JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 23, No. 6, pp. 1159-1166, Jun. 2022

Check for updates

# LSTM Autoencoder 기반 내부자 데이터 유출 징후 탐지

김 서 준[1] · 손 태 식[2*]
[1]아주대학교 정보통신대학원 정보통신공학과 석사과정
[2*]아주대학교 정보통신대학원 사이버보안학과 교수

# LSTM Autoencoder-Based Insider Data Leak Detection

**Seo-Jun Kim[1] · Tae-Shik Shon[2*]**

[1]Master's Course, Department of Information and Communication Engineering, Ajou University, Korea
[2*]Professor, Department of Cyber Security, Ajou University, Korea

**[요 약]**

오늘 날 기업들은 내부 중요 데이터를 지키기 위해 많은 노력을 하고 있으나, 최근 지속적으로 기업 내 데이터 유출 사고가 발생하고 있다. 특히 악의적인 목적을 가진 내부자는 기업의 시스템에 대한 정보와 권한을 가지고 있기 때문에 보안에 가장 심각한 위협으로 부각되고 있다. 본 논문에서는 내부자에 의한 데이터 유출 징후 탐지를 위한 머신러닝 모델에 관한 연구를 진행하고자 한다. 긴 시간 동안 정보를 기억하여 시계열과 시퀀스 데이터 처리에 용이한 LSTM 알고리즘과 Autoencoder를 결합한 LSTM Autoencoder에 내부자의 직급과 5가지 성격 특성 요소에 따라 패널티를 부가하여 탐지율을 향상시킨 모델을 제안하여 시험을 진행하였고, 이 모델이 내부자에 의한 데이터 유출 징후 탐지 분야에서 실효성이 있는지 검증하였다.

**[Abstract]**

Today, companies are making great efforts to protect important internal data, but recently, there have been continuous data leakage incidents within the company. In particular, insiders with malicious purposes are emerging as the most serious threat to security because they have information and authority over the company's system. In this paper, we intend to conduct a study on a machine learning model for detecting signs of data leakage by insiders. We experiment with LSTM Autoencoder combining LSTM algorithms and Autoencoders that are easy to process time-series and sequence data by remembering long-time information, and we propose a model that improves detection rate by adding penalties according to insider rank and five personality characteristic factors.

**색인어 :** 내부자 위협, 데이터 유출, 이상 탐지, 머신러닝, LSTM Autoencoder

**Keyword :** Insider Threats, Data Leak, Anomaly Detection, Machine Learning, LSTM Autoencoder

**http://dx.doi.org/10.9728/dcs.2022.23.6.1159**



**Received** 07 May 2022; **Revised** 08 June 2022
**Accepted** 08 June 2022

***Corresponding Author; Tae-Shik Shon**

**Tel:** +82-31-219-3321
**E-mail:** tsshon@ajou.ac.kr

## Ⅰ. 서 론

오늘날 기업들은 내부 중요 데이터를 지키기 위해 정보보호 시스템 구축 등 많은 노력을 하고 있다. 하지만 최근 몇 년 동안 의료 산업, 금융, 제약, 미디어 등 다양한 산업에서 지속적으로 개인정보 유출 및 기술 유출 등 기업 중요 데이터 유출이 발생되고 있다. 기업 내 데이터 유출 사고가 일어날 경우 금전적 피해와 신뢰도 하락, 기업의 이미지 실추 등 막대한 피해를 입을 수 있다. IBM 시큐리티에서는 매년 전 세계 500개 이상의 기업 및 조직의 실제 데이터 유출 사례를 분석하여 발표하고 있다. 분석 보고서에 따르면 2020년 악의적인 공격을 통한 데이터 유출 사례 중 악의적 내부자 소행은 7%에 해당하며, 2021년은 8%로 1% 상승하였다. 2021년 한국 기업은 데이터 유출 사고로 평균 368만 달러(약 44억)의 손실을 입었으며, 2020년 대비 56만 달러 상승하였다. 또한, 보안 AI와 자동화를 활용하는 기업에서 신속하게 유출을 탐지하여 손실을 최소화할 수 있었다. 개인의 부주의와 기업의 관리적 부실로 데이터 유출이 발생되고 있다. 특히 악의적 목적을 가진 내부자는 기업의 정보나 네트워크 시스템, 시스템에 접근 가능한 방법 및 권한을 가지고 있기 때문에 보안에 가장 심각한 위협으로 부각되고 있다. 기업들은 데이터 유출을 보호하기 위해 데이터유출방지(DLP), 지능형지속공격(APT) 대응 등 정보보호체계 구축에 많은 투자하였으나, 내부자 위협으로 부터 데이터 보호에는 어려움을 겪었다. 최근 머신러닝을 이용한 내부자에 의한 데이터 유출 징후 탐지 방법에 대한 연구가 꾸준히 이루어지면서 내부자의 이상 징후 탐지를 통해 내부자 위협에 대응하고 있다.

본 논문의 연구에서는 미국 카네기멜론 대학에서 내부자위협연구를 위해 제공하고 있는 Insider Threat Test Dataset을 이용하여 데이터 유출 징후 탐지에 적합한 머신러닝 알고리즘인 LSTM Autoencoder를 통해 학습하고 내부자의 직급, 5가지 성격 특성 요소를 적용한 모델을 이용하여 내부자에 의한 데이터 유출 징후 탐지 분야에서 실효성을 검증해본다.

