情報保護學會論文誌
第 22 卷 第 5 號, 2012. 10

DLP방식의 문제점 극복을 위한 E-DRM 방식의 개인정보 보호 기술

최 종 욱,[1†] 이 용 진,[2‡] 박 주 미[1]
[1]상명대학교, [2]마크애니

E-DRM-based Privacy Protection Technology for Overcoming Technical Limitations of DLP-based Solutions

Jong-Uk Choi,[1†] Yong-Jin Lee,[2‡] Ju-Mi Park[1]
[1]SangMyoung University, [2]MarkAny

요 약

개인정보유출 방지 솔루션으로 사용되고 있는 DLP 방식 기술은 데스크톱 컴퓨터 중심의 온라인 작업환경에서는 고객 정보 유출 방지에 효과적이었으나 스마트 폰이 일반화되고 스마트 워크가 도입되면서 스마트폰에 의한 개인정보 유출과 APT공격에 취약하다는 점이 문제점으로 지적되고 있다.

본고에서는 DLP 기술이 갖고 있는 기본적인 문제점을 해결하기 위해 DLP기술의 내용 검색 기능을 정보 유출 방지 솔루션 E-DRM과 통합한 새로운 개인정보 보호 기술을 제안한다. 내용 검색기능을 활용하되 E-DRM이 갖고 있는 암호화 모듈과 접근제어 기능을 활용함으로 스마트워크 환경에서도 문서가 내부와 외부의 어느 사용자에게 전달되더라도 개인정보 유출을 효과적으로 막을 수 있을 것으로 기대되고 있다.

ABSTRACT

DLP technology has been effectively enforcing privacy protection policy in on-line computing environment. However, with wide spread use of smart devices and promotion of smart-works, it has been pointed out that DLP technology cannot effectively prevent privacy leakage at smart devices and is comparatively weak at APT attack.

In this paper, we suggests a hybrid approach, PPS, which integrates E-DRM system with DLP technology, taking advantages of both technologies. The technology basically uses encryption function and access control of E-DRM system, and thus it can effectively prevent leakage of privacy information of customers, even if the documents are in the hands of malicious third parties.

Keywords: PPS, DLP, E-DRM, contents filtering, APT, privacy

I. 서 론

개인정보 유출에 대한 사회적 관심이 높아짐에 따라 법적인 제제와 기술적인 조치들이 계속적으로 등장하고 있다. 지난해 제정된 개인정보 보호법 시행으로 국내에서는 개인정보 보호를 위한 행정적, 기술적 조치들이 정부 기관과 개인기업에 도입되고 있고, 미국

접수일(2012년 3월 9일), 수정일(1차: 2012년 4월 30일, 2차: 2012년 6월 20일, 3차 : 2012년 8월 20일), 게재확정일(2012년 9월 27일)
† 주저자, juchoi@markany.com
‡ 교신저자, yjlee@markany.com

에서도 올해 2월 백악관의 사생활 권리 장전 (consumer privacy Bill of Right)을 발표하였고[3], EU 역시 개인정보보호지침을 개정, 위반시 처벌을 강화하겠다고 약속하고 있다[5].

지난 해 제정된 우리나라의 개인정보 보호법에서는 개인정보의 수집에서부터 데이터의 가공, 관리와 사용, 사후 조치에 대한 규정을 다루고 있다. 개인정보 보호법에서 제시하고 있는 고객 정보 유출방지를 위한 기술적 조치로서는 키보드 해킹방지 솔루션, 백신 프로그램, 침입차단 시스템과 침입탐지시스템, 웹방화벽과 서버보안, 차등 접근권한 실시와 비밀번호 의무갱신, 개인정보 처리시스템에 관리자 접근 시 PKI 방식 사용, 개인정보 출력물에 대한 워터마킹(watermarking), P2P와 웹하드 (web hard) 등 비인가 프로그램의 접속 차단, 개인정보 처리시스템 로그파일의 생성과 관리 백업, 로그파일의 정기적 분석 및 결과보고, 개인정보 노출 사전차단 필터 링과 개인정보노출 자체모니터링 실시가 있다. 이중 기술적으로 중점을 두고 있는 것은 수집된 고객정보의 외부 유출에 대한 엄격한 규제이다. 2011년 9월에 개정, 공포된 개인정보보호법에서는 "개인정보가 분실·도난·유출·변조 또는 훼손되지 아니하도록" 내부 관리 계획을 수립하고 접속기록을 보관하도록 규정하고 있다 (제29조 안전조치 의미)[11].

현재 개인정보 유출 방지 기술로는 DLP(Data Loss Prevention)방식과 E-DRM(Enterprise Digital Right Management)방식이 주류를 이루고 있다[2][9]. DLP 방식은 사용자의 컴퓨터 혹은 네트워크를 통해서 전달되는 데이터를 체크하여 사전에 입력한 민감한 데이터의 패턴과 일치하거나 유사한 경우 이를 감지(detection)하여 조치(enforcement)를 취하는 기술이다. 데이터 유출방지를 위해서는 매체 제어, 패킷 통제, 프로그램 실행 통제, PC 통제 등의 기능을 수행한다. E-DRM 방식은 유통되는 모든 파일을 암호화하고 사용자 컴퓨터에서 사용자의 권한에 따라 복호화된 데이터의 사용을 제어하는 방식이다. 예를 들어, 동일한 파일을 전달받더라도 사용자 A는 파일을 읽고 출력을 할 수 있지만, 사용자 B는 읽기만 가능하고 자신의 컴퓨터에 저장하거나 출력을 할 수 없도록 기술적으로 강제하는 방식이다. DLP기술과 E-DRM 기술은 조직 내부의 중요한 정보를 유출하지 못하도록 하는 내부 보안 기술로 개발되었다. 즉, 기업의 경영 현황이나 시장 계획, 신제품 개발도면, 경쟁사 분석 등의 중요한 정보를 외부로 유출하지 못하도록 개발되었다. 그러나 최근 개인정보 보호가 사회적인 이슈로 떠오르자 기존의 내부 정보 유출방지 기술이 개인정보 유출 방지 솔루션으로 자리 잡고 있는 것이다.

