Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.5, Oct. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.5.957

내부정보 유출 시나리오와 Data Analytics 기법을 활용한 내부정보 유출징후 탐지 모형 개발에 관한 연구

박 현 출,[1†] 박 진 상,[2] 김 정 덕 [3‡]
[1]삼일PwC(상무), [2]한국인터넷진흥원(연구원), [3]중앙대학교 산업보안학과(교수)

A Study on Development of Internal Information Leak Symptom Detection Model by Using Internal Information Leak Scenario & Data Analytics

Hyun-Chul Park,[1†] Jin-Sang Park,[2] Jungduk Kim[3‡]
[1]Samil PwC(Partner), [2]KISA(Researcher), [3]Chung-Ang University(Professor)

요 약

최근 산업기밀보호센터의 통계에 의하면 국내 기밀유출 사고의 경우 전·현직 직원에 의해 기업기밀유출의 약 80%를 차지하고 이러한 내부자에 의한 정보유출 사고의 대다수가 허술한 보안 관리체계와 정보유출 탐지기술의 이유로 발생하고 있다. 내부자의 기밀유출을 차단하는 업무는 기업보안 부문에서 매우 중요한 문제이지만 기존의 많은 연구들은 내부자에 의한 유출위협보다는 외부 위협에 의한 침입에 대응하는데 초점이 맞추어져 있다. 따라서 본 논문에서는 기업 내에서 발생하는 다양한 비정상 행위를 효과적이고 효율적으로 탐지하기 위해 내부정보 유출 시나리오를 설계하고 시나리오에서 도출 된 유출 징후의 핵심 위험지표를 데이터 분석(Data analytics)함 으로써 정교하지만 신속하게 유출행위를 탐지하는 모형을 제시하고자 한다.

ABSTRACT

According to the recent statistics of the National Industrial Security Center, about 80% of the confidential leak are caused by former and current employees in the case of domestic confidential leak accidents. Most of the information leak incidents by these insiders are due to poor security management system and information leak detection technology. Blocking confidential leak of insiders is a very important issue in the corporate security sector, but many previous researches have focused on responding to intrusions by external threats rather than by insider threats. Therefore, in this research, we design an internal information leak scenario to effectively and efficiently detect various abnormalities occurring in the enterprise, analyze the key indicators of the leak symptoms derived from the scenarios by using data analytics and propose a model that accurately detects leak activities.

Keywords: Internal information leak, scenario, data analytics, anomaly detection, risk indicators

I. 서 론

정보통신기술이 발전함에 따라 정보유출의 수단이 되는 매체와 유출 경로가 다양해지고, 유출 방법이 고도화되고 있다. 기업의 유무형의 기술이나 경영상 정보인 산업기밀 유출은 속해있는 산업 분야에서 경쟁력을 상실하게 할 뿐 아니라 고객 신뢰도를 저하시키고 기업 평판에 악영향을 끼치는 등의 막대한 손해를 수반한다[1]. 기업 정보유출 사고 동향을 보면 기업 내 핵심기술 및 기밀정보에 대한 유출사례는 증가하고 있으며 특히 전·현직 직원과 같은 내부자에 의

Received(09. 04. 2020), Accepted(09. 29. 2020)
† 주저자, hyunchul.park@pwc.com
‡ 교신저자, jdkimsac@cau.ac.kr(Corresponding author)

한 정보유출 사고의 대다수가 허술한 보안 관리체계와 정보유출 탐지기술의 이유로 발생하고 있다. 산업기밀보호센터의 통계에 의하면 국내 기밀유출 사고의 경우 전·현직 직원에 의해 기업기밀유출의 약 80%를 차지하고 협력업체에 의한 유출 또한 꾸준히 증가하고 있는 추세이다[2]. 즉, 국내 기업의 보안은 해킹과 같은 외부 위협을 차단하는데 있어 현저한 노력을 기울이고 있으나 내부자에 의한 기밀유출을 방지하는데 있어 한계를 보이고 있다.

내부자의 기밀유출을 차단하는 업무는 기업보안 부문에서 매우 중요한 문제이지만 기존의 많은 연구들은 외부 위협에 의한 침입에 대응하는데 초점이 맞추어져 있고 내부자에 의한 유출을 사전에 효과적으로 차단하고 대응하는 것에 대한 연구는 부족한 실정이다. 국내 기업 대부분의 경우 출입통제, 저장매체 통제, 문서보안, 이메일 통제 등의 여러 보안시스템으로부터 접근 및 사용 로그를 수집하고 모니터링 한다[3]. 그러나 여러 시스템에서 수집한 로그의 양이 방대하고, 시스템 오탐과 내부자의 보안우회가 많아 기밀정보 유출을 정교하고 신속하게 탐지 및 차단하는 것에 한계가 있다. 또한 기업기밀 유출은 계획적이고 치밀하게 이루어지기 때문에 효과적인 모니터링 체계 없이는 신속하게 유출징후를 잡아내고 대응하기 힘들다.

