자가 직접 탐색경로를 설정하여 탐지 범위를 조정할 수 있도록 한다. 또한 PDRA에 적용된 개인정보 외에 기타 정보에 대한 입력을 받아 탐지하고자 하는 개인정보 항목을 유동적으로 관리할 수 있으며 시스템의 오탐과 속도저하를 방지하기 위해 파일의 유효성과 크기를 검사하도록 한다.

제안한 메커니즘에서 진단하고 있는 검색대상 파일 포맷은 아래 [표 9]와 같다.

[표 12] 검색대상 파일 포맷

분류	파일형식	
MS Office	Microsoft Word	Microsoft Word 97 (doc) Microsoft Word 2000 (doc) Microsoft Word XP (doc) Microsoft Word 2003 (doc) Microsoft Word 2007 (docx)
	Microsoft Excel	Microsoft Excel 97 (xls) Microsoft Excel 2000 (xls) Microsoft Excel XP (xls) Microsoft Excel 2003 (xls) Microsoft Excel 2007 (xlsx)
	Microsoft PowerPoint	Microsoft PowerPoint 97 (ppt)

분류	파일형식	
		Microsoft PowerPoint 2000 (ppt) Microsoft PowerPoint XP (ppt) Microsoft PowerPoint 2003 (ppt) Microsoft PowerPoint 2007 (pptx)
Image File	Microsoft Document Imaging (mdi) Adobe Acrobat PDF ver 1.x (pdf)	
Internet File	HTML/HTM file (html, htm) MIME HTML file (mht) XML file (xml) MS Compiled HTML (chm)	
ETC	Text file(txt) Initial File (ini) Log file (log) Rich Text Format (rtf) comma separated value (csv) ZIP (zip)	

[표 9]에 나열된 포맷을 대상으로 1차적으로 파일을 추출하여 파일의 유효성에 대한 확인을 거쳐 다음 단계인 PIPM(Privacy Information pattern analysis & Parsing Mechanism)에게 파일의 정보를 넘겨주게 된다.

PIDM의 알고리즘은 다음과 같다.

[표 13] PIDM 알고리즘

Algorithm Privacy Information Detection Mechanism

```
// 개인정보 속성
private string[] communication = { "ID", "Password", "E-maile" };
private string[] identity = { "성명","주민등록번호","전화번호","집주소","운전면허번호","
                                                                    여권번호" };
private string[] finance = { "계좌번호", "거래은행","신용카드번호","신용카드비밀번호","
                                                                    카드유효기간" };
private string[] direct = { "P1", "P2", "P3", "P4"};

//디렉토리에서 파일정보 읽음
private void SearchDirectory(string DirectoryPath)
{
        DirectoryInfo directory = new DirectoryInfo(DirectoryPath);
        FileInfo[] file = directory.GetFiles("*.*");

        foreach (FileInfo f in file)
        {
            if (f.Length < 10 *1024 * 1024&& ck_file(Path.GetExtension(f.Name)))
            {
                fileRead1(f.FullName);
            }
        }

        DirectoryInfo[] dir = directory.GetDirectories();
        foreach (DirectoryInfo d in dir)
        {
            this.SearchDirectory(d.FullName);
        }
}

//유효한 파일 체크
private bool ck_file(string exten)
{
    string[] compare = new string[] { ".txt", ".doc", ".docx", ".ppt", ".pptx", ".xlsx", ".pdf",
                                      ".mht", ".xml", ".chm", ".ini", ".rtf", ".csv", ".zip",
                                      ".log" };
    bool flag = false;
    foreach (string a in compare)
    {
        if (exten.Equals(a))
        {
            flag = true;
            break;
        }
        else flag = false;
    }
    return flag;
```

```
}

//사용자의 개인정보 지정
private void userInsert(object sender, EventArgs e)
{
    if (userInsert1.SelectedIndex != 0)
    {
        userInsert2.Visible = true;
        userInsert2.Items.Clear();
        if (userInsert1.SelectedIndex == 1)
        {
            userInsert2.Items.AddRange(communication);
            userInsert2.SelectedIndex = 0;
        }
        else if (userInsert1.SelectedIndex == 2)
        {
            userInsert2.Items.AddRange(identity);
            userInsert2.SelectedIndex = 0;
        }
        else if (userInsert1.SelectedIndex == 3)
        {
            userInsert2.Items.AddRange(finance);
            userInsert2.SelectedIndex = 0;
        }
        else
        {
            userInsert2.Items.AddRange(direct);
            userInsert2.SelectedIndex = 0;
        }
    }
    else
    {
        userInsert2.Items.Clear();
        userInsert2.Visible = false;

    }
}
```

3.2.2 PIPM(Privacy Information pattern analysis & Parsing Mechanism)

사용자의 개인정보 항목을 해당 패턴을 통해 탐지 및 분석하는 기능으로, 검사항목은 주민번호, 카드정보, 계좌정보, 핸드폰번호 이며 PID에서 입력받은 키워드는 기타정보 항목으로 설정한다.

각 항목에 대한 분석은 다음과 같다.

① 주민번호

개인정보가 유출될 경우 가장 큰 피해를 입을 수 있는 항목이다. 주민등록번호는 개인마다 고유하게 부여받은 번호로 실생활에서는 본인확인을 위해 주로 주민등록증을 사용하는 반면, 각종 문서나 통신상에서는 이 번호를 사용한다.

[표 14] 주민등록번호 구조

Y	Y	M	M	D	D	-	A	B	C	D	E	F	G
탄생년		탄생월		탄생일			성별	지역번호				순서	코드

- YYMMDD 여섯자리는 생년월일이며 탄생년도는 뒤의 두 자리로 표현한다.
- A는 성별을 의미하며, 총 6가지 경우가 존재한다.
  -9: 1800 ~1899 년에 태어난 남성
  -0: 1800 ~1899 년에 태어난 여성
  -1: 1900 ~1999 년에 태어난 남성
  -2: 1900 ~1999 년에 태어난 여성
  -3: 2000 ~2099 년에 태어난 남성
  -4: 2000 ~2099 년에 태어난 여성
- BCDE는 지역번호를 의미한다. 이는 주민등록지이며 행정자치부에서 읍, 면, 동마다 고유하게 부여한 번호로 구성된다.
- F는 신고 순서로, 주민등록지에서 해당 생년월일로 신고된 순서를 의미한다.

• G는 오류검출 코드, 즉 검증번호로 앞의 숫자들을 조합하여 계산한 결과이다.

PIPM에서 탐지하는 주민등록번호의 패턴은 아래 [표 12]와 같다.

[표 15] 주민등록번호 탐지 패턴

저장 형식	패턴
구분자(-)를 기준으로 양쪽에 공백이 존재하는 경우	****** - *******
구분자(-)를 기준으로 양쪽에 공백이 존재하지 않는 경우	******-*******
구분자(-)가 존재하지 않는 경우	**********
구분자가 공백()으로 존재하는 경우	****** *******

위의 패턴으로 탐지한 주민번호에 대해 올바른 번호인지 검출하기 위해 오류검출 코드G를 이용해 확인 과정을 거치도록 한다.

② 계좌번호

계좌번호의 분석은 금융결제원에서 제공하는 은행별 계좌번호 체계를 참고하여 국내 9개 은행에 대한 패턴을 설정 하였다.

PIPM에서 탐지하는 9개 은행과 그에 해당하는 계좌번호 체계는 다음과 같다.

[표 16] 계좌번호 체계

은행	계좌번호체계	계정과목코드
국민은행	□□□-■■-□□□□-□□□	01, 04, 05, 21, 24, ,25, 26
	□□□□□□-■■-□□□□□□	01, 02, 04, 25, 37, 90
외한은행	■■■-□□□□□□-□□□	600, 601, 610, 611, 620, 621, 630, 631
신한은행	□□□-■■-□□□□□□	01, 02, 03, 04 05,11, 12, 13
	■■■-□□□-□□□□□□	100, 120, 140, 150, 160
우리은행	□■■■-□□□-□□□□□□	002, 003, 004
	□□□-□□□□□□-■■-□□□	01, 02, 03, 04, 12, 13, 15
기업은행	□□□-□□□□□□-■■-□□□	01, 02, 03, 04, 06, 07, 13
산업은행	■■■-□□□□-□□□□-□□□	011, 013, 019, 020, 022
농협	□□□□□□-■■-□□□□□□	51, 52, 55, 56, 66, 67
	□□□-■■-□□□□□□	01, 02, 05, 06, 12, 17

은행	계좌번호체계	계정과목코드
하나은행	□□□-□□□□□□-□□□■■	01, 02, 04, 05, 07, 08
제일은행	□□□-■■-□□□□□□	10, 20, 30

위 [표 13]의 계좌번호체계에서 검은색으로 표시한 곳이 은행을 구분 지을 수 있는 과목번호를 가르키며 해당하는 과목코드는 계정과목코드에 나열해 놓았다.

③ 카드번호

신용카드번호의 분석은 카드번호의 구조와 카드번호 검증을 통해 패턴분석을 한다.

