개인정보 오남용 추출조건 정교화를 위한 시스템 설계

이기호, 정영철, 이야리(한국보건사회연구원)

차 례

1. 서론
2. 관련연구
3. 개인정보 오남용 추출조건 정교화를 위한 시스템 설계
4. 시험적용 및 평가
5. 결론

1. 서론

1.1 이론적 배경

최근 개인정보는 단순히 한 개인을 식별하기 위한 정보만이 아닌 공공 및 민간기관에서 다양한 목적으로 폭넓게 활용되고 있다. 민간기관에서는 고객의 특성에 맞춘 타깃마케팅 측면에서 기업의 마케팅 및 경제적 이익 등을 목적으로 활용되고 있으며, 공공기관에서는 행정 효율성을 높이고, 맞춤형 서비스 제공 및 소비자 선호도 조사 등 정보 사회의 발전에 따른 부가가치 창출의 핵심 자원으로 이용되고 있다[1][2].

그러나 개인정보의 사용이 증가함에 따라 개인정보 대규모 유출사고 및 프라이버시 침해사고가 빈번하게 발생하면서 사회적으로 개인정보보호 관리에 대한 경각심이 높아지고 있다[3]. 개인정보보호위원회의 연차보고서에 의하면 2011년도 한 해에만 포털사이트의 3,500만 건의 개인정보 유출사고를 비롯하여, 공식적으로 보고된 개인정보 유출사고로만 약 7,000만 명 이상의 개인정보가 유출된 것으로 보고되었으며, 국민의 개인정보보호에 대한 관심과 피해구제 요구도 크게 증가하여 한국인터넷진흥원의 개인정보침해신고센터와 분쟁조정위원회에 접수된 개인정보 관련 피해구제 신청건수도 2010년 54,832건에서 2011년 122,215건으로 전년대비 122.8% 증가한 것으로 나타났다[3][4].

기업에서의 개인정보 유출사고는 집단 손해배상손실, 기업 이미지 손실, 매출손실, 주가 하락 손실, 기업 경쟁력 손실 등 5중고를 겪을 수 있으며, 나아가 지속경영가능성에 까지 영향을 미칠 수 있는 기업 비즈니스 위험요인으로 그 중요성이 크게 부각되고 있다[5]. 공공기관에서도 개인정보 유출 사고 발생여지가 다분히 증가하고 있는데, 국민의 신뢰를 바탕으로 국가 정책을 추진하는 공공기관의 개인정보 유출사고는 심각한 사회적 문제를 유발 수 있다[1]. 이에 2011년 제정된「개인정보보호법」에서는 수집, 저장·관리, 이용·제공, 파기 등 개인정보의 처리단계(생명주기, life cycle)에 따라 공공과 민간부분의 모든 개인정보처리자가 개인정보보호 조치를 의무적으로 이행하도록 제도화하고 있다.

특히 업무의 특성상 국민의 중요한 개인정보를 다수 보유 및 취급하고 있는 보건복지부에서는 본부, 소속 및 산하기관의 내부직원에 의한 개인정보 오남용을 상시 모니터링하고 대응할 수 있는 개인정보통합관제시스템을 2009년부터 구축하여 운영하고 있다.

하지만 지금까지는 개인정보 침해를 방지하기 위해 주로 네트워크를 통한 외부 공격자에 의한 침입을 차단하는 방화벽이나 침입차단시스템 구축 등의 연구에만 주로 집중되어 왔었으며 개인정보의 처리단계 중 이용 및 제공 단계에서 발생할 수 있는 내부 직원에 의한 개인정보 오남용에 대한 연구는 상대적으로 미흡한 실정이다.

이에 본 연구에서는 개인정보의 수집, 저장 및 관리, 이용 및 제공, 파기의 개인정보 처리단계 중 개인정보의 이용 및 제공 단계에서 내부직원에 의한 개인정보 오남용 행위를 효과적으로 탐지할 수 있는 추출조건의 정교화와 이를 활용한 시스템의 설계 방안을 제시하고자 한다. 본 연구의 구성은 다음과 같다. 1장은 논문의 개요에 대해서 간략히 소개하고 2장에서는 개인정보 및 개인정보 오남용에 대한 개념, 개인정보 탐지방법 관련 연구에 대하여 설명한다. 3장은 개인정보 오남용 추출조건 정교화를 위한 시스템 설계에 대하여 설명하며, 4장에서 개인정보 오남용 추출조건 정교화에 따른 시험운영 및 추출조건의

적정성 평가를 설명하였다. 5장은 결론으로 구성된다.

2. 관련연구

2.1 개인정보 오남용

‘개인정보보호법’에 따르면 개인정보란 “살아 있는 개인에 관한 정보로서 성명, 주민등록번호, 영상 등을 통하여 개인을 알아볼 수 있는 정보와 그 자체로는 직접 알아보기 힘들더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보”로 규정하고 있다[6].