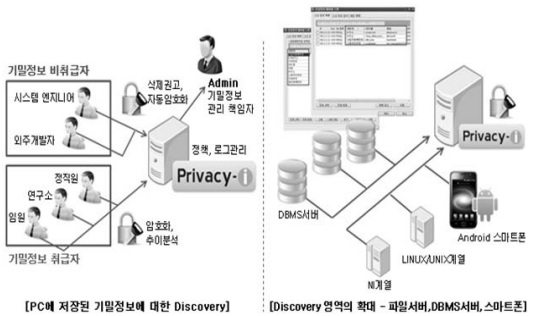
본고에서는 스마트워크 작업 환경에서 DLP방식의 유출방지 기술이 갖는 문제점을 지적하고, 이를 보완하기 위해 새로운 방식을 제안한다. 새로운 방식에서는 기존의 E-DRM기술을 바탕으로 DLP 기술에서 사용하는 내용검색 모듈을 통합하도록 하였다. DLP 기술과 E-DRM기술의 장점을 수용하는 이와 같은 하이브리드(Hybrid) 방식에서는 DLP기술이 가진 개인정보 검색 기능을 활용하고, E-DRM기술이 가진 암호화 기술과 접근제어 기술을 이용하여 시너지 효과를 갖도록 하였다. 이 방식을 사용함으로써 보편화되고 있는 스마트폰의 촬영 기능을 통한 정보 유출을 제한적으로 방지할 수 있고, APT방식의 침해에 근본적인 대응이 가능해진다. 내용검색을 제한적으로 실행함으로써 DLP기술에 대해 논란이 되어왔던 고객 서비스업체 직원들에 대한 '인권침해 가능성'도 줄일 수 있을 것으로 예상한다.

다음 장에서는 DLP방식에 대한 설명과 문제점에 대해 지적하고, 3장에서는 본고에서 제안하고 있는 새로운 방식에 대해 설명한다. 4장에서는 제안된 방식의 장점과 한계점, 결론을 논의하고 있다.

II. DLP방식의 개인정보 유출 방지

DLP방식에 의한 고객정보 유출 시스템은 내용검색에 기반을 두고 있다. 내용검색을 어디에서 하느냐는 방식을 두고, 네트워크 기반 DLP와 엔드포인트(end point) DLP로 구분된다. 네트워크 DLP는 중앙의 통제서버를 통해 외부로 전송되는 트래픽을 검사하는 방식이다. 또 엔드포인트 DLP는 노트북이나 PC 등 클라이언트 기기에 설치된 에이전트에서 사용자의 데이터 활동을 모니터링하여 정보 유출을 통제하는 방식이다[6].

DLP 기술 방식의 제안은 IBM의 기술로 시작되었다[15]. Zamora는 인터넷이 상용화, 보편화되지 않았던 1990년, 오고 가는 이메일을 분석하여 여러 가지 중요한 정보를 추출, 분류, 저장할 수 있다고 주장하였고 정보의 추출 방식으로 언어 의미 분석을 제안했다. 일반인들이 이메일을 사용하기 시작한 것이 1995년 이후라고 생각하면 당시로서는 파격적인 제안이라고 할 수 있다. Zamora는 외부와 연결된 네트워



〔그림 1〕 DLP 시스템 구조도
(출처 : www.somansa.com/dlp/dlo_3)

크를 통해서 오고 가는 메일, 팩스 등으로부터 송신자와 수신자, 그리고 중요한 정보를 의미론적 분석을 통해 항목별로 분류하여 추출하도록 방법을 제시했다. 즉, 상업용 메일의 송신자, 수신자, 날짜, 주소, 주제 등의 파라메타 정보를 구문분석으로 자동으로 추출하고 이를 데이터베이스(Database)에 저장하는 방법을 제시한 것이다. 파라메타 정보 추출(PIE: Parametric Information Extraction)이라고 불리는 이 기술적인 방식은 현재 DLP 방식에서 통상적으로 사용되고 있다.

Zamora의 제안은 DLP라는 이름으로 검색 대상은 사용자 컴퓨터에서 활성화된 모든 임시파일과 저장된 데이터〔23〕, 다양한 범주의 민감정보〔22〕, 사용자 컴퓨터에서 실행되고 있는 다양한 응용 시스템〔20〕, 통용되고 있는 web-mail〔19〕〔13〕로 그 범위가 점차 확대되고 있다.

고객 정보를 많이 다루게 되는 은행권, 통신사, 신용카드, 보험사, 호텔, 백화점 등의 서비스 업체와 공공기관, 정부 기관에서는 직원들의 컴퓨터나 네트워크에서 민감한 정보를 검색하고, 조치를 취하는 DLP 방식은 효과적일 수 있다. 모든 문서와 데이터가 평문으로 유통되기 때문에 암호화된 문서를 통제하는 E-DRM보다 편리하기 때문이다. 그러나 DLP 방식은 근본적으로 저장, 유통되는 정보가 "평문"(plain text)이어야 한다는 점과 사용자 컴퓨터 혹은 네트워크 시스템을 검색하여 유출을 방지한다는 점에서 다음과 같은 문제점이 제기되고 있다.

첫 번째, DLP 기술로서는 빠르게 확산되고 있는 스마트폰에 의한 정보 유출을 막을 수 없다. 스마트폰에는 카메라(camera), 녹음기능(recorder), 저장기능(SD)과 통신 기능(Wifi, 블루투스)이 탑재되어 있다. DLP방식을 사용하려면 저장되거나 유통되는 모든 정보가 평문이어야 하므로 권한을 갖지 않은 사용자가 촬영을 통해 정보를 유출시킬 때 막을 수 있는 방법이 없다. 물론 민감한 고객정보는 담당자의 컴퓨터에서 권한이 없는 타부서 사용자에게 네트워크, 혹은 저장장치를 통해 이동하게 될 때 네트워크와 기기에서 이를 막을 수 있겠지만, 민감한 정보에 대한 접근을 근본적으로 막기는 어렵다.

두 번째, DLP 기술은 평문을 통한 내용검색 기술에 의존하고 있어 외부 해킹에 취약하다. 지난해부터 보안 업계의 가장 심각한 우려 대상이 되고 있는 APT(Advanced Persistent Threat)공격〔21〕〔16〕〔18〕에 대해 기존의 DLP기술로서는 대응이 어렵다. APT는 SK커뮤니케이션즈나 넥슨 사례에서 보듯 초기 기획부터 대상을 명확히 설정, 집중적으로 공격하고, 주로 조직 구성원인 개인을 공격한다〔17〕. 일반적으로 조직 내 개인이 사용하는 PC는 주요 서버에 비해 상대적으로 관리가 취약한 편이기 때문이다. 예를 들면 조직 내부 PC를 감염시키기 위해 악성코드가 삽입된 이메일을 해당 PC사용자 업무와 관련된 내용으로 발송하거나 해당 사용자가 자주 이용하는 웹 사이트를 해킹해 악성코드를 배포하기도 한다〔10〕. APT에서는 이러한 PC를 공격 침입 경로로 이용해 효과적이고 은밀하게 조직 내부에 침투한다〔10〕〔8〕. 최근 SONY 케이스에서 보는 것처럼 APT 공격의 중요 목표가 고객정보 담당자가 되는 경우가 많아지고 있어 고객정보가 유출될 가능성이 높아지고 있다. 기본적으로 APT는 소셜 네트워크 검색과 프로파일(profile) 정보 수집에 의한 패스워드 확보 등, 다양한 경로의 정보 수집 및 철저한 사전 조사로 기획된 공격 대상을 향해 끈질기고 집요한 공격을 수행하기 때문에 방어가 힘들고 APT공격에 의해 일반 사용자 계정을 통해 네트워크가 열리게 되면 평문(plain text)으로 보관된 개인 정보는 유출되기 쉽다. DLP 기술에서는 송수신시 의심되는 유출을 차단함으로 개인정보를 보호할 수 있다고 주장하지 만, 최근 APT 공격은 암호화하거나 위장된 경우가 늘고 있어 이런 경우 DLP 기술로는 막기 어렵다.