따라서 본 연구에서는 기업 내에서 발생하는 다양한 비정상 행위를 효과적이고 효율적으로 탐지하기 위해 내부정보 유출 시나리오를 설계하고 시나리오에서 도출 된 유출 징후의 핵심 위험지표를 데이터 분석(Data analytics)함 으로써 정교하지만 신속하게 유출행위를 탐지하는 모형을 제시하고자 한다.

II. 관련 연구

2.1 기업의 내부정보 유출 탐지 현황 분석

최근 기업들은 내부자에 의한 기밀정보 유출과 정보 시스템 파괴 등의 각종 보안 사고를 겪으며 해당 산업에서의 경쟁력 감소, 기업 평판 감소와 고객 신뢰 하락 등의 막대한 손해를 보고 있다. 기업 내부자에 의해 발생하는 정보유출과 시스템 파괴 등의 보안 위협은 해킹과 같은 외부 공격에 의한 보안 위협보다 증가하고 있는 추세이다. 보안사고의 수뿐만 아니라 피해 규모의 경우에도 외부에 의한 보안 위협보다 내부자에 의한 보안위협이 더욱 크다. 내부자는 기업 내에서 운영되는 시스템과 보안 환경을 외부인 보다 잘 파악하고 있고 접근권한이 있기 때문에 오랜 시간에 걸쳐 치밀하게 정보유출을 할 수 있기 때문이다[4]. 기업에서는 이러한 내부자에 의한 내부정보 유출을 탐지하고 차단하기 위해 관리적 보안통제와 각종 보안 통제 솔루션과 시스템을 이용한 기술적 통제를 수행하고 있다[5]. 기업은 내부 보안 솔루션에서 발생하는 로그를 수집하고 모니터링 하여 내부자의 보안 위협 행위를 탐지한다.

기업뿐만 아니라 로그 수집을 통한 내부자 정보유출 탐지 방안에 대한 학술적 연구도 많이 진행되고 있다. 정귀영[6]은 사용자의 이상 징후를 탐지하기 위해 기업 내부의 다양한 보안 솔루션 및 관련 시스템의 사용자 행위로그를 분석하였고 위협행위 시나리오를 도출하였다. 하지만 해당 연구는 보안 솔루션과 시스템마다 발생하는 로그 정보가 방대하여 실시간으로 탐지시스템과 연동 할 수 없었고 시나리오의 탐지 임계치 정교화 작업이 부재하여 시나리오의 탐지 신뢰성이 부족한 한계를 보였다. 박장수와 이임영[7]은 보안 솔루션에서 발생된 로그를 분석하고 정보유출 시나리오를 설계하는 연구를 수행하였다. 문서보안(DRM), 유해사이트 차단 시스템, 매체제어와 같은 보안 솔루션에서 수집한 로그를 통해 28개의 시나리오를 생성했고 수집된 로그를 분석하는 방안 또한 설명하였다. 그러나 해당 연구가 제시한 시나리오는 업무상 발생하는 예외적인 상황을 고려하지 않아 오탐에 대한 해결방안을 제시하지 못하였다.

김은선[8]은 이상 징후를 탐지하기 위해 사용자의 접근 로그를 사용하여 사용자 특성과 시스템 접속횟수 사이의 관련성을 통계기법을 통해 분석하였고 이에 관련된 이상 징후 탐지모델을 제시하였다. 해당 모델은 시나리오를 이용하여 사용자 특성과 비정상 행위와의 유의미 관계를 입증하였지만 비정상행위에 영향을 미치는 다양한 변수들을 고려하지 못하였고 오탐이 과도해지는 문제점을 가지고 있다. 최재혁[9]은 보안 위협을 탐지하기 위해 총 4 단계의 시나리오 개발 가이드라인을 제안하였다. 이 연구는 기업 보안 시스템에서 발생하는 로그의 상관분석을 하여 패턴을 발견하고 보안 시나리오 설계를 위한 방법을 체계적으로 설명하지만 유출탐지 알고리즘이 부재하여 알려지지 않은 위협에 대응하기 어렵고 정교한 분석이 이루어지지 못한다. 뿐만 아니라 각종 이기종 솔루션과 시스템에서 수집되는 로그를 기계가 아닌 사람이 수동으로 형식을 통합해서 분석해야 하는 비효율성의 한계가 있다.

기업의 내부정보 유출 탐지에 관한 연구 동향을 분석한 결과, 대부분 내부자에 의한 기밀 유출을 탐지하기 위해 보안 시나리오를 설계하고 보안 솔루션과 시스템에서 발생하는 로그를 수집 및 분석한다. 보안 시나리오 설계는 보안 관리자가 유출징후를 추출하기 위한 지표 선정에 유의미한 정보를 제공하고 유출 대상에 따른 위험도 분류가 수월하기 때문에 효과적인 방안이다. 하지만 이기종의 보안 솔루션과 시스템에서 수집되는 막대한 양의 로그를 사람이 일일이 통합하고 분석하는 것은 정교하지 않을뿐더러 비용 효과적이지 않다. 또한 알려지지 않은 위협이나 유출자의 특이 행위에 대해서는 단일 시나리오 및 로그 분석만으로는 탐지하기 힘든 한계가 있다.

