개인정보보호법 개정에 따른 개인정보 시스템에서의 이상 징후 탐지 기법

지승우 [1], 유헌창 [2]
[1] 고려대학교 소프트웨어보안학과 석사과정
[2] 고려대학교 컴퓨터학과 교수

wltmddn8410@gmail.com, yuhc@korea.ac.kr

Anomaly Detection Technique in Personal Information Systems Following Amendments to the Personal Information Protection Act

Seungwoo Chi[1], Heonchang Yu[2]
[1]Dept. of Software Security, Korea University
[2]Dept. of Computer Science & Engineering, Korea University

요 약

2023 년 개정된 개인정보보호법(PIPA)의 요구사항을 기반으로 표준화된 로그 형식을 활용하여, 관계형 데이터베이스(RDB)에서 리스크(risk) 기반의 모델을 적용함으로써 최근 증가하고 있는 대내외 개인정보 유출을 범용적으로 탐지할 수 있도록 하였다. 이를 통해, 중소기업도 대기업과 유사한 수준의 보안 대응 능력을 갖출 수 있도록 지원하고, 로그 데이터를 활용하여 비정상적인 접근 패턴을 식별할 수 있도록 하며, 최소한의 리소스로도 효율적인 모니터링 체계를 구축할 수 있는 방법론을 구현하고자 한다. 이러한 접근 방식은 중소기업이 법적 요구를 충족할 뿐만 아니라, 비용 효율적인 보안을 통해 대기업 수준의 보안 강화 효과를 얻을 수 있도록 모델을 제시한다.

1. 서론

2023 년 개정된 개인정보보호법 [1]은 국내 기업이 개인정보 처리시스템에 대한 보안 조치를 강화하는 것에 큰 변화를 가져왔다. 그중 권한 로그, 로그인 로그, 접속 로그와 같은 시스템의 어플리케이션 로그들의 기록 및 관리가 법적으로 의무화되었으며, 이러한 로그는 모든 기업들이 최소 3 년간 보관하도록 라이프사이클을 정의하였다. 이를 통해 기업은 개인정보 처리시스템으로의 비인가된 접근이나 잠재적인 데이터 유출 사고에 있어 신속하게 식별하고 대응할 수 있게 되었다. 이러한 입법적 변화는 증가하는 사이버 위협 시대에서 데이터 보호의 중요성을 반영한 것이다.

또한 IT 를 기반으로 하는 중소기업들이 증가하면서 개인정보를 중소기업에서 취급하는 사례가 증가하였지만 자본규모가 적은 중소기업은 실질적으로 보안 장비를 도입하기 어려운 환경에 처해져 있으며, 중소기업의 입장에서 적용하는 개인정보보호법의 법적 요구사항도 데이터 라이프사이클과 유출 신고 등에 초점을 맞추고 있어, 자체적으로 내부 모니터링에 대한 구체적인 가이드 조차 부족한 상황이다. 그에 따라 최근 보안 관리가 취약한 중소기업을 대상으로 공격 사례가 증가하고 있으며, 내부자 유출 또한 점검되지 못하는 보안의 사각지대에 놓인 경우가 많다. 이렇게 장벽이 낮은 곳을 통해 유출된 개인정보는 공격자들이 크리덴셜 스터핑(Credential Stuffing)기법으로 보안이 잘 갖춰져 있는 다양한 서비스에 무단으로 접근할 수 있는 틈을 만들 수 있다. 이는 보안에 취약한 기업에서 시작된 유출이 더 큰 피해로 이어질 수 있음을 의미한다.

이를 해결하기 위하여 표준화된 로그 데이터들을 활용하여 중소기업에서도 이상 징후를 탐지하는 방법을 데이터베이스를 기반으로 리스크 기반의 모델을 적용하여 별도의 SIEM(Security Information and Event Management)이나 ESM(Enterprise Security Mnagement)와 같은 로그 분석 시스템을 도입하지 않고도 비정상적

인 접근 패턴과 내부 위협을 효과적으로 탐지할 수 있는 방안을 연구한다.

따라서 본 연구에서는 중소기업들이 비용 부담 없이 기존 시스템에서 수집된 표준화된 로그 데이터를 활용해 비정상적인 접근 및 내부 위협을 실시간으로 모니터링할 수 있는 체계를 제안하고, 이를 통해 중소기업도 대기업 수준의 보안 성능을 확보하도록 하는 것이 목적이다.