세 번째, 기존의 DLP 기술로는 내부직원의 개인정보를 보호할 수 없다. 개인정보 보호(privacy protection)라는 관점에서 보면 은행이나 통신사, 호텔이나 백화점의 경우 '고객정보' 보호도 중요하지만, 한편으로 그 곳에서 일하고 있는 '직원 정보'보호도 중요하다. 미국이나 유럽에서 DLP시스템에 대한 회의적인 의견이 계속적으로 피력되는 이유는 직원들의 사생

활 침해 때문이다. 국내에서도 DLP에 의한 직원들의 사생활 정보 유출에 대한 연구가 이루어진 적이 있다[4]. DLP는 원래 기업 비밀 유출방지를 위해 개발된 것이고, 그 목적을 실현하기 위해 '내용검색'기술을 개발하게 된 것이다. 그러나 "이러한 과정에서 모든 패킷의 내용을 확인하는 행위 및 내부 정보 유출을 막기 위해 제시되고 있는 기술들은 정보 유출방지 시스템과 연결되어 있는 직원들의 프라이버시 침해의 위험성을 가지고 있다"는 것이 김진형, 김형종의 주장이다[4]. 그래서 검색 시 고객정보 외에 도 여러 가지 회사 정책에 위배되는 사생활 정보가 포함될 수 있다는 점이 논란이 되고 있다. 사업체에서 정책적으로 금지하고 있는 키워드를 포함시킬 경우, 예를 들면 性에 관련된 내용이나 직장 이전에 관련된 내용이 검색될 경우, 직원들의 개인 정보가 보호되기 힘들다는 점이다. 즉, 개인정보 보호가 '고객 정보' 뿐만 아니라 그 기관에서 일하고 있는 '직원의 정보'로 확대될 경우 DLP기술은 대단히 위험하다는 주장이 제기되고 있다[12][14].

이처럼 기존 DLP기술로는 기관 '내부 직원'의 개인정보(Privacy) 권리를 보호할 수 없고 외부 해킹에 의한 '정보 탈취'를 막을 수 없다는 점, 그리고 스마트폰의 기능 향상(카메라, 저장장치)을 이용한 정보탈취 시도에 효과적으로 대처할 수 없다는 문제점이 있다.

III. PPS: DLP기술과 E-DRM 기술의 하이브리드 접근 방식

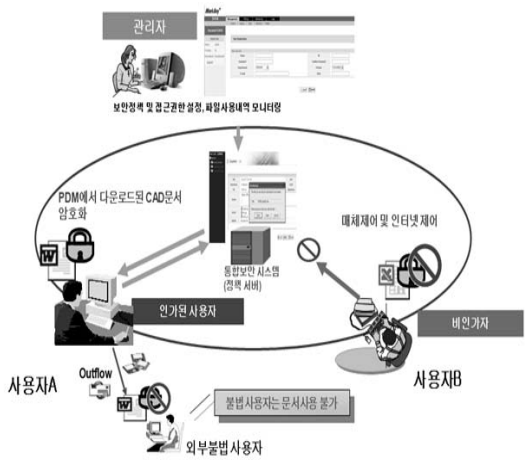
본고에서는 내용기반 검색 기술과 기업내부의 기밀 유출방지를 위해 개발된 E-DRM을 통합한 새로운 방식의 개인정보 보호 시스템을 제안하였다. 이 시스템(PPS: Privacy Protection Safer)에서는 내용기반 검색엔진과 E-DRM이 강점으로 갖고 있는 암호화와 클라이언트 제어(client-control)기능을 통합하여 운영하고 내용 검색 결과를 국지화(localization)함으로써 기업내부 직원의 개인 프라이버시를 최대한 보장하면서, 고객 정보 유출을 최대한 차단하도록 하고 있다.

3.1 E-DRM기술

컨텐츠 DRM (C-DRM: Content Digital Right Management) 기술은 90년대 후반 인터넷의 보급과 디지털 컨텐츠의 공급 시장이 열리면서 개발되었다. 그러나 Napster, e-Donkey등의 P2P사이트가 인기를 끌면서 미국이나 유럽에서는 점차 사라지게 되었다. 그러나 C-DRM 기술은 한국에서 내부정보 보안 기술(E-DRM)로 기술이 개발되고 진화하면서 세계적인 기술로 발전되어 왔다.

E-DRM(Enterprise DRM)이라고 불리는 내부정보보호 기술은 인터넷 환경에서 디지털 컨텐츠 (음악, 영화, 만화 등)를 안전하게 공급하기 위해 개발된 기존의 C-DRM기술과는 달리 조직 내부의 인트라넷 환경에서 주요 문서를 안전하게 유통시키기 위한 기술이다. E-DRM기술은 자동차, 조선, 반도체, 전자 기기 등의 제조업으로부터 금융과 통신, 정부기관으로 적용 대상이 확대되고 있으며 스마트폰의 보급으로 적용분야가 더욱 넓어지고 있다. E-DRM은 기본적으로 (1) 인터넷 환경이 아닌 내부 네트워크(Intranet)에서 문서의 작성, 유통, 폐기를 관리한다는 점에서, 그리고 기존의 컨텐츠 DRM이 주로 컨텐츠의 play기능만을 지원하는데 반해 (2) 문서의 작성, 읽기, 편집, 출력, 전송, 사용기간 등의 다양한 작업들을 제어 한다는 점에서 전통적인 컨텐츠 DRM(C-DRM)과 다르다고 할 수 있다.