2.2 데이터 분석(Data Analytics)

데이터 분석은 비즈니스의 목적을 달성하기 위해 기존의 데이터를 분석하여 시사점을 얻고 이를 통해 의사결정을 하는 프로세스이다. 기술적인 측면에서 데이터 분석은 사용자가 데이터로부터 패턴을 파악하여 새로운 가치가 있는 데이터를 창출하는 기술이나 프로세스를 의미한다. 데이터 분석의 유형으로는 첫째로, 추세, 패턴 및 관계를 식별하기 위해 대량의 데이터를 분류하는 데이터 마이닝(Data mining)이 있고 둘째로, 고객의 행동, 장비 고장 및 기타 미래 사건을 예측하는 예측 분석(Predictive analytics)이 있다. 그리고 셋째로, 사용자가 기존에 수동으로 분석 모델링을 하는 것보다 더 신속하게 데이터 세트를 특정 알고리즘에 대입하여 자동화 된 결과를 도출하는 기계 학습(Machine learning)이 있다. 데이터 분석은 데이터 마이닝, 예측 분석 및 기계 학습 도구를 이용하여 구조적 및 비구조적 데이터를 포함하는 대량의 데이터 세트에서 사용자가 찾고자 하는 의미 있는 데이터를 재생성하는 과정이다.

데이터 분석을 통한 내부자의 기밀유출 탐지에 관한 연구들이 활발히 진행되고 있다. 임원기 등[10]은 데이터 유출 탐지를 위한 이상행위 탐지 방법의 정확도 향상을 위해 프로파일 기반의 탐지 방법과 기계학습 기반의 탐지 방법의 장단점을 분석하였다. 해당 연구는 두 가지의 탐지 방법을 비정상 행위의 경계성, 비적합성, 유연성, 다양성, 유효성, 민감성에 따라 장단점을 분석하였지만 내부자의 기밀유출 탐지를 위한 구체적인 대안을 제시하지 않는다. 박성만과 정충교[11]는 기업내부자의 정보유출 방지를 위해 행위 이상점 분석을 하였다. 이 연구는 DLP 서버에서 수집된 로그를 분석하고 유출자의 행동 속성에 따라 횡단면 분석 및 시계열 분석을 통해 특이행동을 탐지하고자 하였다. 하지만 해당 연구는 유출 징후를 탐지할 때 유출자의 행동속성을 고려해야 한다는 시사점은 도출하였지만 구체적인 유출자의 행동속성 추출 방안이나 탐지를 위해 수집해야 하는 주요 지표는 제시하지 못 하였다.

김권일과 장병탁[12]은 내부자의 이상탐지를 위해 Product of Experts(PoE) 모델을 이상탐지에 적용하여 학습 알고리즘을 제안하고 정상/이상 패턴을 추출하는 모델을 제시하였다. 하지만 해당 연구는 탐지 모델에서 분류 성능에 초점을 맞추기 때문에 실제 모델에 적용할 내부자의 기밀유출 행위 속성에 대한 연구가 부족하였다. 하동욱 등[13]은 기계학습 기반의 내부자위협을 탐지하기 위해 RNN Auto encoder를 이용하여 비정상행위 탐지 실험을 하였다. 기존 연구들이 사용자의 행위를 주 단위로 탐지하기 때문에 탐지의 효율성이 떨어지는 문제가 발생하였는데 해당 연구는 유출행위 5개의 시나리오를 세우고 사용자 행위를 일 단위로 탐지하여 내부자 위협행위가 일어난 후 신속하게 조치를 취할 수 있게 하는 유효성을 검증하였다. 하지만 해당연구의 실험은 시나리오 별 추출한 표본이 30명으로 표본 수가 매우 적고 오탐의 문제를 해결하기 위한 대안을 제시하지 않았다.

연구의 동향을 분석한 결과, 내부자의 기밀유출 탐지를 위한 데이터 분석은 대부분 규칙 기반, 통계 기반과 기계학습 기반의 탐지를 이용하고 각 데이터 분석 방법마다 한계를 가지고 있다. 규칙 기반의 탐지는 매번 사람이 임계치와 규칙을 만들어야하기 때문에 비효율적이며 알려지지 않은 위협에 대한 탐지가 미흡하다. 통계 기반 탐지는 규칙 기반 탐지에 비해 오탐이 많고 대량의 데이터 축적과 많은 학습시간이 필요하다. 기계학습 탐지는 알고리즘을 구축하는 초기 비용이 매우 크고 알고리즘에 사용되는 변수를 선택하는 과정이 어렵다. 또한 변수가 증가할수록 기계학습 모형의 성능이 저하된다는 단점이 있다. 따라서 본 연구는, 오탐을 줄이기 위해 내부자의 행위속성을 고려하여 정보유출 시나리오를 설계하고 자동화된 탐지 시스템 구축을 위해 통계 기반 탐지와 기계학습을 혼합한 내부 기밀유출 징후 탐지 모형을 제안한다. 규칙 기반의 탐지, 통계 기반의 탐지와 기계학습 기반의 탐지의 특징은 〈표 1〉과 같이 요약할 수 있다.