2. 관련연구

2.1 기존 보안 솔루션의 한계

기존의 범용적인 보안 솔루션들은 주로 외부의 위협을 차단하고 방어하기 위한 목적으로 설계되어 사용되고 있다. 방화벽(Firewall), 침입 방지 시스템(IPS), 웹 방화벽(WAF) 등 네트워크 단에서의 보안장비들은 외부 공격을 차단하는데 효과적이지만, 내부에서 발생하는 비인가 접근이나 개인정보 등의 유출 시도 같은 내부자 행위에 대해서는 충분한 대응하지 못한다. 그에 대응하기위해 내부 유출 방지를 위한 DLP(Data Loss Prevention) 와 같은 솔루션들이 사용되지만, 이는 주로 엔드포인트와 네트워크 상에서의 데이터 이동을 모니터링하고 차단하는 데 초점을 맞추고 있어, 로그 데이터를 기반으로 한 접근 패턴 분석이나 비정상적인 접근 탐지와는 거리가 있다. 보안 솔루션의 로그를 활용하여 사용자 이상 징후 탐지를 위해 기업 내부의 다양한 보안 솔루션과 관련 시스템에서 발생하는 사용자 행위 로그를 분석하고, 이를 통해 위협 행위 시나리오를 도출하는 연구[6] 또한 보안 솔루션과 시스템마다 발생하는 로그의 양이 방대하여 실시간으로 탐지 시스템과 연계하는 데 어려움이 있었다. 보안 솔루션에서 수집된 로그를 기반으로 내부정보유출에 대한 시나리오를 설계하는 연구 중 문서 보안솔루션(DRM), 유해 사이트 차단 시스템과 같은 보안 솔루션에서 수집된 로그를 활용하여 시나리오를 개발하고, 이를 기반으로 로그 분석 방안을 제시한 연구도 업무 상 발생할 수 있는 예외적인 상황을 충분히 고려하지 않아, 오탐에 대한 문제를 해결하지 못하였다[7]. 이와 같은 기존의 연구들은 내부정보유출 탐지를 위한 시나리오 기반 접근법을 제시하였으나, 다양한 이기종 간의 로그로 인한 탐지 임계치 설정의 부재, 오탐에 대한 해결책 부족 등의 문제를 보였다. 특히, 중소기업과 같은 소규모환경에서는 이를 적용하기 위해서는 자원과 비용이 부족한 경우가 많아, 그보다 더욱 실용적이고 효율적인 접근이 필요하다.

2.2 내부 위협 탐지 기술의 한계

현재 연구된 내부 위협 탐지 기술은 사용자 행동 분석(UBA: User Behavior Analytics), 엔드포인트 탐지 및 대응(EDR: Endpoint Detection and Response) 등의 솔루션을 통해 주로 이루어진다. 이러한 성격의 솔루션들은 사용자의 행동 패턴을 분석하여 비정상적인 행위를 탐지하는 데 효과적이다. 예를 들어, PC 보안 솔루션 로그를 활용하여 사용자 행동을 분석하고, 이를 기반으로 개인정보 유출을 예방하기 위한 보안 정책을 제시한 연구[3]는 대부분 고가의 전문 보안 시스템에 의존하므로, 중소기업에서는 도입하기 어려운 실정이다.

ERP 시스템의 사용자 로그를 활용하여 이상 행위를 탐지하는 연구[8] 또한 사용자 행동과 비정상 행위 간의 관계를 통계적으로 분석하였지만, 사용자 개개인들의 특성과 비정상 행위 간의 유의미한 관계를 입증하는 데 성공했으나, 비정상 행위로의 판단에 영향을 미칠 수 있는 많은 변수들이 충분히 고려하지 않아 높은 오탐률이 발생하는 문제를 보였다. 그 외 보안 위협 탐지를 위해 4 단계의 시나리오 개발 가이드라인을 제안 및 기업 보안 시스템에서 발생하는 로그의 상관분석을 통해 위협 패턴을 발견하고 보안 시나리오를 설계 연구도 다양한 이기종 보안 솔루션에서 수집된 로그를 수동으로 통합하여 분석해야 하는 비효율성이 있어 범용적으로 다루기 어려운 한계가 있다이와 같은 연구들은 내부 위협 탐지를 위한 다양한 접근 방식을 제안하고 있으나, 오탐 문제, 다양한 이기종의 로그 등의 문제로 인해 중소기업에서 적용하기에 실효성에서 한계를 보인다.

