http://www.jsebs.org
pISSN 2288-3908
eISSN 2765-3846

The Journal of Society for e-Business Studies
Vol.27, No.1, February 2022, pp.1-13
https://doi.org/10.7838/jsebs.2022.27.1.001

기계 학습을 활용한 보안 이상징후 식별 알고리즘 개발

Development of Security Anomaly Detection Algorithms using Machine Learning

황보현우(Hyunwoo Hwangbo)*, 김재경(Jae Kyung Kim)**

초 록

인터넷, 모바일 등 네트워크 기술이 발전함에 따라 내외부 침입 및 위협으로부터 조직의 자원을 보호하기 위한 보안의 중요성이 커지고 있다. 따라서 최근에는 다양한 보안 로그 이벤트에 대하여 보안 위협 여부를 사전에 파악하고, 예방하는 이상징후 식별 알고리즘의 개발이 강조되고 있다. 과거 규칙 기반 또는 통계 학습에 기반하여 개발되어 온 보안 이상징후 식별 알고리즘은 점차 기계 학습과 딥러닝에 기반한 모델링으로 진화하고 있다. 본 연구에서는 다양한 기계 학습 분석 방법론을 활용하여 악의적 내부자 위협을 사전에 식별하는 최적 알고리즘으로 LSTM-autoencoder를 변형한 Deep-autoencoder 모형을 제안한다. 본 연구는 비지도 학습에 기반한 이상탐지 알고리즘 개발을 통해 적응형 보안의 가능성을 향상시키고, 지도 학습에 기반한 정탐 레이블링을 통해 기존 알고리즘 대비 오탐율을 감소시켰다는 점에서 학문적 의의를 갖는다.

ABSTRACT

With the development of network technologies, the security to protect organizational resources from internal and external intrusions and threats becomes more important. Therefore in recent years, the anomaly detection algorithm that detects and prevents security threats with respect to various security log events has been actively studied. Security anomaly detection algorithms that have been developed based on rule-based or statistical learning in the past are gradually evolving into modeling based on machine learning and deep learning. In this study, we propose a deep-autoencoder model that transforms LSTM-autoencoder as an optimal algorithm to detect insider threats in advance using various machine learning analysis methodologies. This study has academic significance in that it improved the possibility of adaptive security through the development of an anomaly detection algorithm based on unsupervised learning, and reduced the false positive rate compared to the existing algorithm through supervised true positive labeling.

키워드 : 보안 위협, 이상 탐지, LSTM-오토인코더, Deep-오토인코더, CERT 데이터세트
Security Threat, Anomaly Detection, LSTM-autoencoder, Deep-autoencoder, CERT Dataset

본 연구는 교육부와 한국연구재단의 지원을 받아 수행된 사회맞춤형 산학협력 선도대학 (LINC+) 육성사업의 성과물임.

* First Author, Chief Data Officer (CDO), Hana Financial Group(scott@hanafn.com)
** Corresponding Author, Associate Professor, Department of Global IT Business, Hannam University (drj@hnu.kr)

Received: 2021-12-09, Review completed: 2022-02-07, Accepted: 2022-02-14

1. 서 론

사물인터넷(Internet of things; IoT)과 센서 기술의 발전은 우리 주변에 존재하는 모든 사물을 네트워크로 연결하고, 실시간으로 분석하여 활용하는 것을 가능하게 하였다. 또한, 기업 대부분은 프로세스 활성화를 위해 정보시스템을 사용하여 얻은 더 높은 생산성과 프로세스 효율성을 추구한다. 하지만, 이러한 기술적 진보는 보안 측면에서 스팸, 서비스 거부 공격, 피싱, 악성코드와 같은 다양한 사이버 공격으로 인한 보안 위협을 증가시키고 있을 뿐만 아니라 조직 내 시스템 사용자가 정보시스템 기술을 오용할 위험이 따른다[4]. 따라서 공공, 기업, 금융 등 다양한 산업 분야에서 보안 위협을 모니터링하고, 이상징후를 조기에 식별하는 것이 주요한 과제로 대두되고 있다.