Table 1. Types of Data Analytics for Information Leak Detection

Division	Feature
Rule-based method	• Rapid detection by predefined threshold criteria • Incomplete detection of unknown threats • Detection does not consider user characteristics
Statistical -based method	• Rapid change detection based on normal behavior • Can detect unknown threats • Required large amounts of data & learning time • High false positives compared to rule-based method
Machine learning -based method	• Self detection by analyzing user behavior • Inadequate implementation complexity and validation • Difficulty in obtaining training data

III. 연구모형 개발

3.1 내부정보 유출징후 탐지를 위한 모형

내부정보 유출징후 분석이란 내부정보 수집, 저장 및 보안우회를 통해 외부로 전송하는 과정을 분석하여 유출징후를 점검하고 조치하는 일련의 보안검토 활동이다. 내부정보 유출 시나리오에 기반 하여 유출징후 위험지표를 추출하고 Data analytics 알고리즘을 통해 데이터를 분석 하고 포렌식을 동반한 실증

Fig. 1. Model for Detection of Internal Information Leak Symptom

조사를 하여 유출 징후자를 색출한다. 내부정보 유출징후 탐지를 위한 모형은 다음 〈그림 1〉과 같다.

3.2 내부정보 유출징후 탐지 모형의 단계별 분석

3.2.1 내부정보 유출 시나리오 설계

기업의 내부정보를 유출하기 위해서 유출 행위자가 접근하고 열람 및 반출하는 모든 행적은 기업 시스템에 로그 데이터로 남는다. 일상적인 업무활동과 다른 비정상적인 행위를 탐지하기 위해서는 기업의 보안 시스템과 업무 시스템에서 발생하는 로그 데이터를 수집 및 분석하여야 한다. 하지만 기업의 보안 시스템과 업무 시스템에서 발생하는 수많은 로그 데이터에서 정상적인 행위와 비정상적인 행위를 구분하는데 어려움이 따르고 기존의 이상치 탐지 알고리즘은 오탐 발생이 많아 내부정보 유출 탐지가 비효율적으로 이루어진다는 한계점을 가진다. 이러한 오탐 문제를 최소화하고 Data analytics를 활용한 내부정보 유출을 탐지하기 위해서는 정상행위와 유출징후를 구분하는 주요 지표추출이 필요하며 이는 내부정보 유출 시나리오를 통해 해결될 수 있다. 단편적으로 수행되

Table 2. Design Criteria for Internal Information Leak Scenario

5W	Who	• Who wants to leak internal information • (Examples: Internal employee, outsourcing)
	When	• When you want to leak internal information • (Examples: Work hours, document discard time)
	Where	• Where you want to leak internal information • (Examples: Business computer, restricted area)
	What	• Internal information you want to leak • (Examples: Electronic information, document)
	Why	• Reason to leak internal information • (Examples: Malice, ignorance)
1H	How	• How to leak internal information • (Examples: E-mail, portable storage)

던 정보유출 예방활동을 정보유출 시나리오에 기반으로 한 유기적인 대응체계로 전환하기 위해서는 내부정보 유출경로 식별을 시스템 및 매체 관점이 아닌 사람의 행위 관점으로 경로를 식별할 필요가 있다. 내부정보 유출의 주체는 사람이기 때문에 "누가", "언제", "어떤 정보를", "어디서", "어떻게", 그리고 "왜"의 5W1H 관점으로 유출경로를 식별하는 시나리오 설계가 필요하다. 내부정보 유출 시나리오 설계기준은 〈표 2〉와 같다.

3.2.2 데이터 수집 및 분석환경 구축

Data analytics를 활용하여 데이터를 분석하고 내부정보 유출 징후자를 추출하기 위해서는 내부정보 유출 시나리오에서 기 정의된 유출징후 핵심 위험지표에 따라 관련 데이터를 수집하고 데이터 분석환경을 구축해야 한다. 내부정보를 유출 할 가능성이 높은 부서 및 인력과 유출을 위해 접근하는 관련 내부 시스템을 식별하고 이와 관련된 로그 데이터를 수집해야 한다. 예를 들어 정보자산 관리자가 기업의 핵심기술 문서를 출력하여 외부로 반출하려는 시나리오의 경우, 해당 유출 징후자의 사용자 ID, 접속 시간, 문서 조회 시간, 출력통제 솔루션의 로그 등을 수집하여야 한다. 이와 같이 시나리오에 따른 관련 로그를 수집한 후 DB에 적재하기 위해서는 로그의 종류별로 데이터베이스 테이블을 생성하고 수집한 로그의 형태와 시간을 표준화하는 데이터 전처리 작업이 필요하다. 분석 데이터베이스를 구성을 할 시, 어플리케이션 관련 로그, 네트워크 관련 로그, 데이터베이스 관련 로그, 기타 보안 솔루션 로그와 같이 주요 필드를 식별하고 선정된 필드에 따라 데이터베이스 테이블을 생성한다. 또한 기업에 존재하는 다양한 시스템과 보안 솔루션마다 발생하는 로그의 형태가 상이하기 때문에 수집한 로그들의 표준화 작업을 해야 한다. 이러한 데이터 수집과 데이터베이스 구성 그리고 데이터 전처리 작업을 완료하면 내부정보 유출 징후 분석을 위한 데이터베이스에 해당 데이터들을 적재한다. 데이터 수집 및 분석환경 구축은 다음 〈그림 2〉와 같다.

