감사데이터를 이용한 지능적인 이상행위 감지 기법에 관한 연구1)

송인수*, 이대성**, 김귀남*
*경기대학교 산업보안학과
**경기대학교 산업기술보호특화센터
e-mail:trumpetsis@naver.com

A Study on Abnormal Behavior Intelligent Detection Method Using Audit Data

In-Su Song*, Dae-Sung Lee**, Gui-Nam Kim*
*Dept of Industry Security, Kyoung-gi University
**Center for Industry Security, Kyoung-gi University

요 약

정보통신 기술과 저장 매체의 발전으로 많은 분야에 편리함과 더불어 산업기밀유출사고의 위험이 늘어나고 있다. 보안사고 중 80% 이상이 인적 보안 유출 이였으며 현직 직원의 유출은 약 25%정도의 부분을 차지하고 있었다. 기존의 단순한 시스템 로그 정보를 이용한 사용자 감사기술, DRM을 이용한 데이터 보호기술방법 보다는 진보된 방법이 필요하다. 사용자 정보와 시스템 정보, 시스템 콜 정보 수집을 통한 구분된 감사데이터의 통계기법을 이용한 지능적인 이상행위 탐지 기법을 제시한다.

1. 서론

인터넷과 저장 매체의 발전으로 인하여 다양한 방법으로 기업 및 기관의 주요 산업기밀 정보가 유출되고 시스템이 위협을 받고 있다. 위협의 대부분은 외부에 의한 소행보다는 내부사용자에 의한 소행이 많으며 피해 규모는 천문학적인 숫자 이르고 있다. 내부자 보안에 관한 현재까지의 기술을 간략하게 살펴보면, 데이터 접근 사용자를 관리하는 시스템 접근 통제 기술과 데이터 유출 방지 기술, 시스템을 사용하는 사용자를 모니터링 할 수 있는 기술이다. 접근 통제 기술과 데이터 유출방지기술을 내부 사용자에 의한 보안 사고를 어느 정도는 예방 해주지만, 기밀정보 유출이나 공격을 계획한 공격자가 시스템을 관리하고 권한을 가진 사용자라면 유출을 막을 수 없게 된다.

최근에 주목 받고 방법이 데이터 손실 방지 기법이다. 이는 시스템을 사용하는 사용자의 모든 정보를 수집하여 분석하기 때문에 정확한 내부 위협을 미연에 방지 하고 차단의 장점이 있지만 많은 자원과 인력이 요구되어 대기업 및 주요 기관을 중심으로만 일부 구축되고 있는 실정이다. 또한 시그니처를 이용한 단순 감사 데이터 분석은 다양한 우회기법을 이용하여 우회할 수 있기 때문에 사용자 행위의 정확한 분석과 감사의 어려움이 있다. 따라서 이런 문제를 해결하기 위해 사용자 정보와 시스템 정보 및 시스템 콜 정보를 종합적으로 수집하여 감사데이터를 생성하고 확률 기법을 이용해 지능적인 이상행위 탐지 방법을 제시한다.

1) 본 연구는 지식경제부 지역혁신센터사업인 산업기술보호특화센터 지원으로 수행되었음

2. 관련연구

2.1. 기존의 정보 유출 방지 기술

현재 일부 기업이나 기관에서 사용하고 있는 산업기밀 정보 유출 방지를 위한 사용자 감사 기술(DLP)들은 사용하는 사용자의 모든 정보를 수집하여 분석하기 때문에 정확한 내부의 위협을 방지하고 차단할 수 있는 장점이 있지만. 사용자 정보를 바탕으로 감사데이터를 만들고 분석해야 하기 때문에 많은 자원과 인력이 요구된다. 또한 시그니처 기반의 분석은 다양한 우회기법을 이용하여 우회할 수 있기 때문에 사용자 행위의 감사에 어려움이 있다.

2.2. DRM을 이용한 데이터 보호 기술

내부자에 의한 정보 유출은 저장매체인 USB메모리나 외장 HDD, CD등을 이용하거나 PDA, FTP, P2P 를 이용하여 유출 되어 지고 있다. 이러한 정보 유출을 방지하는 DRM 기술[1]은 전직의 전산직원이 관리영역과 비관리 영역으로 명확히 구분되도록 관리 대상을 식별하고, 관리영역 내의 정보 체계에서 비관리 영역으로의 정보체계로 자료 흐름을 통제하는 구조이다. 민감한 데이터를 보호하는데 유용한 방법으로 기업이나 기관에서 사용 중에 있지만 전적으로 수동 작업에 의존해야 하고 대규모로 적용하기가 힘든 면이 있다.

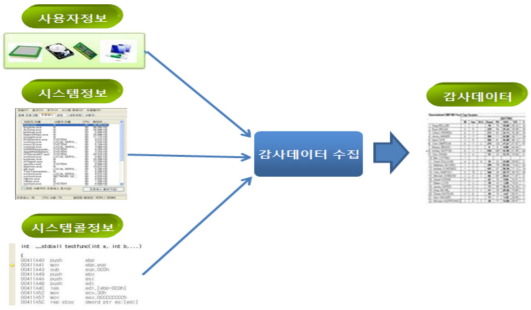
2.3. Native API 수집 방법

Windows 운영제체의 커널모드에서 동작하는 system service를 Native API라고 한다. Windows 운영체제에서

Native API를 수집하기 위해서는 커널 후킹 기법[2]이 필요하다. 커널 후킹 기법에는 IDT를 변조하여 Native API를 얻는 기법, system call 번호 순서대로 해당 함수의 메모리를 가지고 있는 SSDT 를 이용한 기법, 커널 레벨에서 동작하는 디바이스 드라이버의 Major Function Table을 이용한 기법 등이 있다.