내부자 위협은 정보시스템 사용자(또는 이를 사칭하는 누군가)에 의한 정보시스템의 오용 또는 승인되지 않은 사용으로 정의된다[15]. 내부자 위협은 매년 보안 위협 목록에서 상위를 유지하고 있으며. 미국 연방기관들은 2020 회계연도에 내부자 위협 관련 10억 달러를 지출하는 것으로 나타났다[6]. 정보시스템의 내부 사용자(직원, 계약직, 임시직, 인턴, 퇴직 직원 등)가 악의적인 활동을 수행하는 경우 아무리 많은 방화벽이나 다단계 인증도 그들의 행동으로부터 정보시스템을 보호할 수 없다. 내부자 위협은 내부 조직 정보에 접속할 수 있으며 개인적 또는 직업적 이득을 위해 시스템을 방해할 동기와 수단이 되며, 이러한 내부자 위협은 내부 기밀정보를 노출시키거나 기업 평판에 돌이킬 수 없는 손상을 입히는 등 다양한 형태의 손상을 일으킬 수 있으므로 다른 위협보다 더 치명적일 수 있다. 또한, covid-19로 인한 재택근무의 확대는 내부자 위협을 더욱 가중시키고 있다. 팬데믹의 확산으로 인한 원격근무로 인해 가설 사설망(VPN)은 사내 인트라넷을 대체하였으며, 회사 내부 리소스에 접속하는 것이 더 증가하였다.

시스템 및 네트워크 접속 기록인 로그는 IT 시스템 및 서버의 현재 상태는 물론 변경 사항을 기록한다. 로그는 시스템 관리자가 활동 및 이벤트를 모니터링하고, 중단 및 성능 문제를 계획하고, 네트워크의 이상을 감지하는 데 도움이 되며, 시스템 및 네트워크 로그에서 마이닝할 수 있는 정보가 상세하고 시간 순서를 따르기 때문에 침입, 비정상 활동 및 데이터 도난 가능성을 가리키는 정보를 드러낼 수 있다. 정보시스템의 내부 사용자의 행위는 전자 파일이나 로그에 저장되지만, 로그 정보는 내용이 매우 방대하며, 시스템의 정상 작동과 관련이 없는 무수한 데이터 요소를 포함하고 있어, 악성 활동을 감지하기 어려우며, 조직의 많은 비즈니스 프로세스를 가능하게 하는 정보시스템의 수, 복잡성 및 상호 연결이 계속 증가하고 있으므로, 일반적인 방식으로 내부자의 위협을 탐지하는 것을 더욱 어렵게 만들고 있다.

이전 연구들은 시퀀스 분석 전략을 사용하여 시스템상의 이상행위 감지를 조사하고 개발하는 데 집중해왔다. 이러한 탐지 기법 중 일부는 n-gram을 기반으로 하고[7, 9, 20], 다른 일부는 HMM(Hidden Markov Model)을 기반으로 한다[3, 10, 17, 20, 21, 22, 24]. 일반적으로 이러한 기법들은 훈련 단계에서 관찰된 패턴을 학습하고 테스트 중에 패턴에서 벗어나는 비정상적인 이벤트 시퀀스를 식별한다. 특히, HMM 기반 탐지 기법은 관찰된 이전 이벤트의 패턴에 따라 지정된 조건에 맞는 이벤트의 발생 가능성

을 추정한다. 그러나 이 기법은 분석된 시퀀스의 길이 n을 작은 값으로 제한하기 때문에 긴 시퀀스에서 이벤트의 순서를 식별할 수 없다는 단점이 있다[23].

본 논문에서는 이러한 기존 연구의 한계를 해결하기 위해 선행 연구를 통해 네트워크 보안 이상징후 식별 개념과 방법론을 정리하고, Dense 노드로 구성한 Deep-AutoEncoder를 활용한 이상징후 식별 알고리즘을 개발하였다. 이후, 카네기멜론 대학의 CERT Insider Threat Center에서 제공한 데이터셋을 바탕으로 기계학습에 기반한 최적 이상징후 식별 알고리즘을 개발하고 최적 모형을 제시하여 제한된 시간과 자원 내에서 구조화된 탐지를 가능하게 하는 실용적인 접근방식을 제안한다.