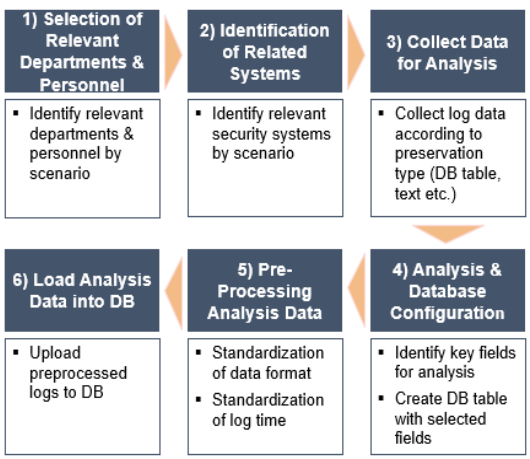

Fig. 2. Data Collection & Constructing Analysis Environment

3.2.3 데이터 분석 및 유출 징후자 추출

데이터 분석 및 유출 징후자 추출은 Data analytics 기법을 활용하여 다수의 내부정보 유출 위험지표, 직원 그리고 시간 정보를 토대로 비교 가능하도록 정규화 및 계층화 작업을 통해 정상 행위의 범위를 벗어난 직원을 탐지하고 추출하는 단계이다. 앞서 관련연구에서 설명한 것과 같이 내부정보 유출 징후를 탐지하기 위해 기존의 많은 연구들이 규칙 기반의 탐지 방법론을 사용하였다. 하지만 해당 방법론은 알려지지 않은 위협에 대처하기 힘든 단점을 가지고 있기 때문에 본 연구에서는 내부정보 유출 시나리오에서 정의된 유출징후 위험지표를 통해 오탐을 줄이면서 통계 기반과 기계학습 기반의 내부 기밀유출 징후 탐지 방법론을 통해 자동화된 내부정보 유출 탐지를 하고자 한다. 또한 통계 기반 탐지를 위해 본 연구는 이상감지(Anomaly Detection)기법을 사용하고자 한다. 이상감지는 데이터마이닝의 일부로 시계열 데이터에서 과거 또는 현재 시점의 보편적인 패턴에서 벗어나거나 벗어나려는 징후가 있는 개체를 찾아내는 데이터 분석 기법이다. 내부정보 유출 시나리오에서 도출된 위험지표 분석 데이터를 이상감지 기법에 적용하여 위험지표 별 크기와 위험 크기 간 거리 계산을 하여 정상행위의 군집을 형성하고 이러한 군집에서 벗어난 위험 징후자를 추출한다. 위험지표 분석 데이터를 이상감지 기법에 적용하는데 앞서 개별 위험지표의 크기와 방향 측정을 통해 변수의 벡터화 및 시계열화 작업이 필요하며 이상감지 기법에 적용 후에는 상이한 특성을 가진 위험지표들을 비교하기 위해 정상행위 군집의 정규화 및 계층화 작업이 필요하다. 이러한 Data Analytics 기법을 활용한 데이터 분석 및 유출 징후자를 추출하는 단계는 다음 〈그림 3〉과 같다.