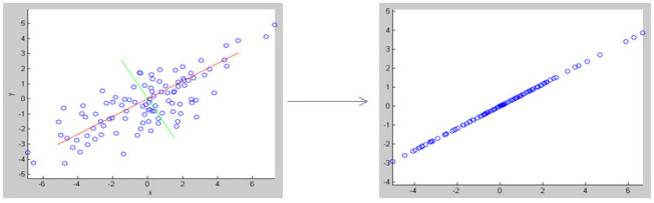
3. 제안 방법

본 연구에서는 시스템을 사용하는 사용자 정보 및 시스템 정보, 시스템 콜 정보를 수집하여 효과적인 감사데이터를 생성하고 SVM 알고리즘을 이용하여 이상행위 사용자를 분류 하는 방법을 제시한다.



[그림 1] 감사 데이터 수집

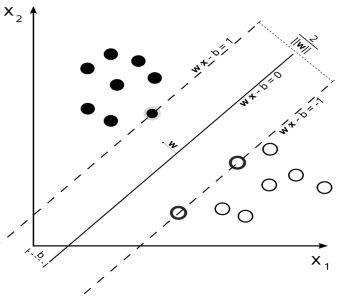
시스템 및 사용자, 시스템 콜 수집으로 얻은 데이터는 다양하기 때문에 많은 특징이 다차원적으로 발생하게 된다. 특징이 많으면 잡음 특징들까지 포함되어 잘못된 분류와 판단을 하게 되고 패턴 분류기에 의한 학습과 인식 속도가 느려진다. 또한 모델링에 필요한 학습 집합의 크기가 커지게 된다.



[그림 2] 주성분 분석을 통한 차원 축소

수집된 데이터를 빠르고 효율적으로 처리하기 위해 데이터의 높은 차원을 낮은 차원으로 축소 시켜야 한다. 주성분분석은 다차원 특징 벡터로 이루어진 데이터에 대한 높은 차원에서의 정보를 유지하면서 낮은 차원으로 차원을 축소 시키는 다변량 데이터 처리 방법 중 하나이다. 주성분 분석[3]은 한 개의 축으로 사상시켰을 때 그 분산이 가장 커지는 축이 첫 번째 좌표축으로 오고, 두 번째로 커지는 축이 두 번째, 다음 축이 세 번째 등과 같은 방법으로 차례로 놓이도록 새로운 좌표계로 데이터를 선형으로 변환하는 방법으로 다차원의 정보를 낮은 차원으로 축소 시키는 방법을 통해 특징을 효과적으로 추출한다.

SVM알고리즘[4]은 명백한 이론적 근거에 기반하므로 결과 해석이 용이하고 실제 응용에 있어서 인공신경망 수준의 높은 성과를 내는 장점이 증명되었다.



[그림 3] SVM을 이용한 사용자 감사 데이터 분석

SVM은 특징들을 Vector 공간에 Vector로 표시하고 선, 면 다차원으로 구성된 무수히 많은 hyperplane 중에서 두 범주를 효과적으로 분류할 수 있는 maximum hyperplane과 이때 maximum hyperplane과 가장 가까운 support vector를 구하여, vector들을 두 그룹으로 분류한다. SVM은 두 그룹간의 거리를 최대로 하는 가중치 값들을 정하는 것으로 다른 학습 알고리즘과 유사한 측면이 있다. 하지만, 비선형성이 보다 높은 차원에서 볼 때 선형성을 띈다고 하면 SVM을 이용하여 보다 신속하고 정확하게 두 그룹으로 분류 할 수 있다. 본 연구 개발에서도 주성분 분석으로 차원이 축소된 사용자 감사 데이터의 특징을 vector로 구성하고 정상 사용자 데이터를 이용하여 SVM의 maximum hyperplane을 만들며, 이를 이용하여 감사데이터에서 정상 행위와 비정상 행위 사용자를 분류한다.

4. 결론

본 연구 기술 개발은 기존의 패턴 매칭 및 시그니처 방식의 사용자 감사 분석 기술에서 벗어난 한 차원 지능화된 감사 기술을 개발하는 것이다. 수집된 데이터를 주성분 분석을 통한 전처리로 특징은 살리고 차원은 축소함으로 분류에 용이한 감사데이터를 만든다. SVM 알고리즘을 이용해 사용자 감사 데이터뿐만 아니라 사용자 행위도 감사할 수 있는 특징이 있다. 본 기술의 알고리즘은 악성코드 탐지 및 침입탐지 등 여러 보안 분야에 활용 될 수 있다. 현재 효율적인 전처리를 통한 감사데이터 확보가 필요하며 확보 된다면 더 효율적이고 명확한 이상행위 탐지를 할 수 있을 것이다.

참고문헌

[1] 이창열 DRM 표준 기술동향, TTA저널 제 82호 91~94
[2] 권오철, 배성재, 조재익, 문종섭, “Native API 빈도 기반의 퍼지군집화를 이용한 악성코드 재그룹화 기법연구”, 정보통신학회 논문지 제18권 제6호, pp. 117, page, Dec 2008.
[3] 김한성, 권영희, 차성덕, “SVM 기반의 효율적인 신분위장기법 탐지”, 정보통신학회 논문지 제 13권 제5호, pp. 117, page, 97 ~ 100 Oct.2003.
[4] I.T Jolliffe, Springer Series in statistics, "Principal Component Analysis second Edition", Springer Ltd. 2002