3. 리스크 기반 모델 설계

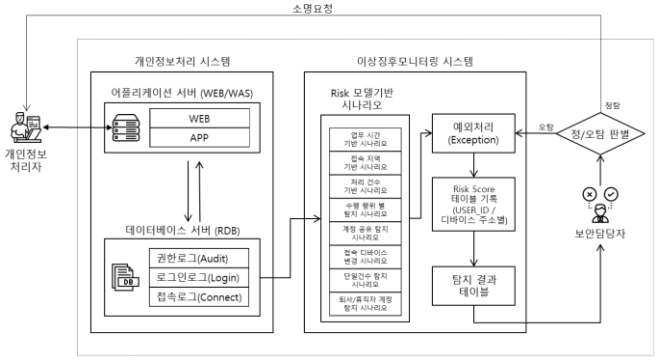
본 연구에서 제안하는 리스크 기반 모델은 개인정보 처리 시스템에서 기록하는 로그를 기반으로 분석하여 비정상적인 행위를 탐지하는 것에 목적을 둔다. 개인정보 유출 시도는 특정한 패턴이나 이상 징후를 동반하기 때문에, 로그 데이터를 통해 이를 분석하고 탐지하는 것이 가능하다. 명확한 이상행위를 하지 않더라도 의심스러운 행위를 리스크형태로 누적 기록하여, 위협을 가시화한다. 개정된 개인정보보호법에 맞춰 모든 개인정보 처리 시스템에서 권한 로그, 로그인 로그, 접속 로그를 의무적으로 기록/보관을 함에 있어 표준화된 원천데이터를 활용하여 비정상적인 행위를 추가적인 전처리를 요구하지 않는다. 리스크 기반 모델은 기본적으로 실시간 로그 분석을 통해 사용자의 접근 패턴, 위치, 시간대, 접근 빈도 등을 평가하여 리스크점수를 산출한다. 이 점수를 바탕으로 임계값을 초과하는 행위를 이상 징후로 간주하고 경고를 발생시킨다.

이러한 모델은 별도의 고가 장비 없이도 대부분의

개인정보 처리 시스템에서 적용할 수 있는 범용성을 제공하며, 기존의 SIEM 이나 EDR 솔루션에 비해 설정과 유지 관리가 간단하여 범용적으로 쉽게 적용할 수 있다.

3.1 전체 시스템 구성도

리스크 기반 모델을 사용한 탐지 시스템은 (그림 1)과 같이 설계되었다. 크게 개인정보 처리시스템과 이상징후모니터링 시스템 두 가지의 시스템으로 나뉘며, 이상징후모니터링 시스템은 개인정보 처리 시스템으로부터 로그 정보를 받게 된다.



(그림 1) 이상징후 탐지를 위한 전체 시스템 구성도

(1) 데이터 수집

연구에서 사용하는 개인정보 처리 시스템은 표준에 맞는 권한로그, 로그인로그, 접속로그를 DB 에 남기며 이를 분석 가능한 형태로 전처리하는 과정을 수행하지 않는다. 로그 데이터는 개인정보 처리 시스템에서 실시간으로 수집되며, 수집된 데이터는 데이터베이스에 저장된다. 수집되는 로그 데이터에는 사용자 ID, 타임스탬프, IP 주소, 이벤트 유형(예: 로그인 성공/실패, 권한 부여/삭제, 파일 접근 등)이 포함된다. 이러한 로그 데이터는 표준에 맞춰 시스템에서 정상적으로 수집을 하는 것을 가정한다. 적재된 로그데이터는 이후 이상징후모니터링 시스템으로부터 호출되어 리스크 기반 이상 징후 탐지 모델의 입력 데이터로 사용된다.