[표 17] 카드번호의 구조

1	2	3	4	-	5	6	7	8	-	9	10	11	12	-	13	14	15	16
Bin 번호 (Bank Identifier Number)							발행기관의 일련번호											검증

첫째 자리부터 여섯째 자리까지의 번호(BIN 번호)는 해당 카드가 어느 나라의 어느 카드사가 발급한 카드인지, 카드 회원이 일반, 골드, 개인, 법인인지 알 수 있도록 되어 있다. 그 다음 7번째 자리부터 15번째 자리까지는 각 카드사가 임의의 규칙에 따라 사용하도록 되어 있고 16번째 숫자는 특정한 공식에 의해 카드번호를 검증하는 값이다.

PIPM는 16번째 검증번호를 통해 비자/ 마스터/ JCB등의 국제카드를 검증하도록 설계하였으며, 국내전용카드는 검증번호 산출 방식이나 위치가 카드사별로 상이하게 발급되고 있으므로, Bin번호를 통해 분류하고 있다.

[표 15]는 신용카드의 시작번호에 따른 카드종류의 구분을 나타낸다.

[표 18] 카드번호 패턴

카드시작번호	Bin 번호
3	**35(JCB카드)** 3562 96 - 신한카드 (체크) 3569 11 - 국민은행 (실버) 3569 12 - 국민은행 (골드) **36(다이너스카드)** 3616 - 현대카드 **37(아메리칸 엑스프레스)** 3779 82 - 신한카드 3791 83 - 삼성카드
4 (비자카드)	4028 57 - 현대카드의 Q 멤버스카드 4046 78 - 신한카드 비씨카드 (실버) 4063 57 - 씨티은행의 신세계 플래티늄카드 4057 53 - 신한카드 (선불) 4182 36 - 외환은행의 시그니쳐 카드 4214 17 - 수협중앙회 (체크) 4214 20 - 우리은행 비씨카드(체크) 4214 68 - 롯데카드 (체크) 4227 27 - 씨티은행 비씨카드 (체크) 4386 76 - 신한카드 (플래티늄) 4432 33 - 구,LG카드 4450 - 비자카드 시그니처카드 4473 20 - 우리은행 비씨카드 (체크) 4499 - 신한카드 (체크), 외환은행 (체크) 4512 45 - 삼성카드 (체크) 4518 42 - 신한카드 (실버) 4518 45 - 신한카드 (체크) 4579 72 - 국민은행 (실버) 4579 73 - 국민은행 (골드) 4585 32 - 삼성카드 (체크) 4599 00 - 외환카드 (실버) 4628 90 - 씨티은행의 씨티 프리미어마일 카드 4658 87 - 신한카드 (플래티늄)

카드시작번호	Bin 번호
	4658 89 - 수협중앙회 (체크) 4705 88 - 삼성카드 (체크) 4902 20 - 우리은행 비씨카드 (신용) 4906 XX - 비씨카드
5(마스타카드, Maestro)	5124 62 - 롯데카드 (골드) 5148 76 - 씨티은행 (플래티늄) 5155 94 - 신한카드 (플래티늄) 5238 30 - 외환카드 (체크) 5387 20 - 우리은행 비씨카드 5388 XX - 비씨카드 5409 26 - 국민은행 (골드) 5522 20 - 우리은행의 플래티늄 비씨카드 5543 46 - 국민은행 프리패스카드 <발급중단> 5021 23 - 국민은행 (체크) 5029 28 - 하나SK카드 (체크) 5886 44 - 씨티은행 국제직불카드 5898 - 우리나라의 모든 Maestro직불카드 6060 45- 신한카드 6360 XX - 비씨카드 6361 89 - 하나SK카드의 티드림체크카드
9 (국내 고유 카드)	9409 51 - 롯데카드 (체크) 9410 XX - 비씨카드 (실버 일반) 9410 61 - 신한카드 (신용) 9420 XX - 비씨카드 (실버 우량) 9420 61 - 신한카드 (체크) 9420 90 - 삼성카드 (선불) 9430 XX - 비씨카드 (골드) 9436 45 - 국민은행 (체크) 9440 XX - 비씨카드 (체크) 9441 16 - NH카드 (체크) 9445 41 - 국민은행 (실버) 9445 42 - 국민은행 (골드) 9445 47 - 국민은행 (선불) 9490 28 - 현대카드 (체크)

④ 전화 번호

최근 이슈가 되고 있는 보이스피싱 및 스팸 문자는 유출된 개인정보로부터 획득한 전화번호를 이용한다. 따라서 [표 16]과 같이 전화번호 종류에 따른 패턴을 탐지하도록 하였다.

[표 19] 전화번호 패턴

종류	구분	번호	구분	번호	패턴
유선전화	서울	02	울산	052	지역번호-(*)***-****
	경기	031	대구	053	
	인천	032	경북	054	
	강원	033	경남	055	
	충남	041	전남	061	
	대전	042	광주	062	
	충북	043	전북	063	
	부산	051	제주	064	
휴대전화	SKT	011	KTF	016	통신사-(*)***-****
		017		018	
	LGT	019	3G 통합	101	

⑤기타정보

기타정보는 사용자가 직접 지정한 개인정보 항목으로, 정해진 패턴을 가지고 있지는 않다. 따라서 입력받은 키워드를 직접 탐지하도록 한다.

사용자는 탐지하고자 하는 개인정보 항목을 입력 할 때 개인정보의 태그 혹은 개인정보의 민감도를 함께 입력한다. 이러한 민감도 설정에 대한 설명은 다음 절에서 다루며 아래 [표 17]은 기타정보에 지정 할 수 있는 개인정보 태그이며 PIPM에서 자동으로 탐지하는 주민번호, 계좌번호, 신용카드번호, 전화번호 또한 포함 하고 있다.

[표 20] 사용자 지정 개인정보 항목

구분	항목
통신정보	ID
	Password
	Email
신상정보	성명
	주민등록번호

구분	항목
	전화번호
	집 주소
	운전면허번호
	여권번호
금융정보	계좌번호
	거래은행
	신용카드번호
	신용카드 비밀번호
	카드 유효기간

PIPM의 알고리즘은 다음과 같다.

[표 21] PIPM 알고리즘

Algorithm Privacy Information Patten Analysis & Parsing Mechanism

```
//파읽을 읽음
private void fileRead1(string search_path)
{
        TextReader reader = new FilterReader(search_path);
        if (reader != null)
        {
            using (reader)
            {
                string result_str = reader.ReadToEnd();
                string InputText = result_str;
                string protection = "";
                // 개인정보 검색
                int jumin = SearchJumin(InputText, search_path);
                int credit = SearchCreditCard(InputText, search_path);
                int account = SearchAccount(InputText, search_path);
                int mobile = SearchMobile(InputText, search_path);
                int user = SearchUser(InputText, search_path);
                int sum = jumin + credit + account + mobile + user;
                //위험도 계산
                protection=RiskManagement(jumin,credit,account,mobile,user);

            }
        }
    }

//주민번호 검색
private int SearchJumin(string InputText, string search_path)
{
        string jumin_pattern = @"\d{6}-[1234]\d{6}";
        Regex exp = new Regex(jumin_pattern);
        MatchCollection MatchList = exp.Matches(InputText);
        int sum=0;
```

```
            if (MatchList.Count != 0)
            {
                for (int j = 0; j < MatchList.Count; j++)
                {
                    Match FirstMatch = MatchList[j];
                    //주민번호 유효성 검사
                    if (VarInfo.CheckSocialNumber(FirstMatch.Value))
                    {
                        sum++;
                    }
                }
                return sum;
            }
        return 0;
    }

    //전화번호 검색
    private int SearchMobile(string InputText, string search_path)
    {
            string mobile_pattern = @"01[016789]\-\d{3,4}\-\d{4}";

            Regex exp = new Regex(mobile_pattern);
            MatchCollection MatchList = exp.Matches(InputText);
            if (MatchList.Count != 0)
            {
                return MatchList.Count;
            }
        return 0;
    }

    //카드 검색
    private int SearchCreditCard(string InputText, string search_path)
    {
            string card_pattern = @"(\d{4}-\d{4}-\d{4}-\d{4})|(\d{16})";
            Regex exp = new Regex(card_pattern);
            MatchCollection MatchList = exp.Matches(InputText);
            int sum=0;
            if (MatchList.Count != 0)
            {
                for (int j = 0; j < MatchList.Count; j++)
                {
                    Match FirstMatch = MatchList[j];

                    //카드번호 유효성 검사
                    string CardVal = CardInfo.GetCardIssuer(FirstMatch.Value);
                    if (!CardVal.Length.Equals(0))
                    {
                        sum++;
                    }
                }
                return sum;
            }
```

```
        return 0;
    }

    //계좌번호 검색
    private int SearchAccount(string InputText, string search_path)
    {

            string bank_pattern =
                                @"((\d{6}|\d{4}|\d{3})-(\d{6}|\d{3}|\d{2})-(\d{6}|\d{5}|\d{4}))|(
                                \d{3})-(\d{6}|\d{4}|\d{2})-(\d{4}|\d{2})-(\d{3})";
            Regex exp = new Regex(bank_pattern);
            MatchCollection MatchList = exp.Matches(InputText);
            int sum = 0;
            if (MatchList.Count != 0)
            {
                for (int j = 0; j < MatchList.Count; j++)
                {
                    Match FirstMatch = MatchList[j];

                    //거래은행 검사
                    string BankVal = BankInfo.GetBankIssuer(FirstMatch.Value);
                    if (!BankVal.Equals(""))
                    {
                        ListViewItem item1 = new ListViewItem(search_path);
                        item1.SubItems.Add(BankVal);
                        item1.SubItems.Add(FirstMatch.Value);
                        listView4.Items.Add(item1);
                        sum++;
                    }
                }
                return sum;
            }
        return 0;
    }

    //사용자 지정 개인정보 검색
    private int SearchUser(string InputText, string search_path)
    {
            int sum=0;
            if (userInsert.TextLength != 0)
            {
                string user1 = userInsert.Text;
                if (InputText.Contains(user1))
                {
                    sum++;
                }
              return sum;
            }
        return 0;
    }
```

3.2.3 PIRM(Privacy Information Risk Management Mechanism)

PIPM에서 탐지한 개인정보 항목들을 대상으로 개인정보의 민감도에 따라 위험수준을 측정하는 기능을 담당한다.