E-DRM 기술은 서버에서 각 개인의 접근 권한을 설정하고 단말기에서 그 권한을 제어하는 구조를 갖추고 있다. E-DRM에서는 통상적으로 [그림 2]에서와 같이 서버에서 암호화된 문서가 사용자의 권한 정보(access rights)와 함께 사용자 컴퓨터로 전달되고, 사용자 컴퓨터에 설치된 에이전트(agent) 프로그램에 의해 사용자의 문서 사용이 통제, 관리된다. 이러한 운영은 암호화된 문서가 의도하지 않은 불법적인

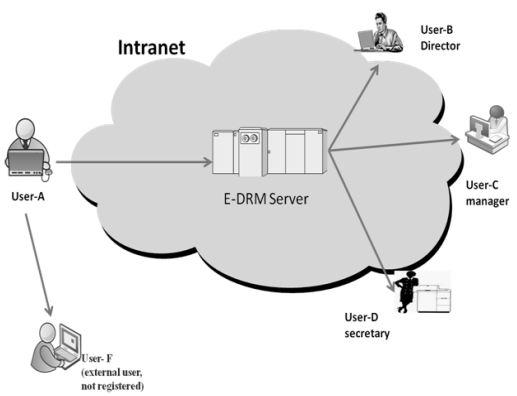


[그림 2] E-DRM의 접근권한 제어

사용자에게 전달되는 경우나 저장장치에 의한 복제와 인터넷에 의한 유포의 경우, 문서 기밀이 유출되지 않도록 하고 접근제어를 통해 사용자와 문서에 따라 각기 다른 접근과 사용을 제어할 수 있도록 한다.

권한 제어는 사용자(user)와 파일/문서(document), 그리고 제어 ('읽기', '쓰기', '편집', '출력', '전송', '기간 제어', '붙여 쓰기', '화면 캡처' 등)의 3차원 조합으로 이루어진다. 이에 따라 같은 사용자라 하더라도 부서에 따라서, 직위에 따라서, 혹은 동일부서에서도 수행하는 일(job)에 따라서 동일한 문서를 읽을 수도, 읽지 못할 수도 있게 된다. 예를 들어, 〔그림 3〕에서 내부 사용자 A가 보낸 문서를 사용자 B(이사급)는 읽거나 편집하거나, 출력을 할 수 있다. 출력의 경우 기업이나 기관의 로고가 문서의 뒤편에 출력되는데 이 로고 속에 워터마크 기술을 이용되고 출력자의 정보와 출력 일시, 출력기 정보가 숨겨져 있다. 동일한 문서를 전달받은 사용자 C(계장급)는 문서를 읽거나 편집할 수 있지만, 중요한 문서를 출력할 수 없고 사용자 D(비서)는 동일한 문서를 읽기는 할 수 있지만, 컴퓨터나 기기에 저장할 수 없다. 이처럼 E-DRM에서 권한 제어는 사용자, 문서, 작업(operation)의 3차원 조합 매트릭스(matrix)에 의해 제어되고 문서의 성격에 따라서 권한을 다르게 줄 수 있다. 예를 들어, 회계 관련 문서의 경우 회계 관련 부서나 임원진이 아니면 '읽기'가 불가능하게 할 수도 있고, 제품 기술에 관한 CAD파일인 경우, 회계 부서에서는 읽거나 출력을 하지 못하도록 통제할 수 있다.

E-DRM 시스템의 가장 중요한 기능은 자동적 암호화와 권한제어 기술이다. 문서의 등급에 따라서 보안 분류가 되면서 일정 수준의 보안 문서에 대해서는



〔그림 3〕 E-DRM에서의 사용자별 권한 제어

자동으로 암호화되어 사용자에게 전달되기 때문에 외부에 유출되더라도 불법적인 사용자는 이 파일을 열거나 사용할 수 없게 된다. 고객정보를 다루는 부서의 문서는 생성과 동시에 모두 암호화되기 때문에 외부의 제 3자에게 전달되더라도 사용이 불가하고, 내부 사용자에게 전달되는 경우에는 사용자 직위와 소속부서, 맡은 업무에 따라 '읽기' '저장' '편집' '출력' '전송' 등의 권한이 통제된다. 권한 통제가 문서의 종류와 성격, 사용자의 직위와 부서, 맡은 업무들을 고려하여 결정되면 사용자 PC 혹은 기기에서 수행되도록 한다.

3.2 PPS: E-DRM 기반의 내용분석 기능을 가진 개인정보 보호 시스템

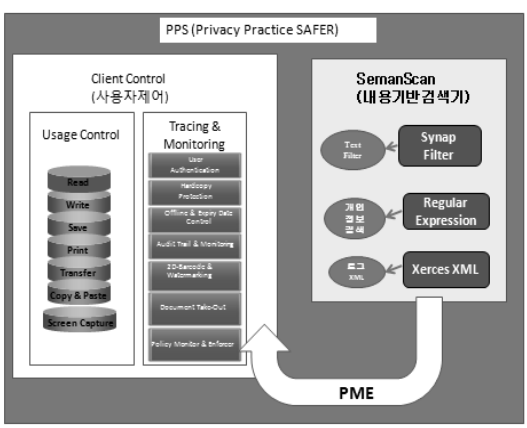
본고에서 제안하고 있는 새로운 방식의 개인정보 보호 기술은 내용분석 기술과 E-DRM기술을 통합한 Hybrid형 개인정보보호 시스템인 PPS(Privacy Protection SAFER)이다. 전체적으로 보면 PPS는 서버 시스템과 클라이언트 시스템으로 이루어진다. 서버 시스템은 문서를 유통하는 경우 개인정보 모듈이 작동되며, 클라이언트 시스템은 새로운 문서의 생성이나 Wifi, Bluetooth, web을 통한 문서 전송이나 USB, CD 등을 통한 문서 저장시 개인정보 모듈이 작동된다.

PPS 서버는 각 기관의 보안 정책에 따라 전달되는 문서나 파일의 사용자 접근권한을 결정하고 부여하는 역할을 한다. 각 기관의 문서나 정보는 문서관리 시스템(EDMS: Electronic Document Management Systems), 전자결제 시스템(Goupware, BPM), 전사자원관리 시스템(ERP: Enterprise Resource Planning), 제품 설계 시스템(PLM: Product Life Management), 포털 시스템, 혹은 이메일 시스템을 통해 사내 혹은 협력업체들과 유통을 위해 사용되는데 PPS는 이러한 사내유통시스템 (Enterprise Communication)에 통합(integration)하여 개인정보에 관련된 문서들이 외부로 유출되지 않도록 보안 정책을 실행하는 한편, 업무의 효율성과 편리성을 유지하는데 목표를 두고 있다. 따라서 PPS 서버는 각 기관의 보안 정책에 따라 사내 유통시스템을 통해 다운로드 받거나 이메일을 통해 첨부된 문서를 수신하는 경우 내용검색을 통해 문서의 개인정보 등급을 결정하고, 사용자의 지위, 부서, 맡은 업무를 고려하여 문서의 사용권한을 결정한다. 기관의 보안 정책과 수신자의 정보에 의해 결정된 접근 권한(access right)정

보를 파일 헤더에 붙이고 암호화한 뒤 다운로드 요청자(download requester), 혹은 이메일 수신자에게 파일을 전송하게 된다. 기존의 E-DRM에서의 서버와 동일한 역할을 수행한다. 클라이언트 모듈은 각 파일이나 문서의 헤더에 삽입되어 있는 사용자의 접근권한(access right)정보를 바탕으로 사용자의 활동을 제어한다.