본 논문의 구성은 다음과 같다. 2장에서 기존 내부자에 의한 데이터 유출 탐지에 관한 연구에 대해 소개하고 3장에서 본 논문에서 제안하는 데이터 유출 징후 탐지 모델을 서술한다. 4장에서는 실험에 사용할 데이터 소개와 전처리과정, LSTM Autoencoder를 직접 구현하여 유효성을 검증한다. 5장에서는 본 논문에 대한 결론과 한계, 향후 연구에 대한 방향을 제시한다.

## Ⅱ. 관련연구

### 2-1 기존 관련연구

내부자에 의한 데이터 유출 징후 탐지에 대한 사회적 관심이 많아지면서 머신러닝 알고리즘을 활용한 다양한 연구가 수행되고 있다. Dongwook Ha는 카네기멜론 대학의 CERT 데이터를 RNN(Recurrent neural network) 알고리즘으로 구성한 Autoencoder 모델에 적용하여 내부자 위협 탐지에 대한 연구를 수행하였다.[1] 기존 loss 값 외에도 USB의 사용횟수가 평소 대비 많을 경우 추가적인 패널티를 부여하여 탐지율을 향상시켰다. Minhae Jang은 카네기 멜론 대학의 CERT 데이터를 활용한 RNN Autoencoder 모델을 개선하기 위해 seq2seq를 사용한 RNN Autoencoder 모델을 구현하는 연구를 수행하였다.[2] 추가적으로 HBOS 모델을 사용하여 CERT 데이터의 분류 난이도를 검증하였고 Attention 기법을 적용시켜 오탐율을 감소시켰다. Jeongmin Lee 등은 카네기멜론 대학의 CERT 데이터를 LSTM(Long short term memory) 알고리즘으로 구성한 Autoencoder 모델에 적용하여 내부자 위협에 대한 연구를 진행하고 있다.[3] Yangwoo Lee 등은 카네기멜론 대학의 CERT 데이터를 비지도 학습 방법인 Clustering 모델 중 하나로 각 군집의 평균을 구해 K개의 군집을 형성하는 K-Means Clustering 알고리즘과 Bisecting K-Means Clustering 알고리즘을 적용하여 정상행위 군집에 속하지 않는 데이터를 이상행위로 탐지하는 내부자 이상행위 탐지 기법 연구를 수행하였다.[4] Hyunsoo Kim은 국내 조직 환경을 고려하여 실제 조직의 DRM, DLP, NAC, 보안 USB 등의 보안시스템 및 그룹웨어 로그를 활용하여 내부자 정보 유출 탐지 기법에 대한 연구를 수행하였다.[5] 해당 연구에서는 데이터 내 내부자 행위 특징을 18가지로 세분화하여 HMM(Hidden markov model)을 통해 내부자 정보 유출 행위 패턴 특징을 추출하여 탐지했다.

데이터 유출 징후 탐지에 대한 연구 외에도 다양한 연구가 진행되고 있다. Sangmok Kim은 가상의 네트워크 트래픽 데이터를 수집하여 소규모 조직을 위해 오픈소스를 활용한 이상 징후 탐지기법에 대한 연구를 수행하였다.[6] 해당 연구에서는 시스코사의 Netflow를 이용하여 네트워크 패킷을 수집하고 LSTM 알고리즘과 평균값에 근거한 군집화 알고리즘으로 네트워크 이상 데이터를 탐지하여 DDoS 공격을 탐지하는 모델을 만들었다. 또한, 내부자 정보 유출 행위를 사전에 탐지하기 위해 시나리오의 위험도 및 가중치에 기반을 둔 위협을 계산하는 모델을 제안하는 연구를 수행하였다.[7]

### 2-2 Long Short Term Memory(LSTM)와 Autoencoder

#### 1) Long Short Term Memory(LSTM)

인공 신경망 알고리즘의 한 종류로 유닛간의 연결이 순환적인 구조를 갖고 있는 RNN 기법의 하나이다. RNN의 기울기 소멸(Vanishing gradient) 문제를 해결하기 위해 Cell State 개념을 도입하여 과거의 데이터를 유지하면서 장기 의존 기간을 필요로 하는 학습을 수행하는 능력이 있다. LSTM은 긴 시간 동안의 정보를 기억하여 시계열 및 시퀀스 데이터 처리에 용이하다는 장점을 가지고 있지만, 연산 속도가 느리다는 단점이 존재한다.

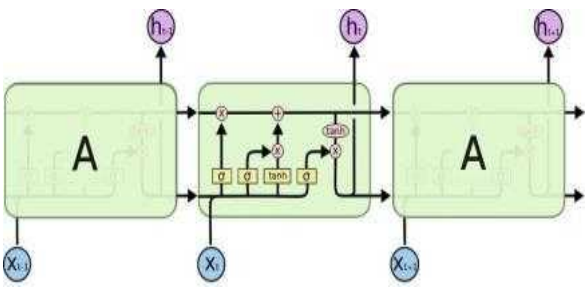


**그림 1.** LSTM 구조
**Fig. 1.** LSTM Structure

LSTM은 RNN처럼 체인과 같은 구조를 가지고 있지만 Hidden State만을 사용하는 것과 달리 Cell State라는 특별한 방식으로 정보를 주고받도록 설계되어 있다. Fig. 1.에서 LSTM의 구조를 볼 수 있듯이 Cell State에는 이전 Cell에서 넘어온 입력 데이터에 대해 과거의 정보 중 어느 것을 반영할 것인지 결정하는 Forget Gate, 현재 입력된 정보를 얼마나 반영할 것인지 결정하는 Input Gate, 다음 Cell에 전달할 데이터에 얼마나 반영할 것인지 결정하는 Output Gate로 구성되어 있다.