개인정보의 속성은 매우 다양하나 정해진 필드를 가지고 있는 DB가 아닌 PC를 대상으로 모든 개인정보를 탐지하는 것은 불가능하다. 따라서 본 메커니즘에서 설정된 개인정보의 등급은 일정한 패턴을 갖고 있는 개인정보와 사용자가 직접 지정할 수 있는 개인정보를 대상으로 하였다.

민감도의 등급은 가장 민감한 등급인 P1에서 가장 낮은 등급인 P4의 단계로 나누었다.

[표 22] 민감도에 따른 개인정보 분류

등급	해당정보	설명	항목
P1	결제 정보 주민등록번호	가장 높은 레벨로 이정보가 제3자에게 노출 될 시에는 금전적 피해가 발생 할 수 있다.	신용카드번호
			비밀번호
			카드유효기간
			주민등록번호
P2	신분확인정보 계좌번호	민감한 정보이며 노출 시 개인정보를 도용당할 수 있다.	운전면허번호
			여권번호
			계좌번호
			거래은행
P3	연락처, 거주정보	어느 정도 공개가 가능한 정보로 구성되지만 사생활 침해가 우려되는 정보이다.	전화번호
			집주소
			E-mail
P4	이름, 아이디	가장 등급이 낮은 레벨로 크게 우려되지 않는 정보이다.	성명
			ID

위험수준의 측정은 위에서 제시한 개인정보의 민감도에 따라 달라지며 하나의 파일 내에서 다수의 개인정보 항목이 조합되어 탐지 될 수 있는 점을 고려하여 다음과 같이 산정하도록 하며 가장 낮은 등급으로 큰 영향을 미치지 않는 수준의 N/A(Not Applicable)에서부터 가장 높은 수준인 Level 8 로 나눈다.

[표 23] 개인정보 위험수준 측정

개인정보 조합	개인정보 조합 설명	위험수준	설명
P4	민감하지 않은 개인정보	N/A	큰 영향을 미치지 않는 수준
P4+P4	민감하지 않은 일반정보가 두 개 이상 조합	Level 1	정확한 개인정보에 대해 알 수는 없지만 개인정보 사칭의 위험이 있음
P3	일반적인 민감정보	Level 2	개인의 신분과 신상정보에 대한 추정이 가능
P3+P4	일반적인 민감정보와 민감하지 않은 일반정보의 조합	Level 3	
P2	신분 및 계좌정보	Level 4	개인의 신분을 알 수 있으며 금전적 피해가 우려됨
P2+P4 P2 +P3	신분 및 계좌 정보와 일반적인 민감정보 이하 등급의 조합	Level 5	
P1	결제 정보 및 주민등록번호	Level 6	실제 금전적 피해를 입을 수 있음
P1+P3 P1+P4	결제 정보 및 주민등록번호와 민감정보 이하 등급의 조합	Level 7	
P1+P2	신상정보와 금융정보 의 조합	Level 8	

[표 20]은 개인정보의 노출 위험수준을 평가하는 방법으로써 단일 개인정보가 아니더라도 조합했을 때 수준이 높은 경우에는 위험 레벨이 높아질 수 있음을 보여준다.

PIRM의 알고리즘은 다음과 같다.

[표 24] PIRM 알고리즘

Algorithm Privacy Information Risk Management Mechanism

```
//위험측정
private string RiskManagement(int jumin, int credit, int account, int mobile, int user)
{
    string risklevel = "";
    int P1 = jumin + credit;
    int P2 = account;
    int P3 = mobile;
    int P4 = 0;
    if (user > 0)
    {
        //사용자 지정 개인정보 위험도 측정
        if (userInsertGrade(userInsert2.Text).Equals("P1")) P1++;
        else if (userInsertGrade(userInsert2.Text).Equals("P2")) P2++;
        else if (userInsertGrade(userInsert2.Text).Equals("P3")) P3++;
        else P4++;
    }
        if (P1 > 0)
        {
            if (P2 + P3 + P4 <= 0) risklevel = "Level6";
            else
            {
                if (P2 > 0) risklevel = "Level8";
                else risklevel = "Level7";
            }
        }
        else if (P2 > 0)
        {
            if (P3 + P4 <= 0) risklevel = "Level4";
            else risklevel = "Level5";
        }
        else if (P3 > 0)
        {
            if (P4 <= 0) risklevel = "Level2";
            else risklevel = "Level3";
        }
        else if (P4 > 1) risklevel = "Level1";
        else risklevel = "N/A";

    return risklevel;
}

//사용자 지정 개인정보 위험도
private string userInsertGrade(string insert)
{
    string grade = "";
    switch (insert)
```

```
        {
            case "신용카드번호":
            case "신용카드비밀번호":
            case "카드유효기간":
            case "주민등록번호":
            case "Password":
            case "P1":
                grade = "P1";
                break;

            case "운전면허번호":
            case "여권번호":
            case "계좌번호":
            case "거래은행":
            case "P2":
                grade = "P2";
                break;

            case "E-maile":
            case "전화번호":
            case "집주소":
            case "P3":
                grade = "P3";
                break;

            case "ID":
            case "성명":
            case "P4":
                grade = "P4";
                break;

            default: grade = "";
                break;
        }
        return grade;
    }
```

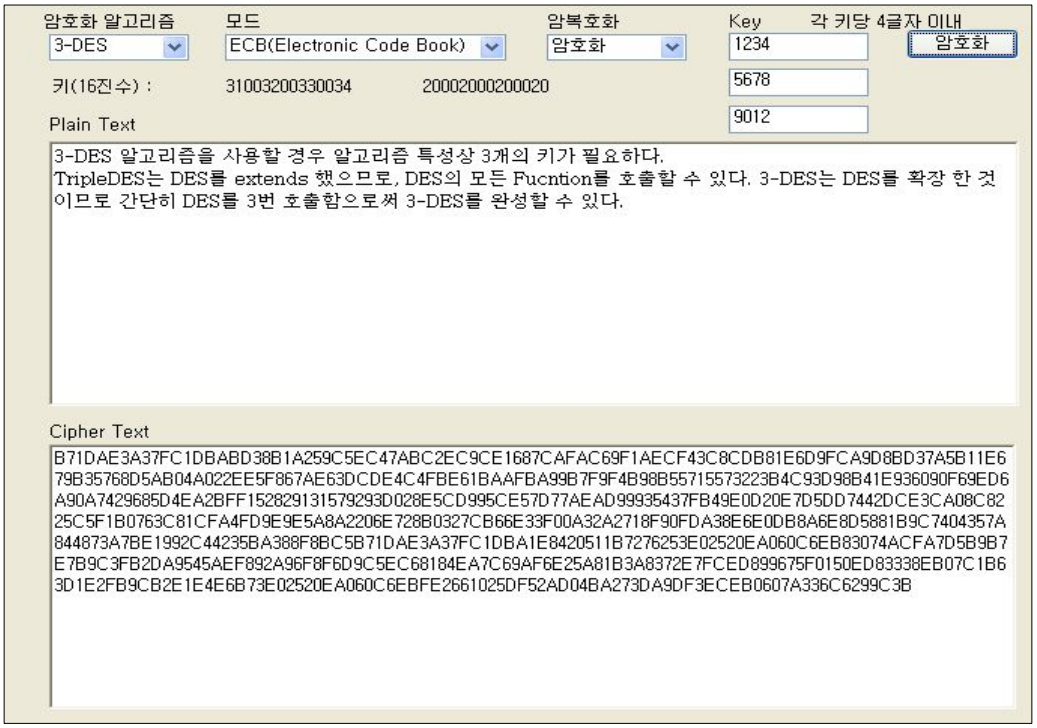
3.2.4 CEM(Crypto Enhanced Mechanism)

개인정보의 노출 위험 수준을 고지 받은 사용자는 선택에 따라 해당 파일을 삭제하거나 암호화 할 수 있다. 대부분의 개인정보는 복합문서에서 검출된다. 복합문서의 경우 사용자의 자의에 의해 작성된 경우가 많으며, 이러한 파일은 재사용의 빈도가 높기 때문에 단순한 삭제보다는 암호화를 통해 안전하게 보관하는 것이 권장된다.