파일이 사용자 컴퓨터에 도착하는 경우, 실제적인 권한제어는 사용자의 기기(컴퓨터 혹은 모바일 기기)에 설치된 PPS Client 모듈에 의해 수행된다. Client module은 〔그림 4〕와 같이 사용자의 활동(read, write, edit, print, keep, transfer)을 통제하고 사용자 활동을 모니터링 하는 Client Control 모듈과 내용기반의 정보를 검색하는 SemanScan 모듈, 그리고 파일의 속성을 조정할 수 있는 PME (Policy Monitoring & Enforcement) 모듈로 이루어져 있다.

PPS Client 모듈은 사용자가 응용 프로그램, 예를 들면 Office 문서나 CAD문서를 클릭하는 순간 자동으로 작동을 개시하지만, 사용하려고 하는 문서가 (1) 사내의 유통 시스템 (Enterprise Communications Systems: EDMS, ERP, groupware, portal, e-mail system 등)을 통해 다운로드 받거나 첨부된 문서인 경우, (2) 새로 작성하는 문서인 경우, (3) Web으로 접근하는 경우, (4) 저장매체를 통해 저장하려는 경우, (5) 주기적인 검사인 경우에 따라 다른 과정을 거치도록 하고 있다. 첫 번째 경우를 제외하고는 모두가 사용자의 client 모듈에서 이루어진다는 점이 중요하다.



〔그림 4〕 PPS Client모듈

PPS Client 모듈은 사내의 유통시스템을 통해 다운로드 받거나 이메일 첨부를 통해 받은 문서를 사용하는 경우, 서버에서 사용자정보와 보안 정책에 의해 '접근 권한' (access right)의 접근 제어를 담당한다. PPS 서버로부터 전달받거나 다운로드 받은 이러한 문서의 접근제어는 서버에서 규정한 (specified) 권한정보에 따라 행해지기 때문에 내용 검색 모듈(Seman Scan)을 구동시킬 필요가 없다. 즉, 서버에서 접근권한을 '읽기 3회'(read), '편집(edit)불가', 혹은 '출력(print) 1회'라고 규정한 경우, PPS Client모듈에서 사용(usage) 제어기능을 작동시켜 사용자의 활동을 감시하고 제어한다. '편집 불가' 혹은 '출력 불가'라는 정보에도 불구하고 네트워크 프린터, 가상 프린터 등을 통해서 출력을 시도하는 경우, 활동정보는 서버로 전달되고, 로그(log)파일에 남게 된다. '출력 불가'에도 불구하고 출력을 시도하는 경우, 클라이언트의 E-DRM agent 모듈이 응용 프로그램(application program)을 제어하거나 utility 프로그램(출력관련)을 제어하여 출력을 막는다. 이는 E-DRM 시스템의 고유한 기술로 E-DRM 기술이 구현되기 어려운 이유이기도 하다. 접근권한이 '출력 가능'인 경우, 문서 출력 시 사용자의 정보 (이름, 사번)과 출력시간, 출력 포트 등의 정보를 기관의 로고(logo)에 은닉하여 문서의 배경으로 출력하고, 문제가 생기는 경우 출력한 사람을 추적할 수 있도록 한다.

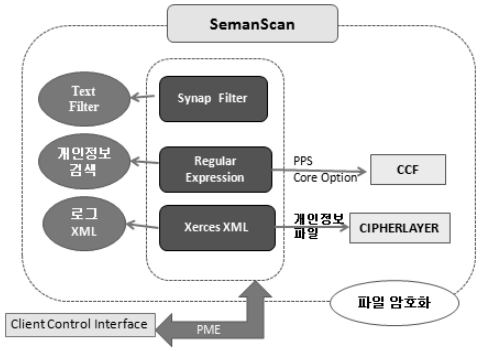
사용자 기기(device: 사용자 PC 혹은 Mobile 기기)를 통해 새로 생성되는 문서와 사용자 기기에서 Web이나 저장매체를 통해 외부로 반출되는 경우, 혹은 주기적인 검색의 경우 내용검색과정(SemanScan)을 반드시 거치도록 하고 있다. 사용자 기기에서 새로 문서가 생성되는 경우, '저장'을 하는 순간 내용 검색 엔진 (SemanScan)이 작동되면서 개인정보 등급이 결정되고, 이를 바탕으로 새로운 문서의 '개인정보 보안조치'가 결정된다. 기존의 내려 받은 문서, 혹은 이메일 첨부된 문서를 편집하여 저장하는 경우에도 이 과정을 따르도록 하고 있다. 만약 내용검색의 결과 개인정보의 수준이 '유의'(meaningful)한 경우, 자동으로 암호화되면서 '개인 정보수준' 필드(field)에 1-5등급으로 정보 수준이 표시된다. Web을 통해 외부의 시스템과 접속하는 경우와 저장매체를 통해 외부 유출의 경우에도 기존 문서의 '접근 권한' 체크와 내용검색, 그리고 개인정보 보안 조치를 수행한다. 주기적인 배치(batch) 검사의 경우에도 내용검색을 통해 개

인정보 수준을 결정하고, 이를 통해 보안조치를 취하게 된다.

개인정보 보안 조치는 (1) 격리 저장, (2) 삭제, (3) 암호화로서 개인정보가 어느 수준 이상인 경우 기본적으로 모두 암호화 등이 있다. 사용자가 로그인(log-in)할 때 서버로부터 내려 받은 사용자 정보 (사용자의 이름과 사번, 지위와 부서, 직책, 소속 그룹 등의 정보)를 참조하고 내용검색에서 얻어진 '개인정보 수준'에 의해 보안조치를 결정하는데, 만약 내용 검색을 통해 개인정보 수준에 대해 어느 정도 이상이지만 강제적인 삭제를 하기에 확신이 서지 않는 경우, 특정 보안 폴더로 파일을 강제적으로 옮겨서 사용자가 임의로 접근할 수 없도록 하고 있다.