### 2) Autoencoder

입력과 출력을 동일하게 하며 답이 주어지지 않은 상태에서 스스로 학습하게 하는 비지도 학습 방법 중 하나이다. 정상 데이터를 Autoencoder를 통해 원래의 데이터보다 작은 차원의 공간으로 압축하고 원래 데이터로 복구하는 과정을 통해 학습시킨다. Fig. 2.의 Autoencoder 구조에서 볼 수 있듯이 입력 데이터가 Encoder를 거쳐 차원이 축소되고, 축소된 입력 데이터가 Decoder를 거쳐 기존 입력 데이터와 동일한 출력 데이터로 변환 시킨다. 이와 같이 Autoencoder의 가장 큰 특징은 입력 데이터와 출력 데이터가 같아야 하는 형태로 입력 데이터와 출력 데이터의 거리인 Loss Function을 최소화하는 것이다. 이러한 특징을 이용하여 이상 탐지, 이미지 복구, 압축 등의 다양한 영역에서 활용되고 있다.

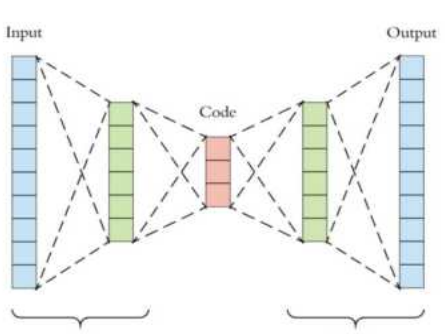


**그림 2.** Autoencoder 구조
**Fig. 2.** Autoencoder Structure

### 2-3 5가지 성격 특성 요소(big five personality traits)

심리학에서 경험적인 조사와 연구를 통해 정립한 성격 특성의 5가지 주요한 요소를 말하며 경험에 대한 개방성(Openness to experience), 성실성(Conscientiousness), 외향성(Extraversion), 우호성(Agreeableness), 신경성(Neuroticism)으로 구성되어 있다. 경험에 대한 개방성은 호기심이 많고 새로운 경험에 대한 열린 자세를 말한다. 성실성은 목표를 성취하기 위해 성실하게 노력하는 성향으로 높을수록 책임감 있고 자기통제와 조절을 잘한다. 외향성은 다른 사람과의 사교적인 활력을 추구하는 성향으로 낯선 사람에게도 스스럼없이 다가간다. 우호성은 타인에게 반항적이지 않고 협조적인 태도를 보이는 성향으로 따듯하고 공감적인 성격을 가진다. 신경성은 분노, 우울함, 불안감과 같은 불쾌한 정서를 쉽게 느끼는 성향으로 정서적으로 예민하고 불안정하다.

## Ⅲ. 제안하는 데이터 유출 징후 탐지 모델

조직에서는 강제 접근 통제(MAC; Mandatory Access Control)를 통해 정해진 직급에 따라 접근 할 수 있는 데이터의 범위가 한정되어 있다. 높은 직급의 경우 높은 보안 등급을 가지고 있어 많은 데이터를 접근 할 수 있는 대신 데이터 유출 징후에 관심을 두어야 하는 대상이다.

2014년 국가 사이버 보안 및 통신 통합 센터(NCCIC; National Cybersecurity and Communications Integraion Center)에서 발표한 내부자 위협 관련 보고서[8]에 따르면 내부자는 데이터 유출 전에 몇 가지 행동지표를 나타낸다고 한다. 휴가, 병가와 같이 부재중에 조직 네트워크에 원격으로 접근하거나 승인 없이 업무 시간 외 근무를 하고 시간 외 근무, 주말 근무 등 비정상적인 업무 일정에 적극적이게 된다. 또한, 약물, 알코올 중독, 재정적 어려움, 도박 등 열악한 정신건강이나 적대적인 행동과 같은 취약성 징후를 유발하고 갑작스런 해외여행을 가거나 결근을 하는 등 경고 신호를 발생하기도 한다. 이러한 행동지표는 내부자들의 5가지 성격 특성 요소를 통해 확인 할 수 있다. 내부자가 데이터 유출을 하게 되는 상황이 발생하면 일에 대한 목표가 사라지게 돼서 성실성이 낮아지게 되고 다른 사람들과의 사교성이 떨어져 외향성이 낮아지게 되며 다른 사람들에게 비협조적이게 돼서 우호성이 낮아진다. 또한, 불안감이 상승하여 신경성이 높아지게 된다.

본 논문에서는 긴 시간 동안의 정보를 기억하여 시계열과 시퀀스 데이터 처리에 용이한 LSTM 알고리즘과 입력 값과 출력 값의 loss 값을 비교하여 이상탐지를 할 수 있는 Autoencoder를 결합하여 데이터 유출 징후 탐지를 위한 LSTM Autoencoder 모델을 구현하고 내부자의 직급과 5가지 성격 특성 요소에 따른 패널티를 부가하여 탐지율을 향상시킨 내부자에 의한 데이터 유출 징후 탐지 모델을 제안 하고자 한다.