암호화된 파일은 실행은 가능하나 일반적인 방법으로 read/write를 할 수 없기 때문에 해킹이나 악의적인 목적으로 노출 될 경우 개인정보 유출을 막을 수 있다.

또한 사용자는 필요에 따라 암호화한 파일을 key를 통해 복호화 함으로써 파일을 정상적으로 read/write할 수 있다.

CEM 에서는 암•복호화 방식으로 TripleDES를 사용한다.



[그림 3] Triple DES를 이용한 암호화 예제

DES는 Data Encryption Standard의 약자로 데이터를 64bit 블록으로 암호화하고, 복호화 하는 과정에서 64bit의 키를 사용한다. 그러나 이 키가 64bit라고 할지라도 효율적인 키 강도는 56bit로 처리되는 수준이어서, 이러한 DES 알고리즘을 보완하기 위해 Triple DES 알고리즘을 사용한다. DES를 3번 수행하여 입력 데이터를 암호화/복호화 하는 방식이다.

[그림 3]은 Triple DES를 이용해 평문을 암호화 한 예제로 3개의 키를 통해 DES를 3번 수행한 64bit 결과 값을 볼 수 있다.

CEM의 알고리즘은 다음과 같다.

[표 25] CEM 알고리즘

Algorithm Crypto Enhanced Mechanism

```
//암·복호화 기본설정
private enum CryptoAction { actionEncrypt = 1, actionDecrypt };
const string EncrFolder = @"c:\Encrypt\";
const string DecrFolder = @"c:\Decrypt\";

//파일삭제
private void FileRemove(object sender, EventArgs e)
{
        if (checkItem.Count > 0)
        {
                foreach (item in checkItem)
                {
                    string filePath = item.Text.ToString();
                    FileInfo fileInfo = new FileInfo(filePath);
                    fileInfo.Delete();
                }
        }
}

//암호화
private void Encrypt(object sender, EventArgs e)
{
    DirectoryInfo di = new DirectoryInfo(EncrFolder);
    if (di.Exists == false)
    {
        di.Create();
    }
        if (checkItem.Count > 0)
        {
                foreach (item in checkItem)
                {
```

```
                        string filePath = item.Text.ToString();
                        ring filePath1 = EncrFolder + Path.GetFileName(filePath);
                        byte[] byteKey;

                        //암호화 /key 설정
                        byteKey = GetKeyByteArray("key1",“key2”,"key3");
                        byte[] byteInitializationVector;
                        byteInitializationVector = GetKeyByteArray("key1",“key2”,"key3");

                        //암호화 시작
                        EncryptOrDecryptFile(filePath, filePath1, byteKey,
                                        byteInitializationVector, CryptoAction.actionEncrypt);
                    }
            }
    }

    //복호화
    private void Decrypt(object sender, EventArgs e)
    {
        DirectoryInfo di = new DirectoryInfo(DecrFolder);
        if (di.Exists == false)
        {
            di.Create();
        }
            if (checkItem.Count > 0)
            {
                foreach (item in checkItem)
                    {
                        string filePath1 = item.Text.ToString();
                        string filePath = EncrFolder + filePath1;
                        string filePath2 = DecrFolder + filePath1;

                        //복호화 key 설정
                        byteKey = GetKeyByteArray("key1",“key2”,"key3");
                        byte[] byteInitializationVector;
                        byteInitializationVector = GetKeyByteArray("key1",“key2”,"key3");

                        //복호화 시작
                        EncryptOrDecryptFile(filePath, filePath2, byteKey,
                                    byteInitializationVector, CryptoAction.actionDecrypt);

                    }
            }
    }
```

3.3 프로토타이핑

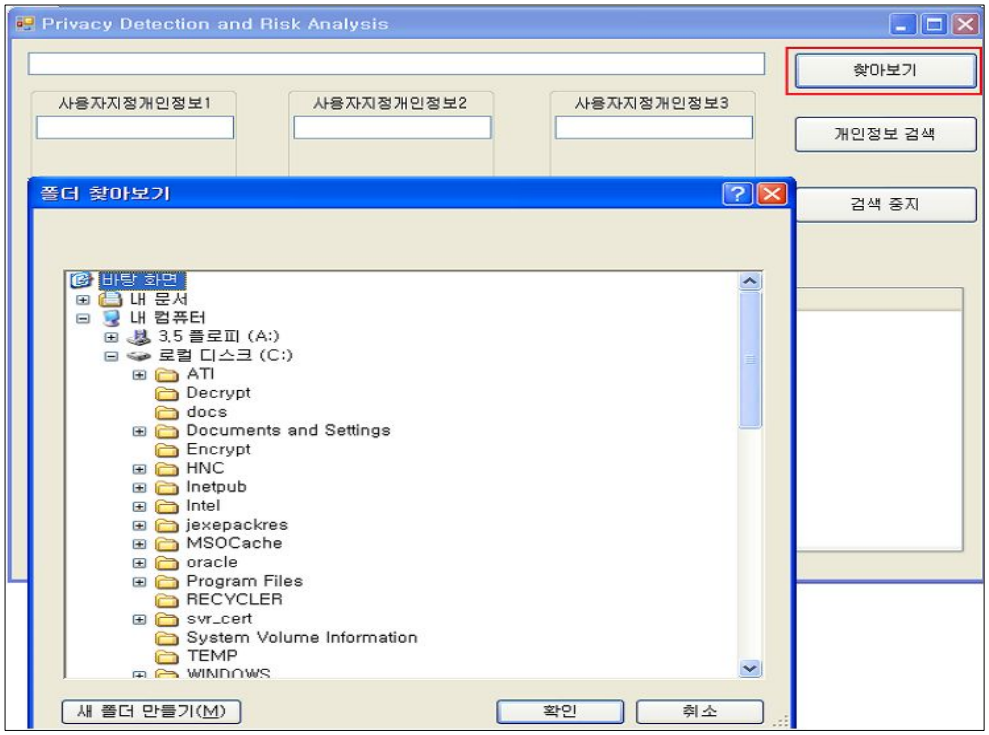
PDRA 모델은 Windows XP 운영체제 하에 C# 언어를 통해 구현하였다. 각 메커니즘별 구현결과는 다음과 같다.

3.3.1 PIDM 구현

PIDM은 사용자가 검색할 대상 디렉토리를 설정하여 해당 디렉토리 내에서 컨트롤 할 수 있는 파일을 추출하는 단계이다.

또한 사용자는 PDRA가 자동으로 검출하는 주민번호, 계좌번호, 신용카드번호, 전화번호 외에 찾고자 하는 개인정보를 직접 지정할 수 있다.

① 검색대상 디렉토리 설정



[그림 4] 검색대상 디렉토리 설정

[그림 4]에 표시해 놓은 찾아보기를 클릭하면 PC의 파일시스템의 구조를 보여주는 FolderBrowser 다이얼로그가 나타나며 사용자는 이를 통해 대상 경로를 지정한다.

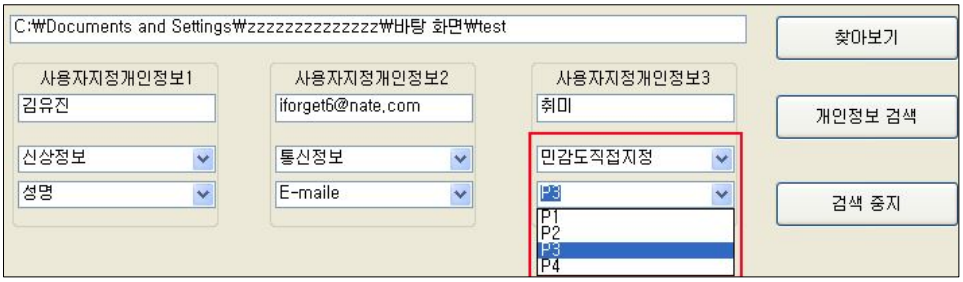
대상 디렉토리는 사용자 PC 내 모든 경로를 포함 할 수 있으며 설정된 경로 안에 속하는 파일 뿐 아니라 하위 디렉토리 내의 파일을 포함한다.

② 사용자 지정 개인정보 설정

[그림 5] 사용자 지정 개인정보 설정

사용자는 탐지하고자 하는 개인정보를 키워드를 통해 3개까지 직접 지정할 수 있다. 필요하지 않을 시에는 입력하지 않아도 되며 키워드를 입력하면 개인정보를 분류할 수 있는 콤보박스가 생성된다.

분류할 수 있는 개인정보는 신상정보, 통신정보, 금융정보 이며 개인정보의 분류를 선택하면 세부 항목을 선택 할 수 있다.