3.3 내용 분석기(SemanScan)

개인정보 추출 모듈인 SemanScan은 Synap Document Filter를 통해 해당 파일의 텍스트를 추출하고, Regular Expression 모듈에서 조건식(CCF)에 정의된 탐색 패턴을 통해 해당 텍스트의 개인정보 포함 여부를 확인한다. 이를 통해 나온 결과를 Xerces XML을 통해 로그 정보를 DSX_PPHistory.xml 파일로 저장하며 해당 개인정보 결과 값을 필요로 하는 PPS Client 모듈에 전달하거나 CCF의 옵션값을 통해 파일 암호화를 수행하는 기능을 제공한다. 개인정보 추출 모듈 SemanScan은 텍스트 추출 라이브러리인 Synap Document Filter를 통해 해당 파일의 텍스트를 추출하는 과정을 거친다. 다음으로 Boost 라이브러리의 Regular Expression으로 CCF에 정의된 정규식 패턴을 통해 해당 텍스트의 개인정보 포함 여부를 확인한다.



〔그림 5〕 SemanScan (내용분석기)의 구조

이를 통해 나온 결과를 Xerces XML을 통해 로그 정보를 DSX_PPHistory.xml 파일로 저장하며 PME를 통해 해당 개인정보 결과 값을 필요로 하는 다른 모듈에 제공해 줄 수 있으며 CCF의 PPS_Core 옵션값을 통해 파일 암호화를 수행하는 기능을 제공한다.

- Synap Document Filter: 파일 및 메모리 상태에서 해당 데이터 내부의 텍스트를 추출하는 기능을 한다. 인터페이스를 통해 파일 데이터에서 실질적인 텍스트 추출을 요청하고 파일 형식을 모르는 임의의 파일을 필터링 하는 기능을 수행하며 파일 경로 및 텍스트 추출에 필요한 SN3BUF 형식의 파라메터를 요구한다.
- Regular Expression (Boost Library): 정규 표현식 패턴을 통해 문자열내의 특정 패턴이 있는지를 검색해 주는 라이브러리이다. Synap Filter의 텍스트 추출 기능을 통해 얻어온 텍스트 데이터를 정규 표현식으로 정의한 패턴을 가지고 해당 텍스트에서의 특정 문자들을 검출해주는 기능을 제공한다. SemanScan모듈 사용 시 개인정보 관련 옵션 값이 있으며 해당 값들은 CCF에 저장 되어 있다. 따라서 모듈 SemanScan을 사용하기 위해서는 CCF가 적용되어져 있어야 하며 필요한 파일 및 옵션정보는 〔표 1〕과 같다.
- Xerces XML: 개인정보검색 로그 정보를 XML 파일로 저장하기 위해 사용되며 DSX_PPHistory.xml파일로 저장한다.

이와 같이 3개의 sub module 외에도 파일에서

〔표 1〕 Default CCF 적용 및 사용 시 필요한 파일

ccf_default.xml	Default CCF 파일로써 기본적인 PPS_Core 관련 옵션들을 가지고 있다.
Command-제품등록.reg	Default CCF 사용의 앞서 "ccf_default.xml"이 저장되고 참조하는 폴더를 레지스트리에 등록해 두어야하며 "Command-제품등록.reg" 파일이 레지스트리 등록하는 기능을 한다.
CCF_TOOL.exe	설정된 CCF값을 수정 및 확인할 수 있게 하는 툴
DSC_Core.dll	암복호화 기능을 담당하는 DLL
DSC_Config_PPS.dll	CCF 설정 값을 읽어오는 DLL
DSC_SystemID.dll	머신키 관련 기능을 수행하는 DLL
xerces_ma.dll	Xerces 를 사용하기 위한 DLL

텍스트 추출 후 개인정보 검색(Personal Info File Search), 메모리에서 텍스트 추출 후 개인정보 검색(Personal Info Memory Search), 문자열에서 개인정보 검색(Personal Info Str Search), 개인정보 추출 로그 저장(Send or Save)기능을 제공한다.

3.4 PME 모듈

PPS에서는 Client 모듈에 Policy Monitor & Enforcer(PME) 서브 모듈이 존재한다. PME모듈은 내용 검색기인 SemanScan에서 전달된 정보를 바탕으로 해당 문서에 '개인정보 등급'(Privacy Grade)을 파일 저장 시 삽입하여 저장하게 된다. 따라서 이 문서가 서버를 통해 다른 사용자에게 전달될 경우, 서버에서는 특별히 Privacy Grade(1-5 등급)를 체크하여 등급이 사용자의 프로파일에 맞추어 파일 접근권한을 조정하게 된다.

예를 들어 Privacy Grade = 5인 경우, 해당 부서 이외의 사용자에게는 문서를 읽을 수 없게 하고, 출력과 재전송을 허용하지 않도록 할 수 있다.

IV. 내용 검색 기반의 PPS 시스템의 장점

내용검색 기능을 추가한 E-DRM기반의 개인정보 보호 시스템인 PPS의 경우 기존 DLP기술이 가질 수 없는 장점들을 갖고 있다.

첫 번째, PPS는 스마트기기에 의한 촬영을 모두 막을 수는 없지만, 암호화와 접근제어에 의해 사용자를 제한할 수 있어, 스마트기기에 의한 피해를 줄일 수 있다. PPS에서는 사내 유통시스템을 통해 전달, 유통되는 파일이 암호화되어 있고, PPS Client모듈 환경 하에서 생성되는 파일 중에서 개인정보 수준(privacy level)이 유의한 모든 파일이 암호화되어 있기 때문에 카메라 촬영이 가능한 대상을 현격하게 줄일 수 있다. 예를 들어, 의료기관에서 개인정보가 포함된 문서의 경우, 문서를 복호화하여 화면을 통해 읽어볼 수 있는 사용자를 '원무과'로 국한하거나 '진료의료진'으로 한정시킬 수 있다. 이와 반대로 DLP의 경우, 평문으로 문서가 전달, 유통되기 때문에 문서를 읽어볼 수 있는 사용자수가 PPS보다 많아질 수밖에 없다. 충분하게 통제되지 않는 경우 의료기관의 조달부서와 회계부서에서도 개인정보 파일을 열어볼 수 있는 구조이다. 민감하고 중요한 개인정보 문서의 경우, PPS에서는 1000명이 근무하고 있는 의료기관이라도 사용자를 5명이내의 제한 할 수 있지만, DLP기술의 경우 '평문'이라는 전제조건 때문에 별도의 보안수단이 없는 한 접근 가능한 사용자의 수가 많아질 수밖에 없다.