개인정보의 개념 및 범위는 사회 환경의 변화와 기술발전에 따라 지속적으로 확대되고 있으며, 정보통신기술 발달로 보호되어야 할 개인정보 유형도 다양화되고 있는 추세이다[6]. 정보생명주기 관점에서 개인정보의 처리단계는 정보의 수집, 정보의 저장 및 관리, 정보의 이용 및 제공, 정보의 폐기 등 4단계로 구분할 수 있으며, 개인정보에 대한 침해는 모든 처리단계에서 발생할 수 있다[7].

개인정보 오남용이란 개인정보를 본래의 목적이나 범위를 벗어나 함부로 사용하거나 잘못 사용하는 행위로 정의할 수 있다. 이는 OECD 프라이버시 가이드라인의 8가지 원칙 중 이용제한의 원칙(정보 대상자가 동의하거나 법에서 허락한 경우를 제외하고는 목적 범위 내에서 적법하게 처리, 목적 이외 활용 금지)을 벗어난 행위로도 정의할 수 있다[8]. 따라서 개인정보 오남용이라 함은 정보의 이용 및 제공단계에 발생할 수 있는 개인정보 침해이다.

2.2 개인정보 오남용 탐지방법

개인정보 오남용 탐지란 개인정보취급자 등의 접속기록을 분석하여 개인정보 유출 및 오남용 의심로그를 찾아내는 것이다.

접속기록이란 개인정보취급자가 개인정보처리시스템에 접속한 기록으로, 개인정보취급자 등이 개인정보처리시스템에 접속하여 수행한 업무 내역에 대하여 식별자, 접속일시, 접속자를 알 수 있는 정보, 수행업무 등 접속한 사실을 전자적으로 기록한 것이라 규정하고 있다[9]. 또한 접속기록은 ‘개인정보보호법’ 및 행정안전부의 ‘안전성 확보조치 기준 고시’에 따라 개인정보 침해사고 발생시 대응할 수 있도록 개인정보처리시스템에 대한 개인정보취급자 등의 개인정보의 입·출력 및 수정사항, 파일별·담당자별 데이터접근내역 등 접속기록을 생성하여 안전하게 관리토록 규정하고 있다[6]. 따라서 접속기록을 통해 개인정보취급자 등 내부직원의 개인정보 유출 및 오남용 여부를 탐지할 수 있으며, 보건복지부의 경우 개인정보 유출 및 오남용을 방지를 위해 접속기록의 수집, 분석, 소명, 판정 등의 일련의 과정을 별도로 규정하고 있다[10]. 이러한 접속기록을 분석하는 방법으로 최근 침입탐지 방법을 이용한 연구들이 수행되고 있다[11][12][13].

침입탐지는 탐지를 분석하는 방법에 따라 크게 오용탐지(misuse detection)방법과 비정상행위 탐지(anomaly detection)방법 두 가지로 구분할 수 있다[14][15]. 오용탐지방법은 이미 알려져 있는 침입에 대한 패턴을 정의하고 패턴의 일치여부를 통해 해당하는 침입 여부를 탐지하는 방법으로 알려진 형태의 공격 순서를 시그너처(signature)화하여 이러한 방식의 순서 혹은 특징을 따르는 상황을 공격이라고 판단한다. 일반적으로 알려진 공격에 대한 탐지능력만을 가지게 되므로 실제 침입이 아닌 경우 침입이라고 판정하는 과탐지 오류(false positive error)가 비교적 적은데 반하여, 공격정보를 계속 수집해야 하며 알려진 공격 형태를 벗어나면 탐지할 수 없다는 한계가 있다[16]. 대표적인 오용탐지방법으로는 전문가의 지식을 규칙으로 표현하여 조건의 일치여부에 따라 처리하는 전문가시스템(expert systems)[17], 침입 패턴의 처음부터 종료까지 상태 전이그래프를 사용하여 표현한 후 이를 통해 침입여부를 판단하는 상태 전이분석(state transition analysis)[18], 기존의 침입 행위에 대한 시나리오를 생성하여 이를 이용하여 침입을 탐지하는 모델 기반 기법(model based technique)[19] 등이 있다.

비정상행위 탐지방법은 정상행위 프로파일링을 추출조건으로 정의하여 정상행위 패턴에서 벗어나는 작업을 비정상행위로 탐지하는 방법으로 축적된 정상적인 데이터를 수집하여 정상적인 형태의 사용에 대한 프로파일링을 완성한 후 이와 다른 형태의 패턴을 가진 데이터를 탐지하는 방식이다. 기존에 알려지지 않은 침입을 탐지할 수 있고 실제의 침입을 침입이 아니라고 판정하는 미탐지 오류(false negative error)를 줄일 가능성이 높은 방법이다[16]. 대표적인 비정상행위 탐지방법으로는 통계적 확률을 이용하여 비정상상태를 탐지하는 통계기반(statistics-based)방법[20]과 데이터들에 대한 발생 빈

도를 기반으로 각 데이터 간의 연관관계를 도출하는 연관규칙(association rule)방법[21], 대용량의 사건들이 기록되어 있는 데이터베이스에서 유사 작업군을 탐색하는 클러스터링(clustering)방법[22], 분류규칙(classification rule) 등과 같은 데이터마이닝 탐지방법이 있다.