(2) 이상징후 탐지 시나리오 설계

이상징후모니터링 시스템으로 로그가 호출된 후 로그 데이터를 기반으로 사용자계정 혹은 접속 디바이스로 로그인 시간, 접근 위치, 데이터 조회 빈도 등을 모델에 적용하여, 이를 통해 이상행위에 대한 여러 시나리오를 적용한다. 예를 들어, 특정 사용자가 평소 근무 시간에 사내 네트워크를 통해 로그인하는 것이 일반적인 패턴이라면, 이와 다른 시간대나 네트워크에서의 접근 시도는 비정상적인 행위로 간주될 수 있다. 이후, 리스크 점수를 산출하는 단계에서는 사용자의 접근 빈도, 접근 시간대, 접근 위치, 권한 변경 내역 등의 요소를 종합적으로 고려하여 각 접근 행위의 위험도를 평가한다. 리스크 점수는 특정 사용자의 행위가 정상 범위에서 벗어난 정도를 수치화한 값으로, 비정상적인 행위일수록 높은 점수가 부여된다.

<표 1> 개인정보 이상징후 탐지 시나리오

항목	Score 기준	하위 항목
업무 시간 외 탐지	업무시간으로부터 시간별 리스크 score 차등 부여	행위 별 score 차등 부여 X 조회 건수
접속 지역 별 탐지	내부대역 혹은 사무실 대역 외 접속 Score 부여	국가 및 권역에 따라 차등 부여 계정 동시 접속에 대한 탐지도 추가가능
개인정보 처리 건수 별 탐지	평균 처리 건수 별로 그룹을 나눠 관리 총 조회 건수 제한으로 관리	사번별 일단위 표준편차 3sigma 일 경우 탐지 혹은 3개월 평균 대비 일정 임계치 이상 탐지
수행 행위 별 탐지	임계치 혹은 평균 이상의 다운로드, 삭제, 수정 등 행위를 탐지	행위 별 차등 score 부여
사용자 계정 공유 탐지	다수의 계정 사용 탐지	동일 시간, 동일 계정, 다른 접속지 기준 탐지
IP / MAC 변경	접속 디바이스에 대한 변화 탐지	-
단일 건수 탐지	호기심에서 비롯된 단순 조회	사용자 기준 평시 평균 대비 단일 건수 조회
퇴사자/휴직자 계정 사용	Score 만점 부여	명확한 부적격 사유

예를 들어, 평소에 접근 권한이 없는 데이터베이스에 접근을 시도하거나, 평소에 사용하지 않는 IP 주소에서 민감한 데이터에 접근하는 경우 리스크 점수가 높아진다. 탐지된 리스크 점수가 일정 임계값을 초과하면 이를 이상 징후로 간주하고 보안 담당자에게 경고 메시지를 전달한다. 이러한 탐지는 규칙 기반 탐지와 리스크 기반 탐지를 결합하여 이루어진다. 예를 들어, 퇴사자/휴직자 계정 사용, 국내 대상 시스템일 경우 해외접근 IP 탐지 등과 같이 확실한 비이상적인 동작에 대한탐지는 사전에 정의된 규칙을 기반으로 탐지되고, 명확히 식별이 불가능한 비정상적인 접근 패턴은 리스크 점수를 통해 탐지된다.

(3) 임계치 설정

각 시나리오 별로 리스크를 부여하는 기준에 대한 임계치를 설정 해야하며, 임계치는 크게 정적과 동적 두 가지의 방식으로 설정한다. 정적 임계치는 단일 건수나 접속 지역 등 기준값이 고정되는 시나리오에서 쓰인다. 그와 반대로 사용자의 행위나 트래픽 변동 등으로 패턴에 따라 동적으로 조정되는 임계치를 말한다. 사용자 별로 장 기간 개인정보 조회 평균을 구하고 직전일에 구한 평균과 차이를 가지고 임계치를 설정한다. 경우에 따라 여건이 될 경우 표준편차를 구하여 3 시그마 이상의 값을 탐지할 수 있다. 임계치는 각 시스템 별 환경이나 서비스 대상, 빈도 등 영향을 받아 시스템 특성에 맞게 조정이 필요하다.

(4) 예외 처리

이상 탐지 시스템에서 정상적인 동작과 명확히 밝혀진 오탐을 구분하고 관리하는 것은 모니터링의 품질을 유지하고, 보안 담당자가 실제 위협에 집중할 수 있도록 돕는 중요한 과정이다. 예외 처리는 여러 단계로 이루어질 수 있다. 먼저, 경고가 발생한 이벤트 중 반복적으로 정상적인 행위로 확인되는 경우를 식별하여 예외 상황으로 분류하고, 시스템 업데이트 등의 작업이 있을 경우 예외처리를 한다. 예외처리의 방식은 해당 USER_ID 혹은 접속 디바이스를 기준으로 스코어 점수를 0 으로 만들거나 특정 값을 빼는 식으로 예외를 진행한다.