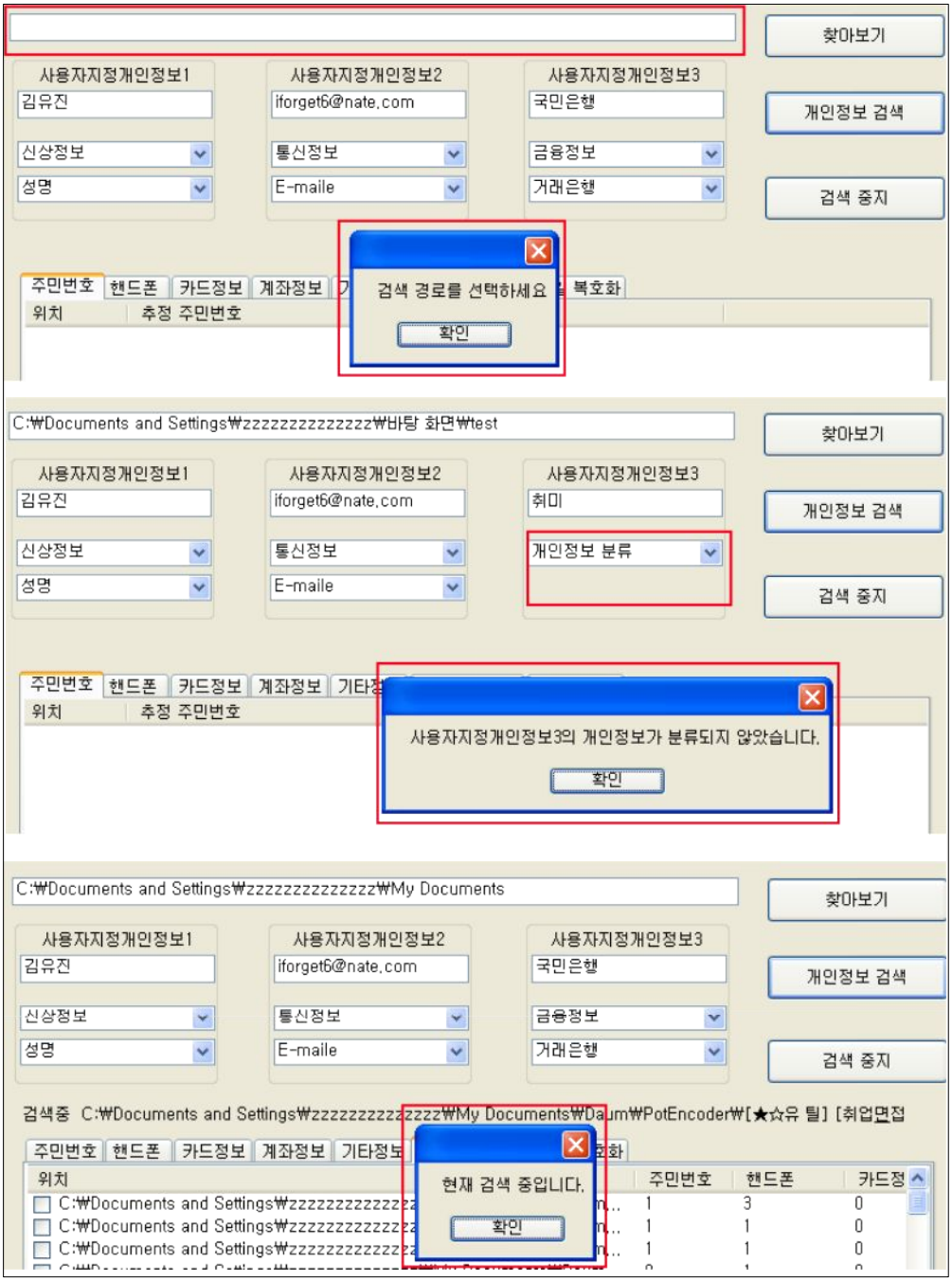
[그림 6] 개인정보 민감도 직접지정

해당하는 개인정보 세부 항목이 없을 시에는 개인정보에 대한 민감도를 직접 설정 할 수 있으며, 이때 민감도는 가장 높은 등급인 P1에서 가장 낮은 등급인 P4의 4단계로 나뉜다.

③ 입력 값 확인

경로설정과 개인정보 입력이 끝나면 개인정보 검색을 클릭하여 다음단계인 PIPM이 진행되어야 한다. 그러나 이 단계가 시작되기 전 사용자가 PIDM에서 설정한 값들을 재확인하여 예외상황을 제어해야 한다. 개인정보 검색이 되지 않는 경우는 다음과 같다.

- 검색 디렉토리를 설정하지 않은 경우
- 사용자 지정 개인정보 입력 후 개인정보 항목을 선택하지 않은 경우
- 이미 검색중인 경우

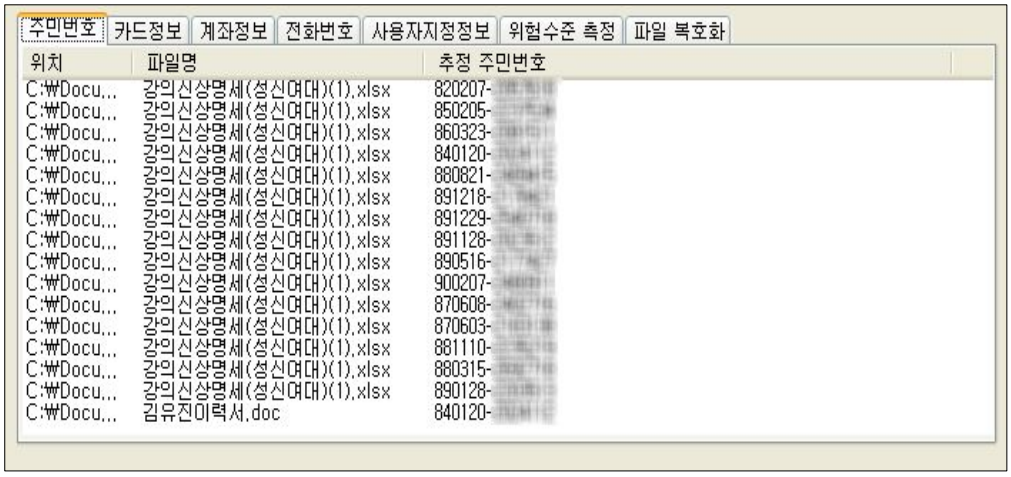


[그림 7] 개인정보 검색 예외 상황

3.3.2 PIPM 구현

파일에서 개인정보 항목을 패턴을 통해 분석 및 분류 하는 기능을 담당하며, 각 항목이 발견 될 때마다 해당하는 탭페이지 리스트에 표시가 된다.

① 주민번호



[그림 8] 주민번호 검색 결과

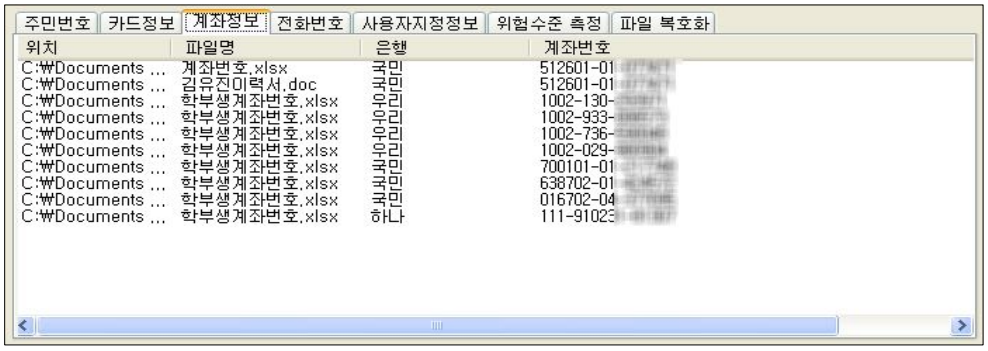
주민번호 검색결과는 리스트에는 탐지된 파일의 위치와 파일 명, 추정 주민번호가 표시된다. [그림 8]에서 보이는 '강의 신상명세' 파일의 경우 여러 명의 주민번호가 탐지 되었다. 이러한 파일이 유출 될 경우 사용자의 주민번호만 노출 되는 것이 아니라 타인의 주민번호도 함께 노출 되는 것이므로 안전한 관리가 필요 할 것이다.

② 카드정보

[그림 9] 카드정보 검색결과

카드정보 검색결과는 카드정보가 탐지된 파일의 위치, 파일명, 카드번호, 그리고 카드를 발행한 카드회사로 나뉜다. 카드번호는 유효기간이나 CVC코드, 혹은 주민등록 번호와 조합되어 발견될 경우 사생활 침해를 넘어 금전적인 피해를 입을 수 있다.

③ 계좌정보



[그림 10] 계좌정보 검색결과

계좌정보 검색결과는 계좌정보가 탐지된 파일의 위치, 파일명, 은행, 계좌번호로 나뉜다. 계좌정보도 카드정보와 마찬가지로 다른 개인정보와 조합된 형태로 노출될 경우 금전적인 피해가 우려되며 [그림 10]에서 알 수 있듯이 타인의 계좌번호 다수가 저장된 파일일 경우 그 피해가 더 클 수 있다.