두 번째, 외부 해킹의 경우 PPS 시스템은 DLP보다 높은 보안성을 보여주고 있다. 사용자 컴퓨터에 전달된 모든 문서는 기본적으로 암호화되어 있기 때문이다. 최근에 문제가 되고 있는 APT공격에 의해 민감한 문서가 제3자에게 전달되거나 유출되더라도 암호화되어 있기 때문에 개인정보 관련 문서가 오용되거나 네트워크를 통해 확산된 가능성은 거의 없다고 볼 수

〔표 2〕 SemanScan 모듈 관련 CCF 설정정보

설정 데이터(Key)	설명
PPSAFER_SAVELOG	개인정보검색을 수행한 후 해당 결과를 DSX_PPHistory.xml로 저장해두거나 DB로 결과 전송을 수행하는 옵션정보이다. 0일 경우 결과저장 및 DB 전송을 수행하지 않으며 1일 경우 수행한다.
PPSAFER_ENCFILEFILTER	개인정보검색을 수행할 시 암호화 파일 경우 개인정보를 검색할지를 결정한다. 0일 경우 수행하지 않으며 1일 경우 수행한다.
PPSAFER_NOTI	개인정보검색 모듈인 PPS_Core.dll을 사용하는 부분에서 개인정보검색 결과를 Message Box나 기타 결과 창을 통해 보여줄 지의 여부를 결정한다. 0일 경우 수행하지 않으며 1일 경우 수행한다.
PPSAFER_FILEDELETE	개인정보검색을 수행한 후 개인정보가 포함된 문서의 경우 해당 파일을 삭제할지를 결정한다. 0일 경우 수행하지 않으며 1일 경우 수행한다.
PPSAFER_REENCRYPT	암호화 파일의 경우 개인정보검색을 수행한 후 해당 암호화 파일을 새로 정의된 정책으로 인해 재암호화를 수행할지를 결정한다. 0일 경우 수행하지 않으며 1일 경우 수행한다.
PPSAFER_ENCRYPT	개인정보검색을 수행한 후 개인정보가 포함된 문서일 시 PPS_Core.dll 모듈에서 제공하는 암호화 기능을 사용할지를 결정한다. 0일 경우 수행하지 않으며 1일 경우 수행한다.
PPSAFER_COMPFILEFILTER	개인정보검색을 수행 시 압축파일의 경우 압축 파일 내 파일들의 개인정보 포함여부를 검색할지를 결정한다. 0일 경우 수행하지 않으며 1일 경우 수행한다.

있다. 더구나 PPS에서 사용하고 있는 암호화는 키값을 사용자 기기로부터 가져다 쓰기 때문에 예외적인 경우를 제외하고는 소속기관이 아닌 다른 기관의 사용자 기기에서 열어 볼 수가 없다. 전술한 바와 같이 사용자 기기에서 생성되는 문서와 Web을 통해 전송하는 문서, 그리고 저장매체(CD, USB, Hard Disk 등)를 통해 저장되는 경우 모두 암호화되기 때문에 외부 유출시 다른 사용자 기기에서 정상적으로 사용이 불가하다. 따라서 APT공격에 의해 개인정보 관련 문서가 유출되더라도 오용되는 경우를 막을 수 있다.

세 번째로, PPS에서는 내용검색 기능(Seman-Scan)이 사용자 컴퓨터에 국한되어 있기 때문에 직원의 개인 정보 보호문제를 야기하지 않는다. 지금도 미국이나 유럽에서 현재 DLP기술에 대해 '직원들의 사생활 침해 가능성'에 대해 끊임없이 의문을 제기하고 있는 바, PPS에서는 '사생활 침해'의 소지를 없앴다. 사용자의 문서에서 고객정보를 발견했다고 하더라도, 고객정보의 민감도 등급을 결정하여 사용자 컴퓨터에서 작동하고 있는 PPS Client에게 고객정보 등급정보를 전달하고, PME 모듈에 의해 파일의 등급이 달라질 뿐이다. 이처럼 고객정보가 발견되는 경우, 그 결과를 바탕으로 문서의 개인정보 등급만을 바꾸도록 하는 것은 이 문서가 다음 사용자에 게 전달되었을 때 그 등급에 따라 사용자의 접근권한이 달라지도록 하기 위한 것이다. 현재 DLP기술이 개인정보를 언제든지 검사하고, 개인정보 검사 시 파일을 삭제하거나 송수신을 중단하는 것과는 다른 조치이다. PPS는 고객정보를 다루는 기관이나 기업 내부 사용자들의 편의성 측면에서 많이 개선된 시스템이라고 할 수 있다.

V. 결 론

본고에서는 현재 심각한 사회적 이슈로 떠오르고 있는 개인정보유출을 방지하기 위한 새로운 시스템으로서 PPS시스템을 제안한다. 본고에서 제안하고 있는 PPS시스템은 기존의 E-DRM 기술에 DLP에서 사용하고 있는 내용 검색 기능을 통합한 솔루션이다. PPS시스템은 모든 유통되는 문서와 사용자 기기에서 생성되는 민감한 정보를 암호화하고, 사용자의 직위와 직책, 부서에 따라 접근을 제한할 수 있기 때문에 외부해킹이나 APT공격으로부터 비교적 안전하다. 유통되는 모든 중요한 문서를 암호화하기 때문에 문서를 복호화하여 읽을 수 있는 사용자 수가 제한적이라는 점에서 DLP기술에 비해 스마트기기의 카메라에 의한 유출에서 비교적 안전한 기술이라고 할 수 있다. 그리고 개인정보 검색을 위한 내용 검색을 사용자 컴퓨터로 제한하고 있고, 해당 문서의 '읽기' 권한을 가진 사용자만을 대상으로 내용검색을 하고 있어 병원, 은행, 신용카드사, 보험사, 통신사, 호텔, 정부기관 등과 같이 민감한 고객 정보를 다루는 기관의 직원 사생활 침해의 논란에서 자유롭다.