<표 2> 성능 평가

탐지방법	탐지 건수	탐지율	비고
SIEM 시나리오를 통한 탐지	138건	98.57%	ML 적용 탐지 (이상 탐지 : 표준편차 3시그마 탐지 적용)
Risk score 기반 RDB사용 탐지	130건	92.85%	리스크 스코어 30 이상 임의 적용

- 총 정탐 건수 : 140건 / 동일 행위에 대한 중복 이벤트 제외
- 리스크 스코어 30 이상 값으로 지정하였을 경우로 산정

4. 성능 평가

실제 기업에서 발생한 동일한 로그 데이터를 사용하여 기존 SIEM 장비에서 탐지된 정탐 결과와 리스크 기반 이상 징후 탐지 모델의 결과를 비교하였다. 총 140 개의 정탐 건수 중 SIEM 을 통한 탐지는 기계학습(ML)을 적용한 시나리오로 총 138 건을 탐지하여 98.57%의 높은 탐지율을 보였다. 이는 SIEM 시스템이 정교한 탐지 알고리즘을 사용하여 다양한 비정상 행위를 정확히 식별할 수 있음을 나타낸다.그에 있어 연구에서 제안한 리스크 기반 모델은 140 건 중 130 건을 탐지하여 92.85%의 탐지율을 기록하였다. 비록 SIEM 보다 낮은 탐지율을 보였으나, 리스크 스코어 임계값을 30 으로 설정한 단순한 규칙만으로 높은 탐지율을 보였다는 점과 고가의 장비 없이도 효과적인 보안 모니터링을 수행할 수 있는 가능성에서 의미가 있다.

5. 결론

본 연구에서 제안한 모델은 중소기업이 별도의 고가의 보안 솔루션 없이도 기존의 로그 데이터를 활용하여 비정상적인 접근과 내부 위협을 실시간으로 모니터링할 수 있도록 제안되었다. 개정된 개인정보보호법을 준수하기 위해 쌓은 로그데이터를 활용하여 중소기업이 실제적인 보안 성능을 강화할 수 있는 효율적인 방안을 제시하였다. 더 나아가, 이러한 로그 기반 리스크 탐지 모델은 환경 변화에 유연하게 대응할 수 있으며, 오탐 및 미탐을 최소화하여 보안 모니터링의 신뢰성을 높일 수 있다.

향후 연구에서는 제안된 모델의 성능을 다양한 환경에서 검증하고, 로그 데이터의 정교한 분석을 통해 더 많은 변수와 조건을 반영한 정교한 임계치 설정 방안을 개발할 필요가 있다.

참고문헌

[1] 개인정보보호위원회, [시행 2023. 9. 22.] 개인정보의 안전성 확보조치 기준

[2] 한국인터넷진흥원, ISMS-P 인증기준 안내서(2023.11) 수정게시

[3] 채현탁, “Security policyproposals through PC security solutionlog analysis : prevention leakageof personal information” , pp. 961-968 한국정보보호학회논문지, Feb. 2015.

[4] 엄정호, “The QuantitativeEvaluation of a Level of InsiderActivity using SFI AnalysisTechniques,” 10(2), pp. 113-122 보안공학연구논문지, Apr.2013

[5] 이광우, 김승주.“Analysis of Trends in Digital Multifunction Device SecurityTechnology from the Viewpoint of Preventing and Protecting BusinessConfidential Information.”, 20(1). pp. 정보보호학회논문지, Feb. 2010

[6] 정귀영, “A Study on the Effective User Anomaly Detection Method through Integrated Security Log Analysis,” master’s thesis, Yonsei University, Feb. 2017.

[7] 박장수, 이임영, “Information Security : Log Analysis Method of Separate Security Solution using Single Data Leakage Scenario,” pp. 65-72 KIPS 정보처리학회논문지, Feb. 2015

[8] 김은선, “A Study on the Anomaly Detection using User Log : ERP System An Empirical Study,” master’s thesis, Korea University, Feb. 2015.

[9] 박성주, 임종인 “A study on the development of SRI(Security Risk Indicator)-based monitoring system to prevent the leakage of personally identifiable information”, pp. 637-644 정보보호학회논문지, June. 2012