④전화번호

[그림 11] 전화번호 검색결과

전화번호 검색결과는 전화번호가 탐지된 파일의 위치, 파일명, 추정 전화번호로 나뉜다. 전화번호와 같은 경우 보이스피싱 및 스팸과 같은 사회공학적 위협에 취약하기 때문에 이러한 정보가 노출될 경우 사생활침해로 이어질 수 있다.

⑤ 사용자 지정정보

[그림 12] 사용자 지정정보 검색결과

사용자 지정정보 검색결과는 파일의 위치, 파일명, 사용자지정정보, 개인정보 분류, 개인정보 항목으로 나뉜다. [그림 12]의 결과는 [그림 13]에서 설정한 개인정보를 토대로 검색되었다.

[그림 13] 사용자 지정 개인정보

결과에서 알 수 있듯이 PIPM에서 자동으로 탐지하는 정보 외에 사용자가 탐지하고자 하는 개인정보를 개별적으로 찾아냄으로써 다양한 개인정보 항목의 탐지가 가능하다.

3.3.3 PIRM 구현

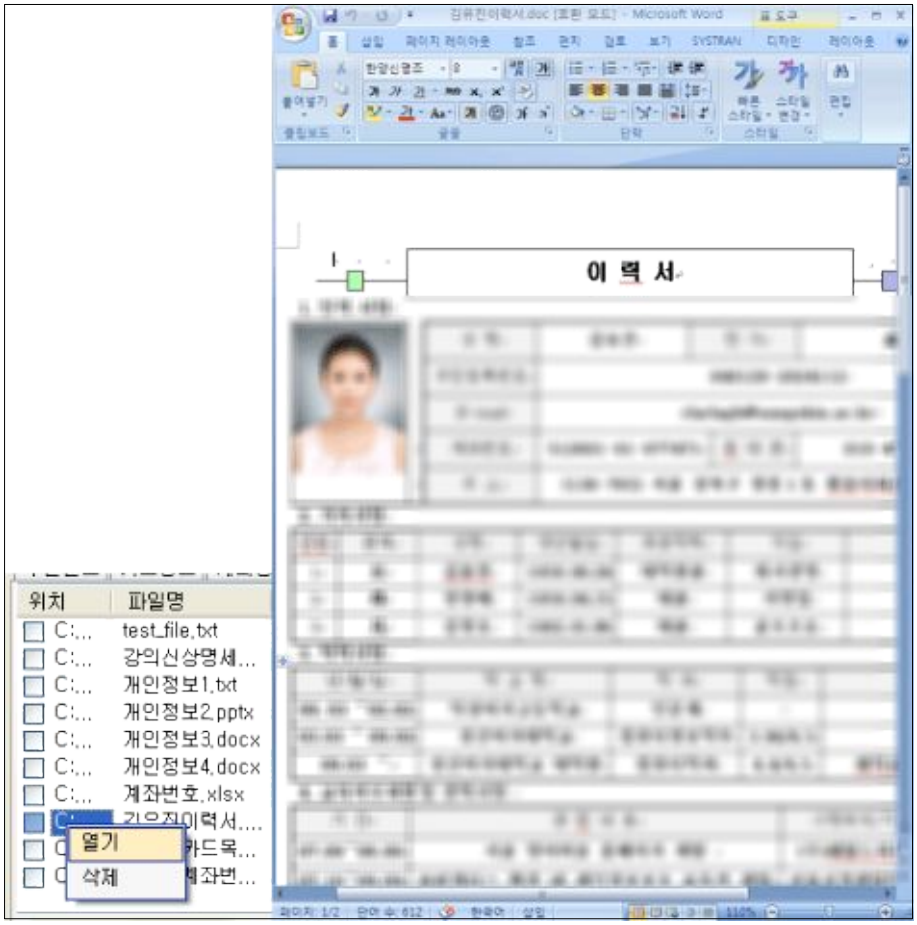
탐지된 개인정보 항목의 민감도에 따라 위험수준을 측정하는 기능으로, 위험 수준은 가장 낮은 레벨인 Level1에서 가장 높은 레벨인 Level8 로 나뉘며, 개인정보의 민감도가 P5등급인 항목의 개수가 하나일 경우는 위험수준을 고려하지 않아도 되는 N/A(Not Applicable) 로 측정된다.

위치	파일명	주민번호	카드정보	계좌정보	전화번호	사용자지정정보	합계	위험수준
C:₩Docu...	test_file.txt	0	0	0	0	1	1	N/A
C:₩Docu...	강의신상...	15	0	0	0	1	16	Level7
C:₩Docu...	개인정보...	0	0	0	1	0	1	Level2
C:₩Docu...	개인정보...	0	0	0	0	1	1	N/A
C:₩Docu...	개인정보...	0	0	0	0	1	1	N/A
C:₩Docu...	개인정보...	0	0	0	1	1	2	Level3
C:₩Docu...	계좌번호....	0	0	1	0	0	1	Level4
C:₩Docu...	김유진이...	1	0	1	1	2	5	Level8
C:₩Docu...	김유진카...	0	4	0	0	0	4	Level6
C:₩Docu...	학부생계...	0	0	8	4	1	13	Level5

[그림 14] 위험수준 측정

결과에서 볼수 있듯이 개인정보가 탐지된 각 파일의 위치, 파일명과 해당 파일에서 발견된 주민번호, 카드정보,계좌정보, 전화번호, 사용자지정정보의 개수와 합계가 표시되며 그에따른 위험 수준을 고지하고 있다.

사용자는 결과를 확인하기 위해 리스트에서 해당 파일을 바로 접근할 수 있다.

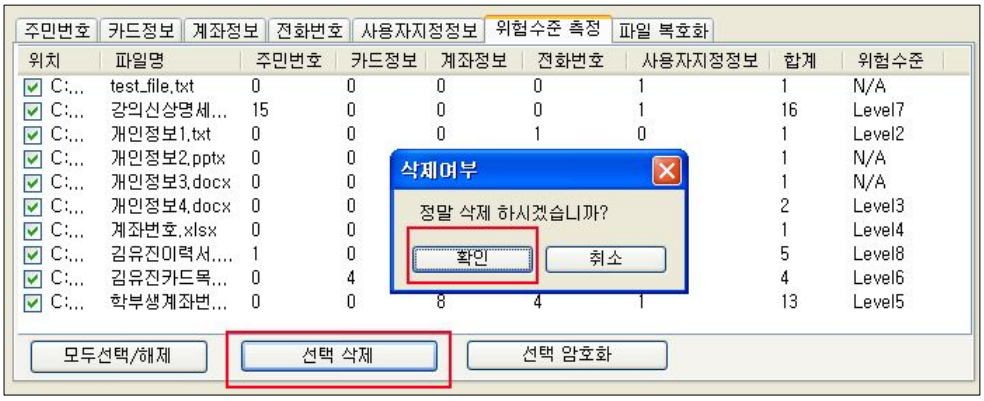


[그림 15] 해당파일 실행

3.3.4 CEM 구현

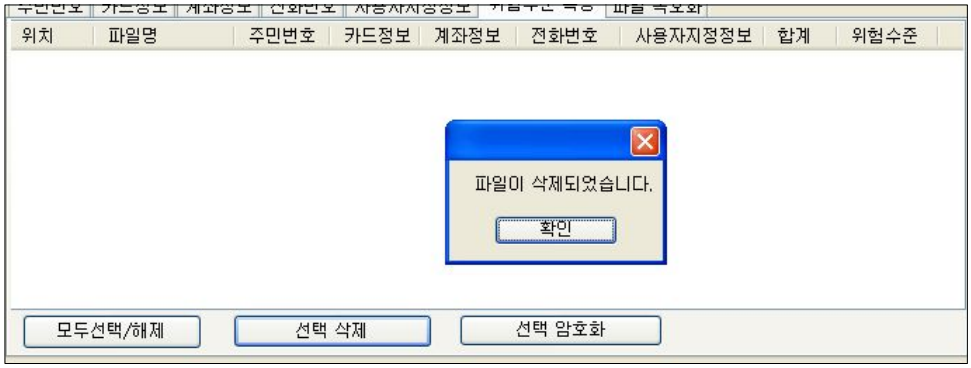
개인정보 노출위협이 있는 파일을 삭제하거나 암호화 하여 안전하게 처리하는 기능으로 사용자는 해당 파일을 암호화 시킨 경우, 복호화 하여 정상적으로 접근할 수 있다.