개인정보 유출 및 오남용방지를 위해 침입탐지방법을 이용한 연구들을 살펴보면, 이미 알려진 위험요인이나 조직에서 관리하고 있는 개인정보를 취급하는 패턴을 분석하여 이상 징후를 탐지할 수 있도록 접근여부나 접근빈도에 따른 임계치를 설정하는 방법과 개인정보 접근횟수를 정규분포와 확률밀도함수 등을 이용하여 통계적 분석을 통해 비정상행위를 탐지하는 방법 에 대한 연구가 수행되고 있다[11][12][13].

3. 개인정보 오남용 추출조건 정교화를 위한 시스템 설계

3.1 개인정보 오남용 추출조건 개요

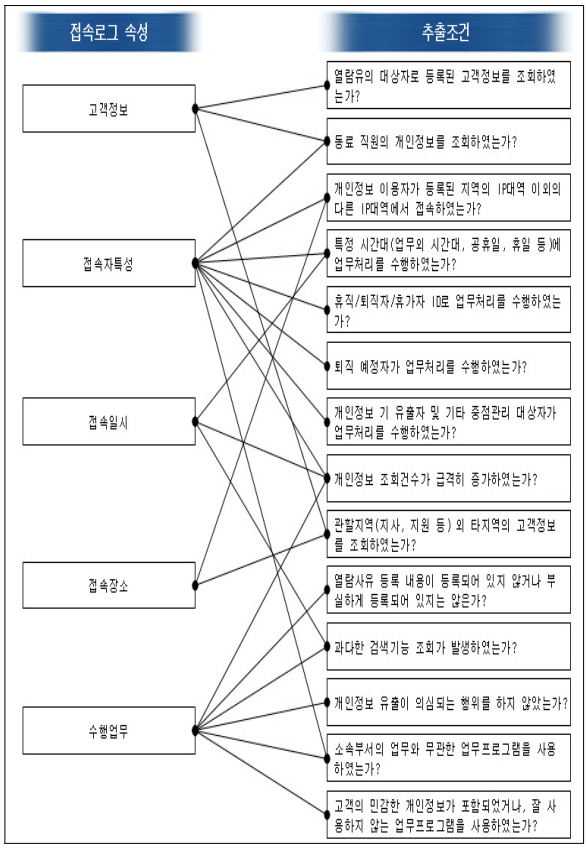
‘개개인정보 오남용 추출조건은 접속기록의 속성인 식별자, 접속일시, 접속자의 특성, 접속 후 수행업무, 접속 및 개인정보 조회빈도 등을 분석하여 접속자의 불법적인 접근 또는 비정상적인 업무처리 등 개인정보 오남용이 의심되는 접속기록을 추출하는 SQL 스크립트이다. 따라서 개인정보 오남용 추출조건의 성능은 분석할 접속기록의 속성 및 형태에 따라 크게 좌우된다.

이러한 접속기록은 기관별・시스템별로 기록되어지는 속성이나 형식이 상이할 수 있는데, 본 연구에서는 보건복지부의 경우를 사례로 연구하였다. 보건복지부는 본부, 소속 및 산하기관의 주요 개인정보처리시스템에 대한 개인정보 오남용 의심로그를 보다 명확하게 탐지할 수 있도록 업무처리 과정을 육하원칙에 따라 기록할 수 있도록 접속기록의 속성과 형식을 표준화 하였다. 만약 보건복지부의 경우와 같이 다수의 기관이나 시스템으로부터 접속기록을 수집하여 분석하고자 한다면, 접속기록의 표준화가 선행되어지는 것이 바람직 할 것이다. 다음의 [표 1]에서는 표준화된 접속기록의 유형별 로그항목들을 나타낸다.

표 1. 표준화된 접속기록 예시

구분	로그항목
접속일시(When)	업무처리일시
접속자특성(Who)	사번, 사원명, 사용자ID
접속장소(Where)	사용자IP, 소속 부서코드, 소속 부서명, 소속지사코드, 소속지사명, 소속 지역본부코드, 소속 지역본부명 등
고객정보(What)	고객주민번호, 고객명, 조회자료건수, 조회조건, 조회순번, 암호화된 피보험자 주민번호 등
수행업무(Why)	버튼처리구분, 화면ID, 화면명
수행업무(How)	프로그램ID, 프로그램명, 요청URL 등
고객정보 (개인정보유형)	가입자 소속코드, 가입자 소속명, 가입자 지역본부코드, 가입자 지역본부명, 관할지부여부, 사업장 소속코드, 사업장 소속명, 사업장 지역본부코드, 사업장 지역본부명 등
기타	감사번호, 권한등급, 현재상태, 일련번호, 시스템코드, 기관코드 등

[표 1]의 내용을 분석하여 접속기록과 개인정보 오남용 추출조건과의 관계로써 개념적으로 재분류해보면접속기록 속성과 추출조건과의 대응관계로 정리되어 [그림 1]과 같다.