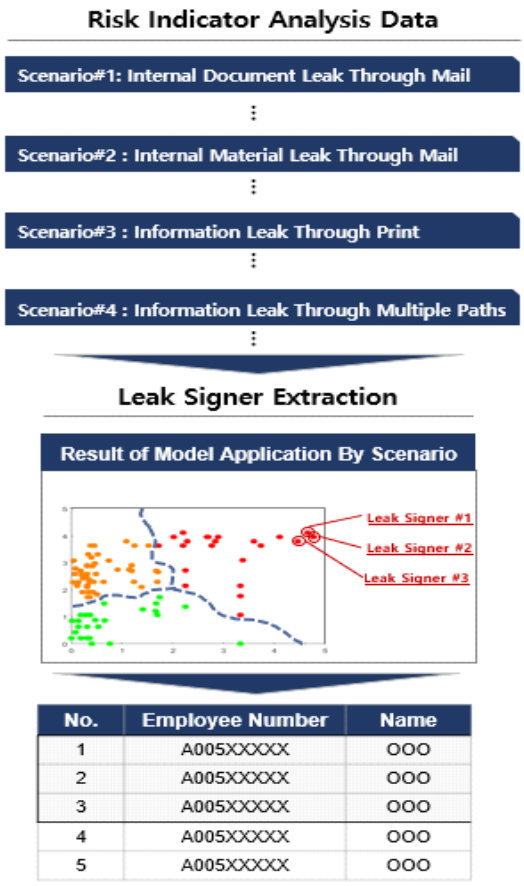

Fig. 3. Data Analysis & Leak Signer Extraction

3.2.4 실증조사 수행 및 내부정보 유출 대책 수립

실증조사는 내부정보 유출 징후자로 탐지된 대상에 대해 실제 내부정보 유출이 발생한 여부를 판단하기 위한 단계이다. 유출 징후자로 의심되는 대상에 대해 포렌식 기법을 적용한 실증조사를 수행하고 실증조사를 위한 동의서가 필요하다면 징구절차를 가져야 한다. 포렌식 이란 PC나 스마트폰과 같은 디지털기기에 들어있는 데이터를 수집 및 추출하고 이를 바탕으로 내부정보 유출과 같은 범죄의 증거를 찾아내는 과학수사 기법을 말한다. 판별된 유출 징후자의 PC, 노트북, 스마트폰, 이메일 발송내역과 기업의 시스템 서버에 존재하는 데이터를 수집하고 분석하여 내부정보가 실제로 유출되었는가의 여부를 확인한다. 이러한 내부정보 유출 탐지 모형과 포렌식 조사 결과를 바탕으로 전사적인 관점의 내부정보 유출 사고 이슈를 도출하고 향후 개선 및 재발방지를 위해 대책을 수립한다.

IV. 연구 검토

본 논문에서 앞서 제안한 내부정보 유출징후 탐지 모형의 실무적 효과성과 효율성을 검토하기 위해 제조회사 Z기업을 대상으로 내부정보 유출 시나리오를 바탕으로 한 총 7개의 모형으로 내부 유출자를 탐지하기 위한 프로젝트를 수행하였다. 프로젝트는 총 5개월 동안 R&D센터 300여명의 연구원을 대상으로 수행, 내부정보 유출 탐지 모형을 적용하였고 총 13명의 내부정보 유출 징후자를 도출하였다. 이에 따라 실증조사를 한 결과, 8명이 실제 내부정보 유출자이거나 유출을 하고자 했던 자였다. 본 연구 검토에서는 여러 프로젝트 사례 중 한가지의 내부정보 유출 적발사례를 소개하고자 한다.

본 프로젝트의 대상 Z기업은 제조/생산 기업으로 신규 R&D 기술이 제품 경쟁력에 큰 영향을 주는 업종으로 이러한 기업 특성상 연구 인력의 인적 보안이 매우 중대한 상황이었다. 이에 따라, R&D센터의 연구 인력을 중심으로 본 논문에서 제시한 내부정보 유출징후 탐지 모형을 적용하였고 실제 데이터 검증을 통해 내부정보 유출 의심자(Suspect)를 도출하였다. 최종적으로 내부정보 유출 징후자를 대상으로 포렌식을 동반한 실증조사를 하여 연구모형의 효과성 및 효율성 검증을 하였다.

데이터 분석에 앞서, 각 조직 별로 유출 대상이나 경로 등의 상황이 다르기 때문에 프로젝트 대상인 A기업 상황에 맞춰 모델링 및 핵심 위험지표를 선정할 필요가 있었다. 이에 따라, Z기업의 제조업 환경에 맞춘 내부정보 유출 시나리오를 정의하고, 이러한 징

Scenario#1: Internal Document Leak Through Mail	
Drawing Security Decryption Over Request	Store
Drawing Security Decryption Over Approval	Store
Increased Storage of Documents in a Short Period	Store
Increase in File Retention with Important Keywords	Store
Increase in Critical Information Downloads	Store
Excessive Document Storage Contrast Decoding	Store
Increase Access to External Mail	Transfer
Send Email with Important Keywords	Transfer
Transfer Important Information with Email	Transfer
Send Excessive Mail to External Email Address	Transfer

Fig. 4. Key Risk Indicators of Leak Symptom

후를 탐지하기 위한 핵심 위험지표(KRI: Key Risk Indicator)를 정의하였다. 모델링을 진행하기 위해 해당 기업에서 내부정보유출이 가능한 경로를 식별하고 그에 대한 내부정보유출 시나리오들을 설계하였다. 설계한 시나리오마다 유출징후를 판별하는 핵심 위험지표들을 산출하였고 해당 위험지표들의 예시는 다음 〈그림 4〉와 같다.

연구모형 절차에 따라 내부정보유출 시나리오 별로 산출한 핵심 위험지표들을 기반으로 Z기업의 실제 데이터 분석을 하였고 각 모델링 별 결과를 통해 상위 Anomaly Score의 유출 의심자 명단을 도출하였다. Z기업의 메일을 통한 내부정보유출 모델링 결과는 다음 〈표 3〉과 〈그림 5〉와 같다.

메일을 통한 내부정보유출 모델링 결과 Rank 1위로 탐지된 A 연구원의 세부 위험 지표 결과는 다음 〈표 4〉와 같다.