## Ⅳ. 실험 및 분석

### 4-1 실험 Dataset 소개

미국 카네기멜론 대학의 소프트웨어 엔지니어링 연구소에 있는 CERT 내부자 위협 센터에서는 사이버 보안 문제를 해결하는데 전념하고 있다. 센터에서는 2001년부터 미국의 국방부, 국토안보부, 미국 비밀검찰국, 연방기관, 정보기관, 민간 업체, 학계 및 공급 업체와 협력하여 내부자 위협에 대해 연구하고 있다. 1000건 이상의 내부자 위협 사례를 수집하고 기술 및 행동 관점에서 검토하여 내부자 사건을 예방, 감지 및 대응하는데 사용할 수 있는 통제수단을 생성하며 악의적인 내부자 활동을 잘 이해하고 탐지할 수 있는 대응책을 수립하는 등 내부자 위협 분야에 기여하고 있다. CERT 내부자 위협 센터는 ExactData, LLC와 협력하고 DARPA I20의 후원을 받아 정상행위와 비정상행위로 구성되어 있는 가상의 내부자 위협 테스트 데이터들을 모아 Insider Threat Test Dataset를 제공하고 있다. 현재 Insider Threat Test Dataset은 r1버전부터 r6버전까지 나와 있다. Insider Threat Test Dataset 내 악의적인 행위를 통해 조직의 데이터를 유출 시키려는 악의적인 내부자가 수행하는 악성행위 시나리오 다섯 개를 Table 1.과 같이 정의하고 있다.

Insider Threat Test Dataset r4.2는 다른 버전의 Dataset보다 악의적인 내부자가 더 많이 구성되어 있기 때문에 실험 데이터로 가장 적합한 버전이다.

**표 1.** 데이터 셋에 대한 악의적인 행위 시나리오
**Table 1.** Malicious Behavior Scenarios for Dataset

| Scenario | Contents |
|---|---|
| Scenario 1 | An insider who has not previously used a removable drive or worked overtime works overtime or uploads data to the website (wikileaks.org) using a removable drive. |
| Scenario 2 | After hiring, insiders attempt to use more mobile drives to their competitors than ever before, surf the turnover-related websites, and leak data. |
| Scenario 3 | A system administrator who is dissatisfied with the organization downloads the keylogger and stores it on his supervisor's computer. It causes confusion by sending a large amount of mail by stealing the account of the boss collected through the keylogger. |
| Scenario 4 | An insider logs in to another employee's computer, finds files of interest, and sends them to him/her via mail. |
| Scenario 5 | A fired insider uploads files to dropbox, etc. for personal gain. |

**표 2.** 데이터 셋 파일의 구조
**Table 2.** Structure of Dataset Files

| File | Primary Field | Content |
|---|---|---|
| logon.csv | activity | Logon / Logoff |
| http.csv | url / content | Accessed url and page content |
| file.csv | filename / content | About moved file names and headers |
| device.csv | activity | Connect / Disconnect |
| email.csv | to / from / attachments | Who sent which email to whom |
| psychometric.csv | O / C / E / A / N | Big five personality traits of insiders |
| LDAP | functional unit / department / team | Insider's department, position information |

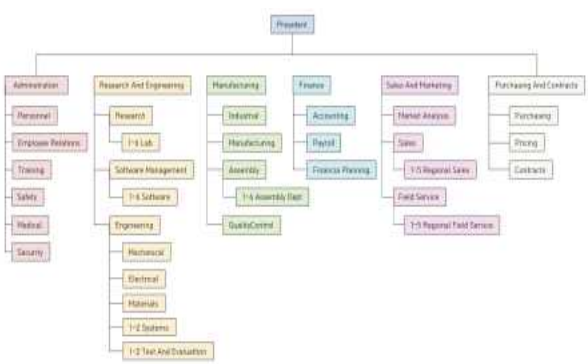

**그림 3.** Insider Threat Test Dataset의 조직도
**Fig. 3.** Organization chart of Insider Threat Test Dataset

총 1000명의 내부자 데이터 중 시나리오 1에 해당하는 내부자 30명, 시나리오 2에 해당하는 내부자 30명, 시나리오 3을 시도하는 내부자 10명으로 구성되어 있으며, 시나리오 4와 5는 r4.2 버전에서 해당되는 데이터가 존재하지 않는다. Dataset에는 내부자의 행위를 분류하여 logon.csv, http.csv, file.csv, device.csv, email.csv, psychometric.csv, LDAP 파일로 구성되어 있다. logon.csv 파일에는 컴퓨터 장치에서 로그온 및 로그온에 기반한 사용자 활동 로그, http.csv 파일에는 내부자가 웹페이지에 접속한 로그, file.csv 파일에는 이동식 장치(USB 등)에 복사된 파일의 정보(이름, 헤더 등) 로그, device.csv 파일에는 이동식 장치(USB 등)의 연결 및 분리에 대한 로그, email.csv 파일에는 email을 통해 전송된 내용, 첨부파일 등에 대한 로그, psychometric.csv 파일에는 내부자 1000명에 대한 5가지 성격 특성 요소에 대한 정보, LDAP 파일에는 가상의 조직을 구성하고 있는 내부자의 직책, 부서 등의 정보를 가지고 있으며 Fig. 3.과 같이 1명의 사장 아래 6개의 기능 단위와 그 아래 부서, 팀으로 운영되고 있다.