▶▶ 그림 1. 접속기록과 추출조건과의 관계

개인정보 오남용 의심로그를 탐지하기 위한 추출조건 개발 프로세스는 [그림 2]와 같이 현황분석, 개발계획 수립, 추출조건 개발, 검증, 운영의 5단계이다. 우선 현황분

석 단계에서는 관련 법률이나 지침, 언론 등에서 보고된 사례분석, 전문 인력의 경험 등을 토대로 개인정보 오남용 의심사례를 분석하고 추출조건의 개발 필요성을 검토한다. 개발계획 수립단계에서는 대상기관의 로그분석과 업무분석을 통해 대상기관의 특성을 파악하게 된다. 추출기관 개발 단계에서는 개인정보 오남용 의심 패턴이나 정상행위에 프로파일링을 작성하여 추출조건과 제외조건을 개발한다. 검증단계에서는 자체 테스트 및 내부 전문 인력 검증, 그리고 대상기관에 대한 검증이 이루어진다. 마지막으로 운영단계에서는 추출조건을 시스템에 적용하여 운영하고 운영과정 중 발생한 문제점 및 개선사항을 반영하는 추출조건 정교화, 타 기관으로 확대 적용을 위한 검토가 이루어진다.

▶▶ 그림 2. 개인정보 오남용 추출조건 개발 프로세스

3.2 보건복지 개인정보통합관제시스템의 개인정보 오남용 추출조건 분석

본 연구에서는 개인정보보호를 위해 기존의 개인정보 오남용 추출조건을 적용한 보건복지 개인정보통합관제시스템을 분석하였다. 보건복지부 및 소속·산하기관의 업무 특성상 국민의 중요한 개인정보를 다수 취급하고 있어, 개인정보 오남용 여부를 상시 모니터링하고 대응할 수 있는 감시체계 필요성에 따라 2009년 개인정보통합관제시스템을 구축하여 운영 중이며, 2012년 상반기 기준으로 국민건강보험공단, 국민연금공단, 건강보험심사평가원, 대한적십자사 등 주요 7개 기관 35개 개인정보처리시스템에 대한 모니터링을 상시적으로 수행하고 있다.

보건복지부 개인정보통합관제시스템에 수집되는 접속기록은 연간 약 37억 건으로 일일 약 1,000만 건 이상, 근무일 기준으로는 일일 약 1,500만 건의 접속기록이 수집되어 추출조건에 의해 개인정보 오남용 의심로그가 추출되고 개인정보 오남용 여부를 분석하는 전문 인력에 의해 최종 검토 후 해당기관으로 소명요청이 이루어지게 된다.

기존의 개인정보 오남용 추출조건은 [표 2]와 같이 이미 알려진 개인정보 오남용 의심로그에 대한 패턴을 규칙화하여 접속기록 중 의심로그를 추출하는 오용탐지방법을 사용하고 있다. 그러나 일일 수집되는 1,000~1,500만 건의 방대한 접속기록을 고려한다면 정교화 되지 않은 기존 추출조건은 과도한 과탐지 오류를 발생시켜 전문 인력이 검토하여야 할 의심로그를 증가시킴으로써, 업무효율성을 크게 저하시키는 요인이 되고 있다.

표 2. 기존 주요 추출조건

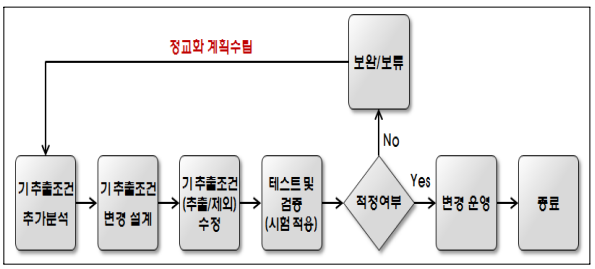
추출조건명	추출조건 정의	접속기록 속성				
		고객정보	접속자 특성	접속일시	접속장소	수행업무
열람유의 대상자 조회	열람유의 대상자로 등록된 고객정보 조회	○				
직원간 조회	동료 직원의 개인정보 조회	○	○			
접속IP 탐지	등록된 지역의 IP대역 이외의 다른 IP대역 접속		○		○	
특정 시간대 업무처리	특정 시간대(업무외 시간대, 공휴일, 휴일 등)에 업무처리		○	○		
휴직/퇴직자 업무처리	휴직/퇴직자 ID로 업무처리 수행		○			
보안 취약자	개인정보 기 유출자 및 기타 중점관리 대상자 업무처리 수행		○			
퇴직 예정자 업무처리	퇴직 예정자가 업무처리 수행		○			
개인별 임계치	개인별 개인정보 조회건수가 임계치 초과		○	○		○
관할지역 외 업무처리	관할지역(지사, 지원 등) 외 타지역 고객정보 조회	○			○	
열람사유 유형분석	열람사유 미등록 및 부실하게 등록하여 의심되는 경우					○
과다한 검색기능 조회	검색기능 조회가 임계치를 초과하여 과다 조회			○		○
개인정보 유출 의심행위	고객정보 조회 후 파일 변환 및 다운로드 등 개인정보 유출이 의심되는 행위					○
타 업무프로그램 사용여부	소속부서의 업무와 무관한 업무프로그램 사용 여부		○			○
중점관리 프로그램 사용여부	고객의 민감정보를 포함하였거나 많이 사용하지 않는 프로그램 사용 여부					○

3.3 개인정보 오남용 추출조건 정교화 프로세스

개인정보 오남용 추출조건은 시스템에 적용된 후 운영과정 중에 정교화 과정이 지속적으로 수행되어야 하며 추출조건 정교화를 통해 과탐지된 개인정보 오남용 의심로그를 최소화하여 전문 인력의 업무효율성을 제고하고, 혹시 발생할 수 있는 미탐지 오류를 제거할 수 있어야 한다.