명단 중 가장 점수가 높은 A연구원의 유출 혐의에 대해 포렌식을 동반하여 실증조사를 수행하였다. 실증조사 일환으로 포렌식 조사를 수행하기 위해 DB 모니터링 솔루션에 의해 수집된 SQL 접근 로그를 수집·분류·분석하였다. DB 접근기록 분석 결과는 다음 〈표 5〉와 같다.

계약직으로 근무 중이었던 Y부서 R&D팀의 A연구원은 재계약이 성사되지 않아 퇴사해야 하는 상황에서 경쟁사 입사 제의를 받았었다. A연구원은 퇴사 전 영업정보, 기술개발 관련 자료 등 사내 주요 문서 150건을 유출 중에 있었다. 비서 X에게 외부저장매체를 전달받은 후 업무 외 시간에 자신의 권한으로 암호화된 자료를 복호화한 후 외부저장매체에 저장하여 반출한 것으로 판별되었다.

Table 3. Modeling Results of Internal Information Leak through Mail

Rank	Employee Number	Name
1	**16129X**	**A Researcher**
2	15014X	B Researcher
3	33167X	C Researcher
4	73155X	D Researcher
5	73045X	F Researcher

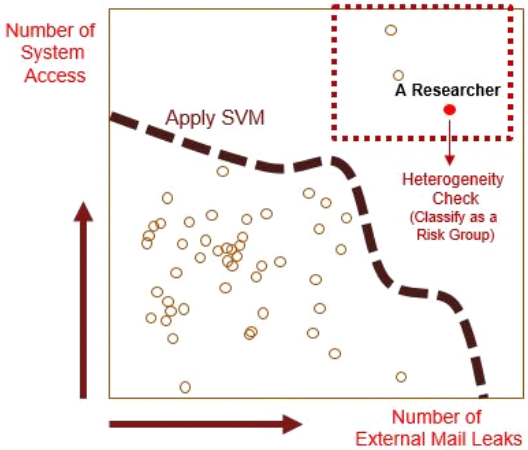

Fig. 5. A Researcher's Modeling Results of Internal Information Leak through Mail

Table 4. Results of the Risk Indicator of the A Researcher

Risk Indicator	Result
Excessive document storage	Caution
Excessive download of confidential documents	Danger
Storage of external documents	Caution
Accessing external mail	Danger
Critical keyword mail	Caution
Excessive attachments in mail	Danger
Excessive external mail address	Danger

Table 5. DB Access Record of the A Researcher

No	Individual Verification Results	DB Access Query Counts
1	Automatically query application execution	34
2	Master Query	1
3	A month inquiry of comprehensive information table	1
4	Access to bulk data outside of business hours	51
5	Create dataset for delivery test operation	4
6	Computerized management operation	10
Sum		101

Table 6. A Case of Internal Information Leak

Division	Description
Who	Contract employee (Scheduled to leave, high authority)
When	Off business hours
Where	External storage medium, business PC
What	Business information, technology development information etc. more than 100 cases
How	Self cancelled DRM

이와 같이, 본 논문의 연구모형을 실제 기업에 적용하였고 실제 유출사고가 일어나기 전에 효과적이고 효율적으로 내부정보 유출자를 탐지하고 적발하였다. 유출 미수 사례뿐만 아니라 기존 기업에 있었던 유출 사고 데이터를 연구모형을 적용했을 때 사고 결과와 동일하게 유출 징후를 탐지하여 연구모형의 효과성이 실질적으로 높음을 입증하였다. Z기업의 내부정보 유출 적발사례 요약은 다음 〈표 6〉과 같다.

V. 결 론

기업 및 기관에서 꾸준히 발생하는 내부기밀유출 사고를 정교하면서 신속하게 차단하기 위해서는 정밀한 보안 관리체계와 정보유출 탐지기술이 구축되어야 한다. 그에 따라 기업과 기관은 여러 보안시스템으로부터 내부 직원의 접근 및 사용 로그를 수집하고 모니터링 하고 있다. 그러나 기존 보안시스템의 오탐이 심하고 계획적이고 치밀하게 이루어지는 내부자의 보안우회 때문에 효과적인 모니터링 체계가 부재한 상황에서 내부기밀유출을 신속하게 차단하기에는 어려움이 있다.

따라서 본 논문에서는 기업 내에서 발생하는 다양한 비정상 행위를 효과적이고 효율적으로 탐지하기 위해 내부정보 유출 시나리오를 설계하고 시나리오에서 도출 된 유출 징후의 핵심 지표를 데이터 분석 함으로써 정교하지만 신속하게 유출행위를 탐지하는 모형을 제시하였다. 기존 연구들은 내부자의 위협보다는 외부 위협을 차단하는데 중점을 두었고 대부분의 연구들이 보안시스템의 오탐 문제를 해결하지 못하여 실무적으로 한계를 보였다. 한편, 본 연구는 기밀유출의 주체가 사람이라는 점에 착안하여, 내부자의 기밀유출 행위를 5W1H 관점으로 시나리오를 설계하고 유출행위를 탐지하는 핵심 위험지표를 추출하여 데이터 분석하는 방안을 제안하였다.