① 파일삭제



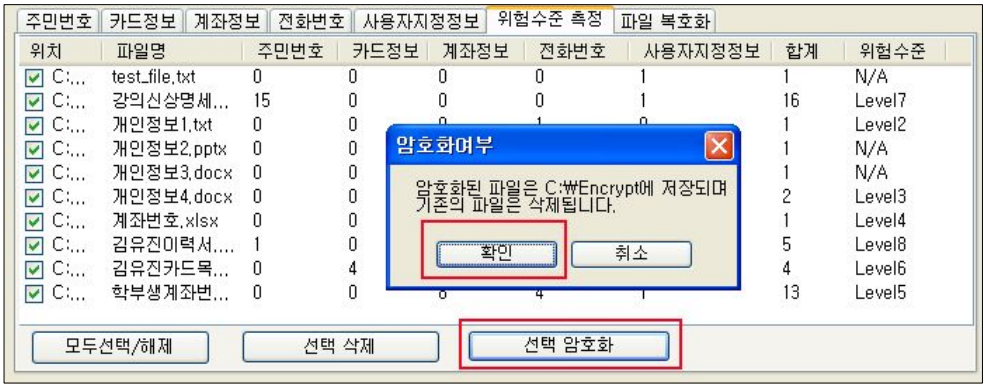
[그림 16] 파일 삭제

리스트에 나타난 결과를 선택적으로 삭제할 수 있으며 삭제 시 삭제 여부를 재확인한다. 파일이 삭제되면 리스트에서 해당항목들이 사라지며 사용자의 시스템에서도 파일이 삭제된다.



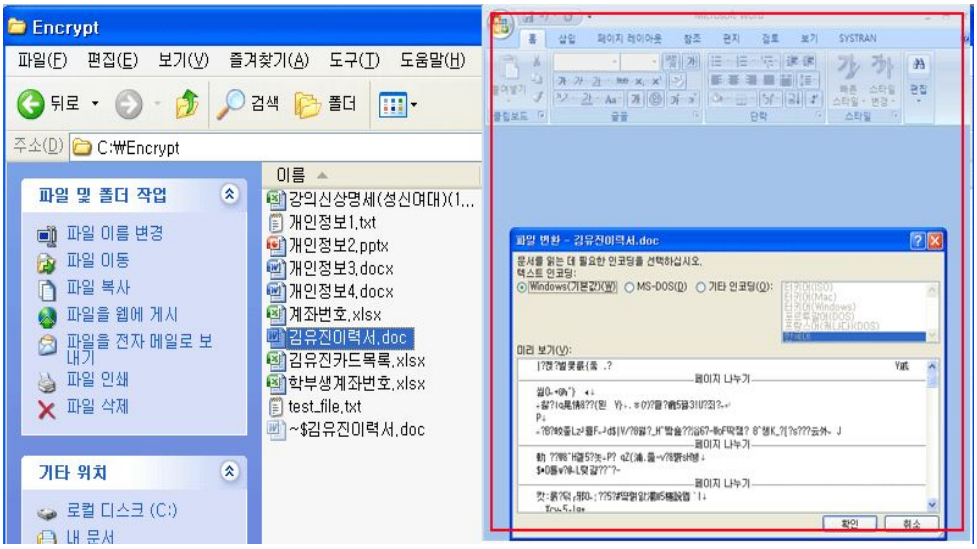
[그림 17] 파일 삭제 결과

② 파일 암호화



[그림 18] 파일 암호화

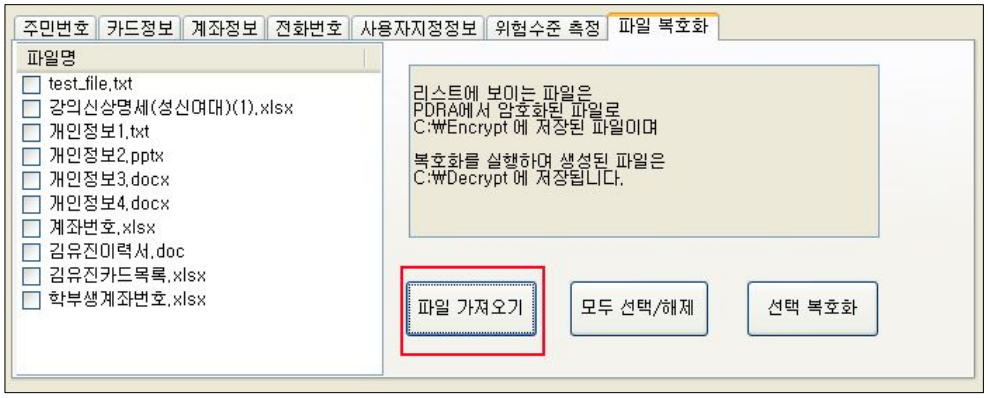
파일을 암호화 할 경우 암호화 된 파일은 C:\Encrypt 에 저장되며 기존의 파일은 삭제된다. 사용자 PC내에 Encrypt 디렉토리가 없는 경우 자동으로 디렉토리를 생성하게 된다.



[그림 19] 파일 암호화 결과

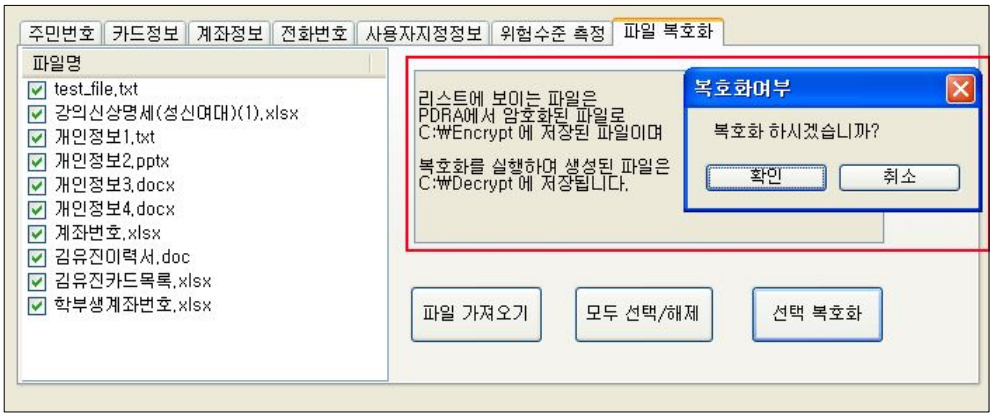
[그림 19]의 결과 화면에서 볼 수 있듯이 C:\Encrypt에 암호화된 파일이 저장되었으며 파일을 실행하면 해당파일의 정상적인 내용이 보이지 않음을 확인할 수 있다.

③ 파일 암호화



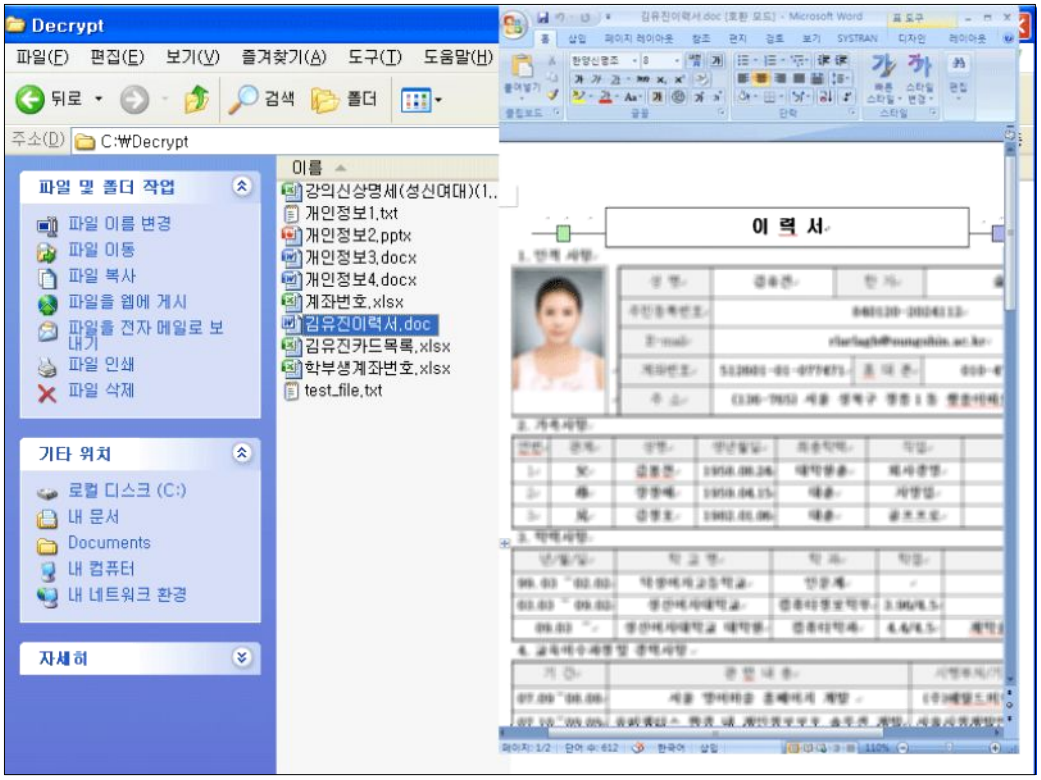
[그림 20] 암호화 파일 가져오기

암호화 된 파일은 복호화를 통해 다시 변환이 가능하다. 파일 가져오기를 클릭하면 C:\Encrypt에 저장된 암호화된 개인정보 파일이 리스트에 보인다.