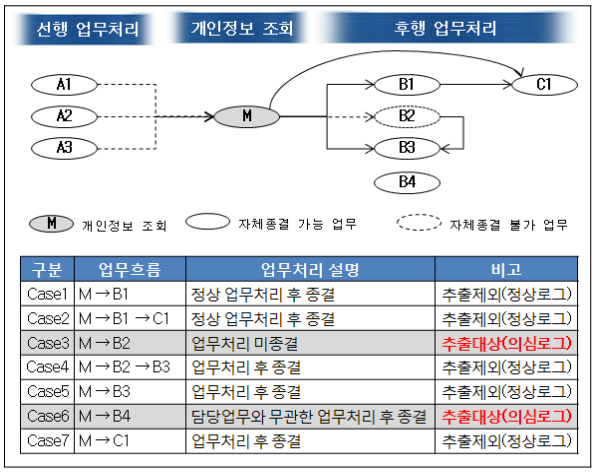
본 연구에서 설계한 시스템에서 적용한 개인정보 오남용 추출조건 정교화 프로세스는 [그림 3]에서 보는 바와 같이 소명판정결과와 대상기관의 업무조사서 분석을 토대로 정상적인 업무행위로 확인된 개인정보 오남용 의심

로그의 과탐지 발생 원인을 파악하고, 이를 추출조건 및 제외조건에 반영함으로써 개인정보 오남용 의심로그 추출에서 제외하기 위한 과정이다.



▶▶ 그림 3. 개인정보 오남용 추출조건 정교화 프로세스

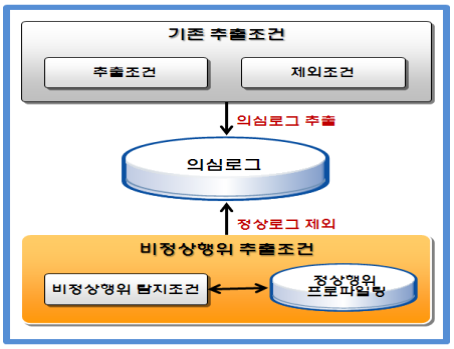
소명판정결과에 대한 심층 분석을 통해 기존 개인정보 오남용 추출조건에 의해 추출된 의심로그라 할지라도 [그림 4]와 같이 개인정보 조회 후에 정상적인 후행 업무처리가 이루어졌다면 결과적으로 정상적인 업무행위로 판단할 수 있으며, 이러한 로그들은 의심로그에서 제외할 필요성이 제기되었다.



구분	업무흐름	업무처리 설명	비고
Case1	M→B1	정상 업무처리 후 종결	추출제외(정상로그)
Case2	M→B1→C1	정상 업무처리 후 종결	추출제외(정상로그)
Case3	M→B2	업무처리 미종결	추출대상(의심로그)
Case4	M→B2→B3	업무처리 후 종결	추출제외(정상로그)
Case5	M→B3	업무처리 후 종결	추출제외(정상로그)
Case6	M→B4	담당업무와 무관한 업무처리 후 종결	추출대상(의심로그)
Case7	M→C1	업무처리 후 종결	추출제외(정상로그)

▶▶ 그림 4. 업무흐름에 따른 정상/비정상행위 분석

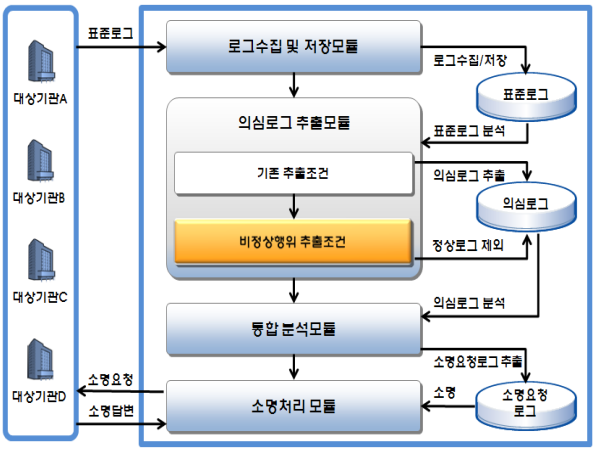
이에 본 연구에서는 기존에 적용되어진 개인정보 오남용 추출조건들의 정교화를 위해 비정상행위 탐지방법의 연관규칙을 도입하였다. 우선 대상기관에 대한 업무조사서를 통해 대상기관 업무담당자가 고객의 개인정보를 열람하는 프로그램을 목록화하여 정상적인 업무처리로 자체 종결 가능 여부를 파악하였고, 연관된 후속 업무처리가 필요한 경우 정상적인 후행 업무처리흐름에 대한 조사를 실시하여 정상행위 프로파일링을 작성하였다.