### 4-2 데이터 전처리

**표 3.** 행위 정보 데이터 치환
**Table 3.** Replace behavioral information data

| Replacement number | Behavior information(Activity) |
|---|---|
| 0 | Logon : Logon to a PC |
| 1 | Http : The act of accessing a Web page |
| 2 | Email : The act of exchanging Email |
| 3 | File : Copying files to removable storage media |
| 4 | Connect : Connecting removable storage media to PC |
| 5 | Disconnect : Disconnecting removable storage media from PC |
| 6 | Logoff : Logoff to a PC |

머신러닝 모델링 작업에서 반드시 거쳐야 하는 과정이며, 모델 성능에 직접적인 영향을 주는 과정이기 때문에 매우 중요하다. 본격적인 실험에 앞서 Insider Threat Test Dataset을 전처리 과정을 통해 데이터 유출 징후 탐지 모델에 적합한 데이터로 변환시킨다. 첫 번째로 Insider Threat Test Dataset 내 행위별로 구분되어 있는 로그 파일을 내부자(User ID)별로 재분류하여 내부자 행위 로그 파일을 만든다. 이 과정에서 실험에 불필요한 데이터는 제거하고 시간정보, 내부자 ID, PC ID, 행위정보만 보존한다. 두 번째로 생성된 내부자 행위 로그 파일에서 누락된 값을 다른 값으로 대체하거나 삭제하는 결측치 처리, 부적절한 값과 이상 값을 제거하는 이상치 제거를 통해 정확성을 높인다. 세 번째로 문자열로 구성되어 있는 행위정보(Activity)를 Table 3.과 같이 모델 학습 시 용이한 데이터 타입인 숫자 형태로 치환한다.

### 4-3 알고리즘에 따른 학습 및 실험 방안

머신러닝 모델을 구현하고 학습과 평가하기 위해서는 데이터를 학습데이터와 검증데이터로 분류하는 과정이 필요하다. Insider Threat Test Dataset에는 약 1년 6개월 동안의 내부자 업무 데이터들로 구성되어 있다.

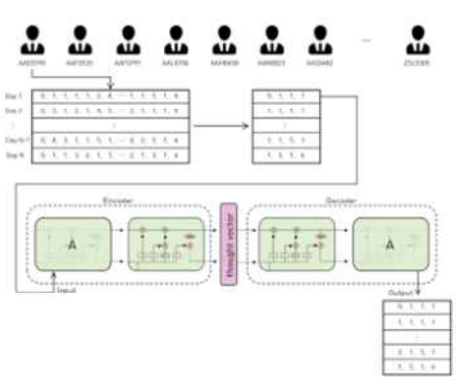

**그림 4.** LSTM Autoencoder 학습 프로세스
**Fig. 4.** LSTM Autoencoder Learning Process

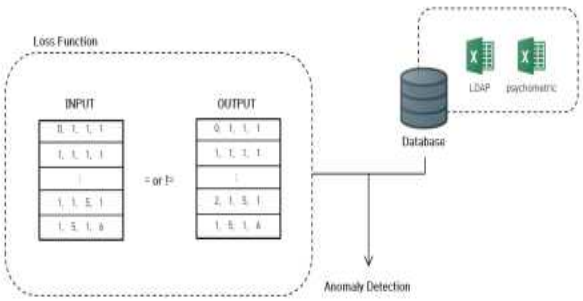


**그림 5.** 데이터 유출 징후 탐지 프로세스
**Fig. 5.** Process for detecting signs of data leakage

해당 Dataset의 내부자별 행위로그 양이 다르기 때문에 특정 로그 양이 아닌 각 로그 파일의 20%에 해당하는 업무 데이터를 정상업무행위라고 가정하고 학습 데이터를 구성한다.

내부자의 행위 로그 파일을 하루 단위의 업무 패턴으로 분류하고 나올 수 있는 최소의 업무 패턴을 추출한다. 최소의 업무 패턴은 한 업무를 수행하는데 필요한 행위들의 집합으로 컴퓨터를 켜고 업무를 수행한 뒤 컴퓨터를 끄는 [logon, http, email, logout], [logon, connect, disconnect, logout] 등 4개의 동작으로 이루어진다. 검증 데이터는 학습 데이터를 통해 학습 한 모델의 성능을 평가하기 위한 데이터로 학습 데이터를 포함한 전체 데이터로 구성한다. LSTM Autoencoder 학습 과정은 각각의 내부자 업무데이터로 만들어진 학습데이터를 이용하여 LSTM Autoencoder 모델을 학습 시키고 검증 데이터를 학습된 모델의 입력 데이터로 넣어 Encoder와 Decoder의 과정을 거쳐 결과 값을 출력 시킨다. 모델의 입력 데이터와 출력 데이터 사이의 loss 값에 각 내부자별 직급 및 5가지 성격 특성 요소가 저장되어 있는 데이터베이스를 통해 패널티를 부여하여 최종적인 loss 값을 구해낸다. 최종 loss 값이 임계치보다 큰 값을 가지는 날을 데이터 유출 징후로 탐지한다.

### 4-4 실험

본 논문에서는 앞서 설명한 LSTM를 이용하여 구현한 Autoencoder를 이용하여 내부자의 정상 업무 행위를 학습하고, 정상 업무 행위에 벗어나는 데이터 유출 징후를 탐지하도록 할 것이다. 사전에 준비한 내부자의 초기 20%의 업무 패턴을 학습 데이터로 구성하고 모델을 학습시킨다. 학습과정은 다음 수식들을 이용한다.

$$y = f(x) = \sigma(Wx + b) \quad (1)$$

$$x' = g(y) = \sigma'(W'y + b') \quad (2)$$

$$L(x, x') = \| x - x' \|^2 \quad (3)$$

**표 4.** 임계값 변경에 따른 성능 평가
**Table 4.** Performance assessment as threshold changes

| | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 |
|---|---|---|---|---|---|
| Accuracy | 0.981 | 0.983 | 0.987 | 0.989 | 0.989 |
| Precision | 0.412 | 0.485 | 0.527 | 0.593 | 0.638 |
| Sensitivity | 0.791 | 0.791 | 0.817 | 0.936 | 0.842 |
| Specificity | 0.997 | 0.998 | 0.998 | 0.998 | 0.999 |

수식(1)은 Autoencoder의 Encoder에 해당하며, 입력 데이터 x를 이용하여 숨겨진 y값을 출력하게 된다. Decoder에 해당하는 수식(2)에 y값을 대입하여 최종 출력 데이터 x'을 출력한다. 마지막으로 입력 데이터 x와 출력 데이터 x'를 수식(3)을 통해 두 값의 같고 다른 정도를 비교하여 같아질수록 0에 가까운 수가 출력되고 다르다면 다른 정도에 따라 큰 수가 출력된다.