본 연구는 학술적 및 실무적으로 의미 있는 결과를 도출하였지만, 각 기업과 기관의 데이터 분석 역량에 따라 연구에서 제안한 모델이 공통적으로 적용되기에는 한계가 존재한다. 또한 시나리오에서 추출한 핵심 위험지표들을 기반으로 한 이상감지는 꾸준히 오랜 기간에 걸친 내부기밀유출 행위를 탐지하기에는 한계가 있다. 따라서 향후 연구로는 데이터 분석역량이 부족한 중소기업에서 적용 가능한 내부기밀유출 탐지 모델 개발과 내부자의 치밀한 유출행위까지도 탐지 가능한 알고리즘 개발에 관한 연구가 필요하다.

References

[1] Gi-Hyouk Lee and Cheol-Gyu Lee. "A Study on the Construction of Leak Prevention System through Analysis of Internal Information Leak Symptom." Journal of The Korea Institute of Information Security & Cryptology, 19(3). pp. 70-79. Jun. 2009

[2] National Industrial Security Center, http://service12.nis.go.kr

[3] Hyun-Tak Chae, "Security policy proposals through PC security solution log analysis : prevention leakage of personal information," master's thesis, Korea University, Feb. 2015.

[4] Jung-Ho Eom, "The Quantitative Evaluation of a Level of Insider Activity using SFI Analysis Techniques," Journal of security engineering, 10(2), pp. 113-122, Apr. 2013

[5] Kwang-Woo Lee and Seung-Joo Kim. "Analysis of Trends in Digital Multifunction Device Security Technology from the Viewpoint of Preventing and Protecting Business Confidential Information." Journal of The Korea Institute of Information Security & Cryptology, 20(1). pp.

47-55. Feb. 2010
〔6〕 Gui-Young Jung, "A Study on the Effective User Anomaly Detection Method through Integrated Security Log Analysis," master's thesis, Yonsei University, Feb. 2017.
〔7〕 Jang-Su Park and Im-Yeong Lee, "Information Security : Log Analysis Method of Separate Security Solution using Single Data Leakage Scenario," KIPS Transactions on Computer and Communication Systems, 4(2), pp. 65-72, Feb. 2015
〔8〕 Eun-Seon Kim, "A Study on the Anomaly Detection using User Log : ERP System An Empirical Study," master's thesis, Korea University, Feb. 2015.
〔9〕 Jae-Hyouk Choi, "A Study on A Scenarios Development Guideline for Detecting Security Threats," master's thesis, Korea National Open University, Feb. 2016.
〔10〕 Won-Gi Lim, Koo-Hyung Kwon, Jung-Jae Kim, Jong-Eon Lee and Si-Ho Cha, "Comparison and Analysis of Anomaly Detection Methods for Detecting Data Exfiltration," Journal of Korea Academia-Industrial cooperation Society, 17(9), pp. 440-446, Sep. 2016
〔11〕 Sung-Man Park and Choong-Kyo Jeong, "Enterprise Data Loss Prevention Using Behavior-Based Outlier Detection," Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, pp. 94-95, Nov. 2016
〔12〕 Kwon-Ill Kim and Byoung-Tak Zhang, "Hybrid Product of Experts Model and Learning Algorithm for Anomaly Detection," Proceedings of Conference of the Korean Institute of Information Scientists and Engineers, pp. 1553-1555, Jun. 2013
〔13〕 Dong-Wook Ha, Ki-Tae Kang and Yeong-Seung Ryu, "Detecting Insider Threat Based on Machine Learning : Anomaly Detection Using RNN Autoencoder," Journal of The Korea Institute of Information Security & Cryptology, 27(4), pp. 763-773, Aug. 2017

〈저 자 소 개〉

박 현 출 (Hyun-Chul Park) 정회원
1998년 2월: 중앙대학교 정보시스템학과 졸업
2000년 8월: 중앙대학교 정보시스템학과 석사
2000년 12월~현재: 삼일 PwC (상무)
2010년 11월~2013년 3월: PwC 호주 (Secondment)
2018년 3월~현재: 중앙대학교 융합보안학과 박사과정
〈관심분야〉 정보보호 전략/ 거버넌스, 내부통제/ 내부감사, 개인정보보호/ 정보유출

박 진 상 (Jin-Sang Park) 정회원
2017년 8월: 중앙대학교 경영학과 졸업
2019년 8월: 중앙대학교 융합보안학과 석사
2019년 10월~현재: 한국인터넷진흥원 연구원
〈관심분야〉 정보보호 거버넌스, 보안관리체계, 위험관리, 5G 보안

김 정 덕 (Jungduk Kim) 종신회원
1979년 2월: 연세대학교 정치외교학과 졸업
1981년 8월: 연세대학교 경제학과 석사
1986년 5월: University of S. Carolina, MBA
1990년 12월: Texas A&M University, Ph.D. in MIS
1995년 3월~2014년 8월: 중앙대학교 정보시스템학과 교수
2014년 9월~현재: 중앙대학교 산업보안학과 교수
〈관심분야〉 정보보호 거버넌스, 정보보호 관리, 디지털 비즈니스