▶▶ 그림 5. 정교화된 개인정보 오남용 추출 모듈

그 결과 기존 개인정보 오남용 추출조건을 통해 추출된 의심로그라 할지라도 정상행위 프로파일에 의해 정상적인 업무처리 흐름이라고 판정되면 의심로그에서 제외할 수 있는 정교화된 추출 모듈을 [그림 5]와 같이 설계하였다.

정교화된 추출 모듈을 통해 제외된 정상로그 이외의 의심로그는 통합분석모듈을 거쳐 소명처리모듈로 전달되어 개인정보 오남용의 결과로 추출된다. [그림 6]는 정교화된 개인정보 오남용 추출 모듈인 비정상행위 탐지모듈을 가진 개인정보 오남용 추출조건 정교화 시스템의 구성도이다.



▶▶ 그림 6. 개인정보 오남용 추출조건 정교화 시스템 구성도

4. 시험 적용 및 평가

4.1 추출조건 적정성 평가 기준

추출조건의 시스템 적용여부는 추출조건의 적정성 판정결과에 따른다. 추출조건의 적정성 판정기준을 [표 3]을 이용하여 설명하면, 추출조건은 의심로그에 대한 과

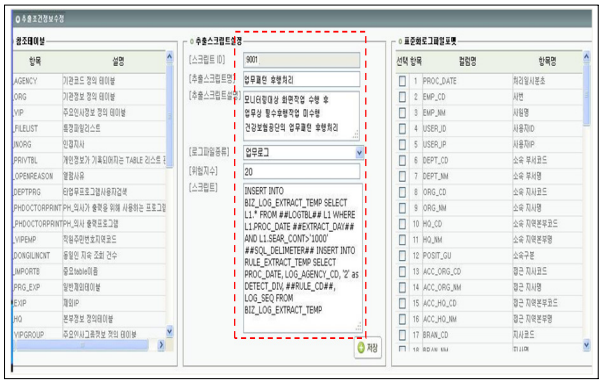
탐지 오류를 최소화하면서, 미탐지 오류가 발생시켜서는 안된다는 것이다. 즉 정상적인 업무처리 로그임에도 불구하고 개인정보 오남용 의심로그로 추출하는 과탐지 오류를 최소화하여야 하며, 오남용 의심로그로 추출되어야 함에도 불구하고 정상적인 업무처리 로그로 판정하여 제외하는 미탐지 오류가 없어야 한다는 것이다.

표 3. 추출조건 적정성 판정기준

구분		개인정보 오남용	
		의심	정상
추출 결과	의심로그 (추출된 로그)	정상탐지 (true positive)	과탐지 (false positive) ※과탐지시 적용불가
	정상로그 (제외된 로그)	미탐지 (false negative) ※미탐지시 적용불가	정상제외 (true negative)

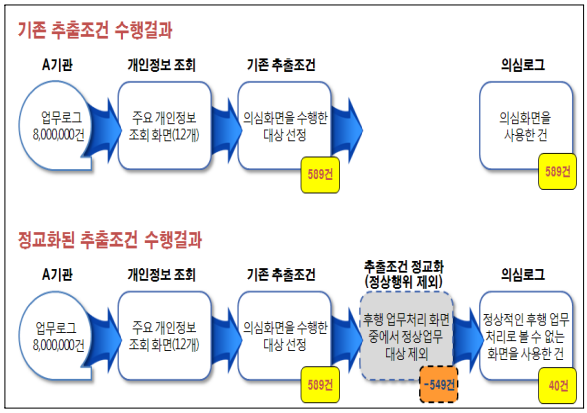
4.2 시험적용

본 연구를 통해 정교화된 추출조건의 시험적용을 위해 [그림 7]과 같이 시스템에 추출조건을 추가하였다.



▶▶ 그림 7. 정교화된 추출조건 시험적용

정교화된 추출조건의 적정성 평가를 위해서는 기존 추출조건과 전문인력에 의해 분석이 완료된 접속기록이 필요하다. 본 연구에서는 A기관의 약 8,000,000건 과거 접속기록을 샘플데이터로 활용하였다. A기관의 경우 총 192종의 개인정보 조회 프로그램이 존재하며, [그림 8]에서 보는 바와 같이 A기관에서 중점관리 중인 12개 개인정보 조회화면에 대한 접속기록 중 기존 추출조건만을 통해 의심로그를 추출하였을 시에는 총 589건 의심로그가 추출되었으나, 추출조건 정교화를 통해 개발된 비정상행위 추출조건을 병행하여 추출하였을 시에는 정상적인 후행 업무처리로 확인되어 549건이 제외됨에 따라 40건의 의심로그만 추출되어 과탐지 오류가 크게 개선된 것으로 평가되었다.