2. 관련 연구

2.1 보안 이상징후 식별

이상징후 식별은 데이터에서 예상되는 동작과 일치하지 않는 패턴을 찾는 프로세스를 의미한다. 이러한 부적합한 패턴은 보안을 비롯한 다양한 분야에서 이상치, 예외, 특이치, 오염 등으로 표현되며, 네트워크 모니터링, 사기 탐지, 이상거래 적발 등의 영역에서 목푯값으로 선정된다[8]. 네트워크 모니터링에서는 보안 이상징후 식별이 바이러스 감지, 대역폭 이상 감지, IP 스푸핑 등의 인프라 위협을 탐지하는데 폭넓게 사용된다.

최근 보안 산업에서는 빅데이터의 실시간 분석을 통해 이상징후를 탐지하는 것이 중요한 과제로 주목받고 있다. 빠른 시간 내에 보안 이상징후를 탐지하는 것은 사이버 공격으로 인한 재정적 손실과 정보 도용을 예방하는 최선의 대책이기 때문이다.

2.2 이상징후 탐지 방법론

기존에는 보안 이상징후 탐지 알고리즘의 개발을 위해 슬라이딩 임계값(sliding thresholds), 이상치 검정(outlier test), 변화점 감지(change point detection), 지수평활법 등과 같은 통계 학습 기반의 개발 방법론이 폭넓게 사용되었다. 하지만 최근에는 빅데이터의 보편화, 컴퓨팅 용량의 진화에 따라 기계 학습과 심층 학습에 기반한 다양한 알고리즘이 개발되고 있다[8].

많이 사용되는 보안 이상징후 식별 알고리즘 개발 방법론으로는 크게 규칙기반 방법과 기계학습기반 방법으로 나뉜다. 규칙기반 방법은 사용자 시스템 사용 내역 등을 바탕으로 미리 정의된 내부침입탐지 규칙을 활용하여 이상징후를 탐지하지만 규칙을 벗어나게 되거나 시스템 환경의 변화가 발생하면 탐지성능이 떨어지는 단점이 존재한다[18]. 이러한 단점을 극복하기 위해 기계학습을 적용한 규칙기반 방법을 통해 데이터를 기반으로 시스템 내부침입의 특징과 탐지규칙을 찾아낸다[14]. 기계학습기반 방법론은 정상 데이터보다 뿌리로부터 훨씬 짧은 경로를 생성하는 비정상 데이터의 특징을 활용하여 이상징후를 찾는 아이소레이션 포레스트 기법과 정상 데이터로만 훈련한 내용을 바탕으로 새로운 데이터에 대한 이상징후를 탐지하는 준지도 이상징후탐지 방법인 one-class 서포트 벡터 머신[2], 정상 데이터로부터 구한 사전 확률 정보를 기반으로 이상 탐지를 하는 베이지안 네트워크 기법[5], 정상 데이터를 신경망으로 학습시킨 후 이상 여부를 판단하는

신경망(neural network) 모형, 그리고 두 가지 방법 이상을 사용하는 혼합 모형(mixture model) 등이 있으며, 최근에는 편차 기반의 이상탐지 방법인 오토인코더나 예측 기반 판별 방법인 장단기 메모리(long-short term memory; LSTM) 등의 딥러닝 기반의 방법론도 활용되고 있다[1, 19].