학습과정을 통해 내부자들의 정상 업무 행위를 학습한 LSTM Autoencoder를 구할 수 있다. 그 다음으로 검증 데이터인 전체 업무 패턴을 통해 얻은 loss 값과 임의의 임계값(Threshold)을 비교하여 임계값 보다 loss 값이 크면 데이터 유출 징후로 탐지한다. 임의로 정해진 임계값은 데이터 유출 징후를 결정하는데 가장 중요한 값이기 때문에 실험에 알맞은 값을 찾기 위해 값을 변경해가며 반복 실험을 하였고 결과는 Table 4.와 같다. 본 실험의 머신러닝 모델의 성능평가를 할 때는 성능 평가 지표 중 오차행렬을 활용하여 평가한다. 정확도(Accuracy)는 전체 데이터에서 예측 데이터가 얼마나 같은지를 판단하는 지표로 예측한 유출 징후 중 실제 유출 징후와 예측한 정상 업무 중 실제 정상 업무의 비율을 나타낸다. 정밀도(Precision)는 데이터 유출이 발생한 날이라고 예측한 날 중에 실제 데이터 유출이 발생한 날의 비율을 나타낸다. 민감도(Sensitivity)는 실제 데이터 유출이 발생한 날 중에 데이터 유출이 발생했다고 예측한 비율을 나타낸다. 특이도(Specificity)는 실제 정상 업무가 이뤄진 날을 데이터 유출이 발생한 날로 예측한 비율을 나타낸다. Table 4.를 보면 임계값이 0.9에서 1.0으로 변할 때 정확도는 0.001 증가, 정밀도는 0.066 증가, 민감도는 0.119 증가, 특이도는 변화가 없었으며 임계값이 1.0에서 1.1로 변할 때 정확도는 변화 없음, 정밀도는 0.045 증가, 민감도는 0.094 감소, 특이도는 0.001 증가하는 것을 알 수 있었다. 데이터 유출 징후를 탐지함에 있어서는 실제 데이터 유출이 발생한 날 중에 데이터 유출이 발생했다고 예측하는 민감도부분이 중요하기 때문에 임계값을 1.0으로 설정하였을 때 가장 좋은 성능을 나타내는 것을 알 수 있었다.

앞서 실험을 통해 알아내 임계값을 1.0으로 설정한 LSTM Autoencoder 모델을 이용하여 내부자의 데이터 유출 징후를 탐지하였을 때 Table 5.와 같은 결과가 나타났다. The day when the signs of leakage occurred는 내부자에 의해 데이터 유출 징후가 발생한 날을 의미하며 모든 데이터 유출이 하루만 시도되는 것이 아니라 수일에 걸쳐 시도되기 때문에 많은 날이 나타난다.

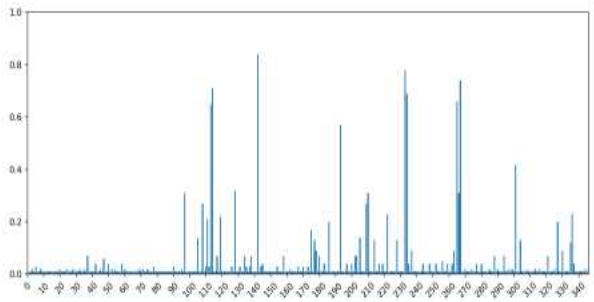

**그림 6.** LSTM Autoencoder 데이터 유출 징후 탐지
**Fig. 6.** LSTM Autoencoder Data Leak Detection

**표 5.** LSTM Autoencoder 탐지 결과
**Table 5.** LSTM Autoencoder Detection Results

| | S1 | S2 | S3 | S4 | S5 | Sum |
|---|---|---|---|---|---|---|
| The day when the signs of leakage occurred | 68 | 207 | 20 | 0 | 0 | 295 |
| Detected | 68 | 183 | 15 | 0 | 0 | 266 |
| Undetected | 0 | 24 | 5 | 0 | 0 | 29 |
| Detection rate(%) | 100 | 88 | 75 | - | - | 88 |
| Precision rate(%) | 72 | 68 | 54 | - | - | 65 |

Detected는 실제 데이터 유출 징후가 탐지된 수, Undetected는 데이터 유출 징후를 탐지하지 못한 수, Detection rate는 데이터 유출 징후가 발생한 날 중 데이터 유출 징후라고 예측한 비율을 나타낸다.

Precision rate는 이후 패널티를 적용시킨 모델과 성능 비교를 위해 데이터 유출이라고 예측한 날 중 실제 데이터 유출이 발생한 날의 비율을 표시한다. 시나리오1에 해당하는 유출 징후는 모두 탐지하였으며, 시나리오2에 해당하는 유출 징후는 88%, 시나리오3은 75%의 탐지율로 총 88%의 유출 징후를 탐지할 수 있었다.