▶▶ 그림 8. 정교화된 추출조건 시험적용 결과 예시

다음으로 추출조건의 적정성 평가 중 미탐지 오류 발생여부를 검증하였다. 개인정보 오남용 의심로그에 대한 과탐지 오류가 크게 개선된다 하더라도 미탐지 오류가 발생하게 되면 추출조건을 적용할 수 없다. [표 4]에서 보는 바와 같이 샘플데이터 중 전문 인력에 의해 분석이 완료되어 과거 A기관에 소명 요청한 의심로그인 정상탐지(true positive)는 총 8건이었으며, 정교화된 추출조건을 통해 추출된 의심로그에서도 정상탐지 8건이 모두 포함된 것이 확인됨에 따라 정교화 과정 중 미탐지가 발생하지 않은 것으로 평가되었다. 따라서 본 연구에 의해 도출된 정교화된 추출조건의 시스템 적용이 적정한 것으로 평가되었다.

표 4. 정교화된 추출조건 적정성 평가결과

구분	기존 추출조건		정교화된 추출조건		소명요청 대상여부 검증결과
	의심로그	소명요청 대상로그	의심로그	소명요청 대상로그	
추출조건 적정성 평가	589건	8건	40건	8건	일치

4.3 정교화된 추출조건 확대적용 검토

정교화된 추출조건의 적용범위 확대를 검토하였다. 정교화된 추출조건은 [표 5]와 같이 기존 추출조건 전체에 반영할 수 있는 것은 아니다. 예를 들어 타지역IP 접속이나 휴직자나 퇴직자의 ID로 업무를 수행하는 경우는 관련법규에서 금하고 있는 ID 공유나 ID도용이 의심되기 때문에 의심로그에서 제외하여서는 안되며, 개인정보 오

남용 위험 가능성이 높은 기 유출자나 중점관리 대상이 업무처리를 수행 등도 마찬가지로 정교화된 추출조건을 적용하는데 적합하지 않은 것으로 검토되었다.

표 5. 정교화된 추출조건 적용여부

추출조건명	정교화된 추출조건 정의	정교화된 추출조건 적용여부
열람유의 대상자 조회	후행 업무처리가 발생하지 않은 열람유의 대상자로 등록된 고객정보 조회	○
직원간 조회	후행 업무처리가 발생하지 않은 동료 직원의 개인정보 조회	○
접속IP 탐지	등록된 지역의 IP대역 이외의 다른 IP대역 접속	×
특정 시간대 업무처리	선행/후행 업무처리가 발생하지 않은 특정 시간대(업무외 시간대, 공휴일, 휴일 등)에 업무처리	○
휴직/퇴직자 업무처리	휴직/퇴직자 ID로 업무처리 수행	×
보안 취약자	개인정보 기 유출자 및 기타 중점관리 대상자 업무처리 수행	×
퇴직 예정자 업무처리	퇴직 예정자가 업무처리 수행	×
개인별 임계치	개인별 개인정보 조회건수가 임계치 초과	×
관할지역 외 업무처리	후행 업무가 발생하지 않은 관할지역(지사, 지원 등) 외 타지역 고객정보 조회	○
열람사유 유형분석	열람사유 미등록 및 부실하게 등록하여 의심되는 경우	×
과다한 검색기능 조회	검색기능 조회가 임계치를 초과하여 과다 조회	×
개인정보 유출 의심행위	고객정보 조회 후 파일 변환 및 다운로드 등 개인정보 유출이 의심되는 행위	×
타 업무프로그램 사용여부	소속부서의 업무와 무관한 업무프로그램 사용 여부	×

5. 결론

사회 전반적으로 개인정보가 다양한 목적으로 폭넓게 활용되면서, 개인정보 유출 및 오남용에 대한 우려도 높아지고 있다. 이에 국가에서는 「개인정보보호법」을 제정하여 수집, 저장·관리, 이용·제공, 파기 등 개인정보의 처리단계에 따라 개인정보보호 조치를 의무적으로 이행하도록 제도화하였다.

본 연구는 개인정보 처리단계 중 개인정보의 이용 및 제공 단계에서 내부직원에 의한 개인정보 오남용행위를 효과적으로 탐지할 수 있도록 비정상행위 탐지방법 중 연관규칙기반의 정교화된 추출조건을 개발하고, 이를 기존의 오용탐지 추출조건과 하이브리드방식으로 적용한 추출조건의 정교화 및 적정성 평가를 실시하였으며, 이를 활용하기 위한 시스템 설계 방안을 제시하였다.

본 연구결과는 보건복지 개인정보통합관제시스템 뿐 아니라 내부직원의 개인정보 오남용행위를 모니터링하기 위한 시스템 구축을 고려중인 타 정부부처나 민간기관에서도 유용하게 활용할 수 있을 것으로 기대된다.

하지만 기존에 발견하지 못한 개인정보 오남용 의심로그에 대한 추출조건 개발과 개인정보 오남용 행위자에 대한 추가적인 분석을 위해서는 연관규칙, 클러스터링, 의사결정나무 등의 데이터마이닝 기법을 이용한 추가적인 향후 연구가 필요할 것으로 사료된다.

참고문헌

[1] 신영진, 정형철, 강원영, “공공분야 개인정보보호 정책 집행 과제의 우선순위 분석: 개인정보보호 수준진단 지표의 선정 및 중요도를 중심으로,” 정보보호학회논문지 22(2), 2012년 4월.