[그림 21] 파일 복호화

복호화 되는 파일은 C:\Decrypt 에 저장되며 복호화를 실행하기 전에 복호화 여부에 대해 재확인한다.



[그림 22] 파일 복호화 결과

[그림 22]의 결과에서 알 수 있듯이 암호화 되었던 파일이 복호화를 통해 정상적으로 변환이 된 것을 확인 할 수 있다.

IV. 기대효과

본 연구는 사용자환경 내에서 개인정보 유·노출을 사전에 예방할 수 있는 솔루션을 제공함으로써, 개인PC를 사용하는 이용자 스스로 개인정보를 안전하게 관리하고 통제할 수 있는 환경을 제공할 것이라 사료된다.

현재 국내에는 개인PC를 대상으로 개인정보의 검색 및 삭제를 제공하는 솔루션들이 이슈가 되고 있지만, 대부분이 다양한 포맷을 대상으로 하고 있지 않으며, 단순히 개인정보가 저장 되어있는 파일의 삭제를 권고하고 있다.

그러나 개인정보를 탐지하여 제거하는 기술적 조치 뿐 아니라 사용자들이 스스로 개인정보에 대한 중요성을 인지하고, 올바른 판단에 따라 안전한 개인정보 관리를 할 수 있도록 하는 것이 우선이라고 할 수 있다.

이러한 요소를 고려하여 본 논문에서 제안한 모델은 사용자의 인지 없이 자동으로 기록된 개인정보 뿐 아니라 자의에 의해 저장된 개인정보 또한 다양한 파일 포맷 내에서 탐지함으로써, 사용자로 하여금 어떠한 개인정보가 유·노출 대상이 되고 있는지 알 수 있도록 한다.

탐지된 개인정보는 민감도와 조합 형태에 따라 위험수준을 측정해 사용자에게 고지되므로, 개인정보의 가치와 개인정보보호에 대한 인식을 제고하도록 할 수 있다.

또한 사용자의 선택에 따라 개인정보 유·노출 위협이 있는 파일을 삭제하거나 암호화함으로써, 해킹 및 불법적인 접근을 통한 개인정보 유출을 원천적으로 차단할 수 있도록 하였다.

그러므로 개인정보의 중요도와 위험수준에 따라 사용자가 개인정보를 효율적으로 관리하도록 하여 개인정보에 대한 중요성을 인지하고, 불법적인 개인정보 수집 및 유출사고를 예방할 수 있다.

V. 결론 및 향후연구

정보화 사회에서 개인정보의 활용가치는 매우 높아졌으며, 정보의 보호보다는 이용측면이 강조되면서 대량의 개인정보를 다루는 개인정보취급자가 고의 또는 과실로 개인정보를 오·남용하거나 관리적 보호대책의 부족, 기술적 보호대책의 적용미비로 개인정보가 유출되는 등 다양한 개인정보 침해 형태가 나타나기 시작하였다.

최근에는 일반 사용자의 개인PC를 대상으로 해킹 및 시스템에 설치된 봇을 이용하여 개인정보를 유출시키는 사례 또한 끊임없이 늘어나고 있다.

개인정보의 유출 사고 시 개인은 정신적, 금전적 피해 뿐 아니라 사회 활동에 까지 악영향을 미칠 수 있으므로 개인정보의 관리는 보다 안전하게 지켜져야 한다.

그러므로 정보 주체인 개인이 스스로 개인정보에 대한 가치를 인식하고 안전하게 관리 하여 정보화의 역기능을 최소화 하는 것이 중요하다.

이에 본 논문에서는 인터넷 및 PC 사용 시 기록될 수 있는 개인정보 관련 항목을 탐지하여 사용자에게 개인정보의 민감도에 따른 위험수준을 분석하는 PDRA(Privacy Detection and Risk Analysis) 모델을 제안하였다.

또한 개인정보의 유·노출 위협이 있는 파일을 사용자의 판단에 따라 효율적으로 관리 할 수 있도록 함으로써 악의적인 목적의 개인정보 수집을 원천적으로 차단할 수 있도록 하였다.

향후 PDRA 모델을 바탕으로 실시간 개인정보 탐지에 대한 방안을 연구할 것이며, 일반 사용자 외에 개인정보를 대량으로 수집·활용하는 기업과 공공기관의 환경에 적용 가능하도록 지능화, 경량화, 고성능화 등의 연구에 주력하고자 한다.

참고 문헌

[1] OECD, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data", 2001.

[2] The Brian Carrier, "File System Forensic Analysis", Addison Wesley, 2005.

[3] Amir M. Hormozi, "Cookies and Privacy", Information Systems Security, 2005.

[4] Microsoft Corporation, "Standard ECMA-376 Office Open XML File Formats", 2nd Edition, 2006.

[5] 한국인터넷진흥원, "APEC ECSG 개인정보보호 논의 동향", 2006.

[6] 한국인터넷진흥원, "개인정보의 안전한 수집, 저장, 및 관리, 이용, 제공, 파기를 위한 개인정보 관리모델 연구", 위너다임, 2006.

[7] Daniel Rentz, "Microsoft Compound Document FileFormat", OpenOffice.org's Documentation, 2007.

[8] 이동훈, "개인정보보호의 중요성과 보호기술", 한국소프트웨어산업협회, 2007.

[9] 한국인터넷진흥원, "인터넷 개인정보 노출방지 및 프라이버시보호방안 연구", 2007.

[10] 한국정보통신기술협회, "개인정보보호정책 설정 및 협상 규격", 2007.

[11] 남기효 외 2인, "개인정보보호기술의 최신 동향과 향후 전망", 한국정보보호학회, 2008.

[12] Microsoft Corporation, "Windows Compound BinaryFile Format Specification, Microsoft Corporation", 2008.

[13] 박보라 외 2인,"Microsoft Office 2007 파일에의 정보 은닉 및 탐지 방안", 한국정보보호학회, 2008.

[14] 박중환 외 3인, “기업내부 개인정보보호 시스템 개발”, 한국정보보호학회, 2008.
[15] 지식경제부, “유비쿼터스 환경에서의 정보보호 정책 방향”, 2008.
[16] 개인정보분쟁조정위원회, “2009 개인정보분쟁조정사례집”, 2009.
[17] 국회문화체육관광방송통신위원회, ″개인정보보호의 침해현황 및 개선방안″, 2009.
[18] 김진형 외 1인, “개인정보 데이터 접근 비정상행위 탐지 기법을 활용한 개인정보보호 기법 연구”, 정보과학회지, 2009.
[19] 방송통신위원회, “개인정보의 기술적·관리적 보호조치 기준 해설서”, 2009.
[20] 한국인터넷진흥원, “2009 정보보호 실태조사”, 2009.
[21] 한국인터넷진흥원, “개인정보의 기술적·관리적 보호조치 기준 해설서”, 2009.
[22] 한국인터넷진흥원, “”2009 정보시스템 해킹·바이러스 현황 및 대응, 2009.
[23] 한국인터넷진흥원, “개인정보의 기술적·관리적 보호조치 기준 해설서”, 2009.
[24] 한국인터넷진흥원, “개인정보 유출 공격 탐지 방안 연구”, 2009.
[25] 한승원 외 3인, “개인정보 저장 형태에 따른 유출 탐지 방안”, 한국정보보호학회, 2009.
[26] 홍승필, “개인정보보호 개론 : 사례연구 및 기술 중심으로”, 한티미디어, 2009.
[27] 국가정보원/방송통신위원회, “2010 국가정보보호백서”, 2010.
[28] 김영삼 외 2인, “u-IT 환경에서의 개인화서비스를 위한 개인정보보호방안 연구”, 전자통신동향분석 , 2010.
[29] 박남제, “스마트 그리드 환경에서의 개인정보 취약점 분석과 보호방안”, 한국정보기술학회, 2010.
[30] 박찬홍, “윈도우즈 기반의 다중 리눅스 서버 시스템 모니터링 및 제어 솔루션 구현”, 한국정보기술학회, 2010.
[31] 행정안전부, “개인정보보호법 제정”, 2010.

Abstract

Privacy Detection and Risk Analysis Model

Kim, Eugene
Dept. of Computer
Graduate School
Sungshin Women's University

In the information society, 'private information' has been developed as property right, the source of social group control or vast added value, and this has increased dysfunctions and negative social phenomena drastically.

Recently, a large amount of private information has been leaked numerous times, concern over personal information leakage is getting higher and higher, and groups or companies dealing with private information are establishing measures to cope with it by detecting the leak of personal information or inducing a system to prevent it.

However, in the environment of users, the actual subjects of information, such systems to cope with it have not been designed concretely. Besides, since it is difficult to recognize the level of risk based on the importance of private information or weaknesses, protective measures for it also have not been fulfilled sufficiently.

Thereupon, this thesis suggests a model to propose a systematic measure responding to the actual misuse and abuse of personal information based on

multi-level personal information protection policies in consideration of the significance of private information protection.

Also, it develops a mechanism to manage personal information safely according to the risk of personal information in the user's PC environment.

Lastly, this thesis shows the possibility through actual prototyping in order to raise the analysis on the risk of personal information and the utility of responsive researches.