기존 LSTM Autoencoder 모델에 내부자의 직급 별 패널티와 5가지 성격 특성 요소에 따른 패널티를 부가하여 탐지율을 개선시킬 수 있는지 실험한다. 직급이 높을수록 접근 가능한 중요 데이터가 많아짐으로 데이터 유출 시 더 큰 피해가 발생 할 수 있기 때문에 직급 별로 추가 패널티를 부가한다. 실험 데이터로 사용하고 있는 Insider Threat Test Dataset에는 사장(President), 부사장(Vice president), 부서장(Director), 팀장(Manager), 팀원(Member)로 구성되어 있으며 부사장 직급의 내부자는 0.3, 부서장 직급의 내부자는 0.2, 팀장 직급의 내부자는 0.1의 패널티를 설정하였다. 사장의 경우 조직의 최고 관리자로 실험의 목적인 데이터 유출 징후 탐지와는 관계가 없고 팀원은 부사장, 부서장, 팀장 대비 접근 가능한 데이터양이 적어 패널티 부가 대상에서 제외하였다.

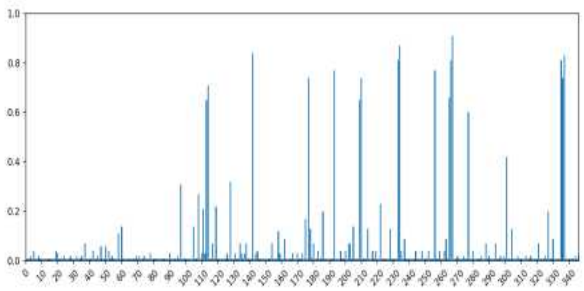

**그림 7.** 패널티를 사용한 LSTM Autoencoder 데이터 유출 징후 탐지
**Fig. 7.** LSTM Autoencoder Data Leak Detection with Penalty

**표 6.** 패널티를 사용한 LSTM Autoencoder 탐지 결과
**Table 6.** LSTM Autoencoder Detection Results with Penalty

| | S1 | S2 | S3 | S4 | S5 | Sum |
|---|---|---|---|---|---|---|
| The day when the signs of leakage occurred | 68 | 207 | 20 | 0 | 0 | 295 |
| Detected | 68 | 198 | 17 | 0 | 0 | 283 |
| Undetected | 0 | 9 | 3 | 0 | 0 | 12 |
| Detection rate(%) | 100 | 96 | 85 | - | - | 94 |
| Precision rate(%) | 58 | 48 | 42 | - | - | 49 |

데이터 유출을 계획하고 실행하려는 내부자는 5가지 성격 특성 요소 중 성실성, 외향성, 우호성이 낮아지고 신경성이 높아지는 변화가 생긴다는 것을 가정하고 내부자별 5가지 성격 특성 요소에 따라 패널티를 부가한다. 실험 데이터로 사용하고 있는 Insider Threat Test Dataset에는 내부자별 5가지 성격 특성 요소를 측정하여 10~50의 수치로 표현한 psychometric.csv 파일이 존재한다. 성실성, 외향성, 우호성 항목이 20 이하로 낮으면 0.1의 패널티를 부가하고 신경성은 40 이상일 때 0.2의 패널티를 부가한다.

성실성, 외향성, 우호성은 성격에 따라 쉽게 변화할 수 있지만 신경성의 변화에는 외부적인 요인이 필요하기 때문에 패널티 부가를 다르게 설정하였다. LSTM Autoencoder 모델에 직급과 5가지 성격 특성 요소에 따른 패널티를 부가하여 내부자의 데이터 유출 징후를 탐지하였을 때 Table6.과 같은 결과가 나타났다. 시나리오1에 해당하는 유출 징후는 모두 탐지하였으며, 시나리오2에 해당하는 유출 징후는 94%, 시나리오3은 85%의 탐지율로 총 93%의 유출 징후를 탐지할 수 있었다.

실험 결과 패널티를 적용하지 않은 LSTM Autoencoder 모델보다 직급과 5가지 성격 특성 요소에 따른 패널티를 부가하여 내부자의 데이터 유출 징후를 탐지하였을 때 탐지율 측면에서 약 6% 향상된 것을 확인할 수 있었다. 하지만 데이터 유출이라고 예측한 날 중 실제 데이터 유출이 발생한 비율인 정밀도 부분에서 약 16% 하락한 것을 알 수 있었다. 이는 실제로는 정상 업무 행위지만 데이터 유출 징후로 탐지한 건수가 늘어난 것이지만 본 논문에서는 유출 징후를 판단하는 것이기 때문에 큰 영향성은 없다. 기존 모델에서는 정상 업무 패턴이 이동식 저장장치를 많이 사용하는 업무를 수행하다가 데이터 유출을 위해 저장장치를 많이 사용하게 될 때도 유출 징후가 아닌 정상 업무로 분류했지만, 해당 내부자의 직급이 팀장인 것과 신경성이 42로 높다는 것으로 패널티를 적용받아 데이터 유출 징후로 탐지해 낼 수 있었다.