[2] 개인정보보호위원회, 2012 개인정보보호 연차보고서, 2012년 9월.

[3] 고유미, 최재원, 김범수, “개인정보 오·남용 방지 및 보호를 위한 정보공유센터 프레임워크,” 정보보호학회논문지, 22(2), pp.391-400, 2012년 4월.

[4] e-나라지표 홈페이지, http://www.index.go.kr

[5] 이기혁, 이철규, “내부정보 유출 징후 분석을 통한 유출방지 체계 구축에 관한 연구,”, 정보보호학회지, pp.70-79, 19(3), 2009년 6월.

[6] 행정안전부, 개인정보 보호법령 및 지침·고시 해설, 2011년 12월.

[7] 송유진, 이동혁, “개인정보 라이프사이클에 따른 프라이버시 보호 프레임워크,” 정보보호학회지, 16(4), pp.77-86, 2006년 8월.

[8] OECD, OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, 2002. http://www.oecd.org/internet/interneteconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[9] 보건복지부, 보건복지부 개인정보보호 기본지침, 2012년 1월.

[10] 보건복지부, 보건복지 개인정보통합관제센터 운영규정, 2012년 1월.

[11] 박성주, 임종인, “개인정보 유출 방지를 위한 SRI(Security Risk Indicator) 기반 모니터링 시스템 개발,” 정보보호학회논문지 22(3), pp.637-644, 2012년 6월.

[12] 조성규, 전문석, “개인정보보호를 위한 개인정보 유출 모니터링 시스템의 설계,”, 정보보호학회논문지, 22(1), pp.99-106. 2012년 2월.

[13] 김진형, 김형종, “개인정보 데이터 접근 비정상행위 탐지기법을 활용한 개인정보 보호 기법 연구,” 정보과학회지, 27(12), pp.60-67, 2009년 12월.

[14] K. Ilgun, R. A. Kemmerer, P. A. Porras, “State Transition Analysis : A rulebased intrusion detection

approach," IEEE Transaction on Software Engineering, 21(3), pp.181-199, March., 1995.
[15] S. Axelsson, "The Base-rate Fallacy and the Difficulty of Intrusion Detection," ACM Transactions on Information and System Security, 3(3), pp.186-205, 2000.
[16] 최윤정, 박승수, "이상탐지(Anomaly Detection) 및 오용탐지(Misuse Detection)분석의 정확도 향상을 위한 개선된 데이터마이닝 방법 연구," 2006 한국컴퓨터종합학술대회논문지 33(1), pp.238-240, 2006년.
[17] B. Mukherjee, T. L. Heberlein and K. N. Kevitt, "Network intrusion Detection," IEEE Network, 8(3), pp.26-41, May/June, 1994.
[18] P. A. Porras, "STAT: A State Transition Analysis Tool For Intrusion Detection," Proc. 17th National Computer Security Conference, pp.11-21, October 1994.
[19] T. D. GarveY, T. F. Lunt, "Model-Based Intrusion Detection." 14th National Computer Security Conference, October, 1991.
[20] H. S. Javitz, A. Valdes, "The SRI IDES Statistical Anomaly Detector," In Proc. of the 1991 IEEE Symposium on Research in Security and Privacy, May, 1991.
[21] R. Agrawal, T. Imielinski, A. Swami. "Mining association rules between sets of items in large databases," In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C., pp.207-216, May, 1993.
[22] 신대철, 김홍윤, "비정상행위 탐지 알고리즘 구현 및 성능 최적화 방안," 한국산학기술학회논문지, 11(11), pp.4553-4562, 2010년.

저자소개

● 이 기 호(Gi-Ho Lee)

- 1997년 2월 : 순천향대학교 환경보건학과(보건학사)
- 1999년 8월 : 연세대학교 보건대학원(보건학석사)
- 2013년 현재 : 숭실대학교 IT정책학과(박사과정)
- 2005년 9월 ~ 현재 : 한국보건사회연구원 전문연구원

<관심분야> : 개인정보보호, 건강정보, U-Health, 금연

● 정 영 철(Youngchul Chung)

- 1984년 2월 : 연세대학교 가정학과(이학사)
- 1987년 2월 : 연세대학교 보건대학원(보건학석사)
- 1996년 8월 : KAIST 테크노경영대학원(경영정보학석사수료)
- 2003년 2월 : 가톨릭대학교 대학원(보건학박사수료)
- 2007년 4월 ~ 현재 : 한국보건사회연구원 연구위원

<관심분야> : 개인정보보호, 보건복지정보시스템 구축 및 평가, 건강정보서비스 제공, 복지정보서비스 제공

● 이 야 리(Yari Lee) 정회원

- 1990년 2월 : 고려대학교 전자전산공학과(공학사)
- 1999년 2월 : 동국대학교 교육대학원(컴퓨터교육학석사)
- 2002년 8월 : 동국대학교 컴퓨터공학과(공학박사)
- 2004년 3월 ~ 2012년 5월 : 대학강사
- 2012년 6월 ~ 현재 : 한국보건사회연구원 초빙연구위원

<관심분야> : 개인정보보호, IT융합, 클라우드컴퓨팅, 모바일프로그래밍, 형식언어