본 연구에서 개발된 기술은 향후 개인정보뿐 아니라 정부기관이나 사기업의 정보유출과 같은 좀 더 광범위한 영역으로 확대 적용이 가능하다는 점에서 중요한 의의를 가지고 있다. 즉, 본고에서 제시하고 있는 기술은 내용 검색 기능을 확장하면 고객의 이름이나 주소, 혹은 전화번호와 같은 개인정보 관련뿐만 아니라 정부기관이나 사기업의 중요한 문서의 유출을 막는 용도로도 사용할 수 있다. 그러나 앞으로 풀어야할 과제도 많다. 스마트 기기의 보편화와 스마트워크의 확대는 기존의 기술로는 대처하기 어려운 도전으로 등장하고 있다. 예를 들어 스마트 기기의 카메라, 녹음기능은 개인정보 보호 분야의 심각하고 중요한 위협이 되고 있다. 본고에서 주장하고 있는 것처럼 PPS방식에서는 암호화와 접근제어로 개인정보 관련 문서를 열람할 수 있는 사용자의 수가 제한적이기는 하지만, 현재의 내용 검색 기능이 문자와 숫자에만 국한되고 있다. 개인 생활에 대한 사진 촬영이나 녹음의 경우, 현재의 내용검색 기능만으로는 관련 파일의 차단이 불가능하다.

현재 개발되고 있는 MDM(Mobile Device Management)기술에서는 스마트기기에서의 사진 촬영, 음성 녹음, Wifi 송수신, Bluetooth기능을 무력화하기 위한 시도가 이루어지고 있지만, 급속히 번지고 있는 BYOD(Bring Your Own Devices)환경에서는 효과적인 대처가 어렵다. BYOD환경에서 개인정보 유출을 효과적으로 할 수 있는 새로운 기술의 개발이 절실히 필요하다 하겠다. 향후 모든 직원이 사무실이 아닌 자택이나 사무실 외부에서 업무를 진행하는 스마트워크 환경에서는 BYOD환경보다 더 광범위한 개인정보 유출이 가능하기 때문에 새롭고 강력한 방식의 정보 유출 기술이 개발되어야 할 것이다.

참고문헌

〔1〕 강태욱, "개인정보보호와 기업의 책임," 디지털 타임즈, May 29, 2012.
〔2〕 금융보안연구원, "개인정보 보호 기술 동향 보고서,"

금보원, 금융보안연구원 동향보고서, 2011-09, Dec. 2011.
〔3〕 김명환, "미국, 온라인 개인정보 유출차단 나선다," 매경 뉴스, Feb. 24, 2012.
〔4〕 김진형, 김형종, "정보유출방지와 프라이버시 침해에 대한 고찰," 정보보호학회, 정보보호학회지 v.21, no.5, pp.45-49, 2011.
〔5〕 김희연, "세계는 지금 '개인정보보호 열풍,'" ZDNet, Jan. 27, 2012.
〔6〕 이민형, "DLP 솔루션을 도입해야하는 이유는…," Ddaily, Ddaily.co.kr, Feb. 3, 2012.
〔7〕 이종현, "SK컴즈 100만원씩 보상땐 35조…해킹 후폭풍," 조선일보, April 26, 2012.
〔8〕 장윤정, "APT공격, 99.999% 당한다: 다계층 보안 및 전 직원 보안교육 등 전방위 대비 필요," 보안닷컴, Nov. 2011.
〔9〕 정영철, "개인정보 유출방지를 위한 기술적 보호모델에 관한 연구," 석사학위논문, 성균관대학교, 2011.
〔10〕 조남용, "APT공격, 지능적 대응이 답이다," 보안닷컴, Jan. 2012.
〔11〕 행정안전부, 개인정보보호법, 법률 제10465호, 2011.3.29. 제정, 시행 2011.9.30., 2011.
〔12〕 Bardin J., "Data Loss Prevention . What the DLP Companies Don't Tell You?," Sept. 2009.
〔13〕 Barzilai Z., Shmulyian S., Feldman S.,"Enterprise Privacy Manager," US Patent 7,225,460, IBM, May 2007.
〔14〕 Curtin-Mestre K., Room S., Yngve S., "Privacy Concern with Adopting DLP Technology," RSA Conference Europe 2009, Oct. 2009.
〔15〕 Elena M. Zamora, Computer Method for Automatic Extraction of Commonly Specified Information from Business Correspondence, US Patent 4,965,763, IBM, Oct. 1990.
〔16〕 Frankie Li, "A Detailed Analysis of an Advanced Persistent Threat Malware," Whitepaper of SAN Institute, Oct. 2011.
〔17〕 Greg Hogland, "Advanced Persistent Threat: What means to your enterprise?," Presentation of ISSA Conference, Feb. 2010.
〔18〕 Mathew J. Schwartz, "Advanced Persistent Threats Get More Respect," Information Week, Feb. 9. 2012
〔19〕 Ra1an B., Dalal C. D., Kabra N., "Method and Apparatus for Detecting Web-based Electronic Mail in Network Traffic," US Patent 7,996,406 B1, Symantec, Aug. 2011.
〔20〕 Russell Stringham, Eduardo Suarez, "Systems and Methods for Processing and Managing Object-Related Data for use by a Plurality of Applications," US Patent 2011/0113466 A1, Symantec, May 2011.
〔21〕 Sam Cury, Bret Hartman, David P. Hunter, David Martin, Dennis R. Morean, Alina Oprea, Uri Rivner, Dana Elizabeth Wolf, "Mobilizing Intelligent Security Operations for Advanced Persistent Threat," RSA security brief, RSA, Feb. 2011.
〔22〕 Wootton B., Dandliker R., Tsibulya A., Brucening O., Kessler D., "Methods and Systems for Normalizating Data Loss Prevention Categorization Information," US Patent 8,060,596 B1, Symantec, Nov. 2011.
〔23〕 Zoppas M., Hermann J., O'Raghallaigh C., Bothwell E., Fontana A., "Method and Apparatus for Detecting Policy Violations in a Data Repository Having an Arbitrary Data Schema," US Patent No. 7,996,373, Aug. 2011.

〈著者紹介〉

최 종 욱 (Jong-Uk Choi) 정회원
1982년: 아주대학교 산업공학과 졸업
1988년: University of South Carolina, MIS 박사
1991년~현재: 상명대학교 소프트웨어대학 컴퓨터 과학부 교수
〈관심분야〉 DRM, Watermarking, 네트워크 해킹 및 시스템 해킹

이 용 진 (Yong-Jin Lee) 정회원
2002년: 청주대학교 전자정보통신반도체 공학부 졸업
2009년~현재: 마크애니 전략사업본부 선행개발실 팀장으로 근무
〈관심분야〉 개인정보보호, DRM, Watermarking, 네트워크 해킹 및 시스템 해킹

박 주 미 (Ju-Mi Park) 정회원
1996년 2월: 상명대학교 정보과학과 졸업
2000년 8월: 상명대학교 일반대학원 정보과학과 석사
2009년 8월: 상명대학교 일반대학원 정보과학과 박사 수료
〈관심분야〉 정보보호, 인공지능