## V. 결 론

본 논문에서는 금전적 피해와 신뢰도 하락, 이미지 실추 등 막대한 피해를 입을 수 있는 내부자 위협에 대응하기 위해 내부자에 의한 데이터 유출 징후 탐지를 위한 머신러닝 모델에 관한 연구를 진행 하였다. 기존 머신러닝 알고리즘을 활용한 내부자에 의한 데이터 유출 징후 탐지 연구에서는 다양한 머신러닝 알고리즘을 활용하여 모델을 구현하는 연구가 진행됐지만 탐지 지표에 다양성이 부족하다는 한계가 있었다. 본 논문에서는 긴 시간 동안의 정보를 기억하여 시계열과 시퀀스 데이터 처리에 용이한 LSTM 알고리즘과 Autoencoder를 결합하여 LSTM Autoencoder를 구현하여 데이터 유출 징후를 탐지하는 모델을 만들고, 직급이 높을수록 중요하고 많은 데이터에 접근 가능하다는 점과 데이터 유출에 따른 5가지 성격 특성 요소를 통한 패널티를 적용시켜 탐지율을 향상시킬 수 있는 방법에 대해 실험을 진행 하였다. 실험 데이터로는 미국 카네기멜론 대학의 CERT 내부자 위협 센터에서 내부자 위협 연구를 위해 제공하고 있는 Insider Threat Test Dataset을 활용하였다. 해당 데이터는 내부자별 행위 로그 파일로 분류하고 결측치 처리, 이상치 제거 등 전처리하여 학습데이터와 검증데이터로 구성하였고 Autoencoder의 특징을 이용하여 모델의 입력 데이터와 출력 데이터 사이의 loss값에 내부자별 직급 및 5가지 성격 특성 요소에 따른 패널티를 적용하여 최종적인 loss 값을 구해냈다. 최종 loss 값이 설정한 임계값보다 큰 값을 가지는 날을 데이터 유출 징후가 발생한 날로 탐지하였다. 실험 결과 패널티를 적용시키지 않은 기본 LSTM Autoencoder 모델보다 패널티를 적용했을 때 약 6% 향상된 탐지율을 볼 수 있었다. 이는 기존 모델과 비교하여 직급과 5가지 성격 특성 요소를 탐지 지표로 활용했을 때 더 많은 데이터 유출 징후를 탐지함으로써 좋은 성능을 갖고 있다고 말할 수 있다. 그러나 정밀도 부분에서 약 16% 하락하여 관리자가 예측된 데이터 유출 징후를 정·오탐 구분할 때 어려움이 발생할 수 있다. 또한, Insider Threat Test Dataset에는 내부자별 5가지 성격 특성 요소를 정리해 놓았지만 실제 기업에서는 내부자별로 5가지 성격 특성 요소 검사를 진행하여 데이터로 정리해놓는 기업이 많지 않다.

향후 연구로는 기업별 파일의 중요도, 내부자가 어떤 웹사이트에 접속하여 어떤 행위를 하는지에 대한 데이터 등 탐지

지표를 더욱 구체화하여 더 높은 성능을 나타내는 모델과 패널티 적용에 의한 정밀도 하락을 더욱 향상 시킬 수 있는 방법을 을 연구할 계획이다.

## 참고문헌


[1] Dongwook Ha, A Study on the Machine Learning Model for the Detection of Insider Abnormal Behavior, Master's thesis, Myongji University Graduate School, Seoul, Feb 2018.

[2] Minhae Jang, Insider Abnormal Behaviour Detection Using Deep Learning, Master's thesis, Myongji University Graduate School, Seoul, Feb 2019.

[3] Jeongmin Lee, Yeji Jeon, Jumin Oh, Ki Young Lee, "An Implementation of the Insider Anomaly Behavior Detection System Using Machine Learning Algorithm," in *2021 Korea Electronics Association Summer Conference*, Jeju, pp. 1801-1802, Jun 2021.

[4] Lee YangWoo, Kim SeongJik, Ha SeungTae, Lee HeonGyu, "Insider Anomaly Detection Method Using Machine Learning," in *2018 Korea Software Conference*, Pyeongchang, pp. 971-973, Dec 2018.

[5] HyunSoo Kim, "A Study on Method for Insider Data Leakage Detection," *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 17, No. 4, pp. 11-17, Aug 2017.
https://doi.org/10.7236/JIIBC.2017.17.4.11

[6] Sangmok Kim, Open Soutce Based Machine Learning Abnormal Detection Techniques for Small Organizations, Master's thesis, Ajou University Graduate School, Suwon, Feb 2020.

[7] Ju Young Lee, Goo Yeon Lee, Ho Yeol Kwon, "Insider Information Leakage Detection Method using Scenario Technique," *The Journal of Digital Contents Society*, Vol. 21, No. 3, pp. 617-626, Mar 2020.
http://dx.doi.org/10.9728/dcs.2020.21.3.617



**김서준(Seo-Jun Kim)**

2018년 : 학점은행제 정보보호학(학사)
2018년～2022년 : 공군 정보통신장교
2019년～현 재 : 아주대학교 정보통신대학원 정보통신공학과 (석사)

※관심분야 : 정보보호, 기계학습(Machine Learning) 등



**손태식(Tae-Shik Shon)**

2000년 : 아주대학교 정보및컴퓨터공학부 졸업(학사)
2002년 : 아주대학교 정보통신전문대학원 졸업(석사)
2005년 : 고려대학교 정보보호대학원 졸업(박사)

2004년～2005년: University of Minnesota 방문연구원
2005년～2011년: 삼성전자 통신·DMC 연구소 책임연구원
2017년～2018년: Illinois Insitute of Technology 방문교수
2011년～현 재: 아주대학교 정보통신대학 사이버보안학과 교수

※관심분야 : Digital Forensics, ICS/SCADA Security, Anomaly Detection