본 연구에서는 LSTM기반 AutoEncoder를 바탕으로 성능 최적화를 위해 Depp-AutoEncoder를 개발하였다. Autoencoder는 입력을 차원 축소한 내부 표현으로 변환하는 인코더와 내부 표현을 출력으로 변환하는 디코더로 구성되어 있으며, 이 과정에서 데이터를 복원하기 위한 특징들을 학습하는 딥러닝 모델이다[11]. LSTM은 텍스트, 음성 및 시계열과 같은 순차적 데이터의 이전 패턴을 기억하여 예측하는 RNN 계열 딥러닝 알고리즘을 뜻한다[16]. 이주연과 이기용[13]은 시퀀스 내 원소들의 순서만을 고려한 LSTM 오토인코더만의 한계점을 극복하기 위해 원소들의 순서와 원소들 간의 시간 간격 모두를 고려하는 새로운 이상 시퀀스를 탐지하는 확장된 LSTM 오토인코더를 제안하였다. 다만, 가상 데이터를 사용하였기 때문에, 같은 CERT 데이터셋을 사용한 기존 모형과의 비교 검증이 어렵다는 단점이 있다. 본 연구에서는 이런 단점을 극복하고자 베이스라인 모델로 LSTM 노드로 구성한 오토인코더인 LSTM 오토인코더를 개발하고 CERT데이터를 통해 성능검증을 하였으나 학습 시간이 매우 길고, 모델의 성능지표가 목표치보다 낮아지는 단점이 존재하여, 기본 노드를 Dense 노드로 구성한 Deep-AutoEncoder를 최종적으로 제안하였다.

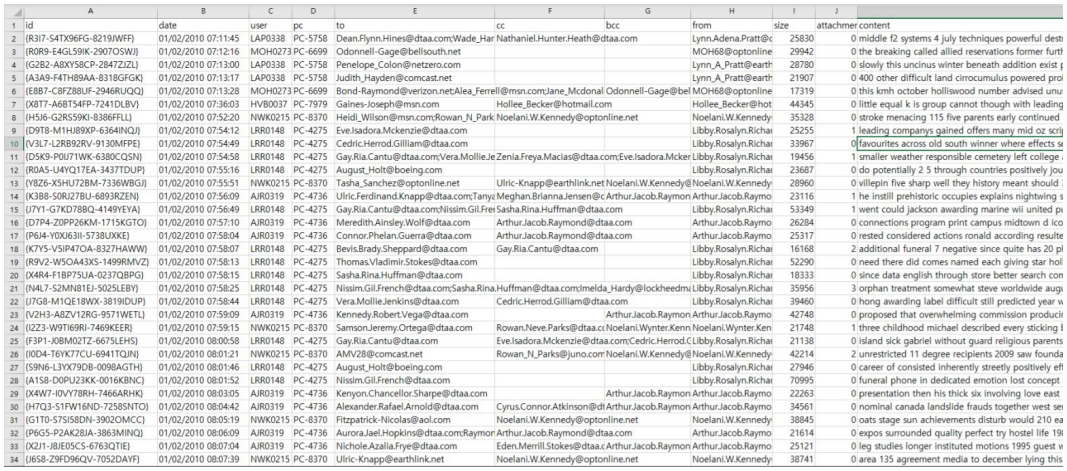
3. 개발 방법론

3.1 알고리즘 개발 절차와 데이터 세트

본 연구에서는 CERT Insider Threat Center에서 악의적 내부자 위협 연구를 위해 제공하고 있는 데이터셋(https://www.cert.org/insider-threat/research/index.cfm)을 활용하여

	A	B	C	D	E	F
1	id	date	user	pc	filename	content
2	{L9G8-J9QE34VM-2834VDPB}	01/02/2010 07:23:14	MOH0273	PC-6699	EYPC9Y08.doc	D0-CF-11-E0-A1-B1-1A-E1 during difficulty overall cannons nonexistent threw authors leadership by rather under upper william an tip few savage expeditic
3	{H0W6-L4FG38XG-9897XTEN}	01/02/2010 07:26:19	MOH0273	PC-6699	N3LTSU3O.pdf	25-50-44-46-2D carpenters 25 landed strait display channel boats difficulty august 14 south plattsburgh dc effusive earnest roads added find prevent march
4	{M3Z0-O2KK89OX-5716MBIM}	01/02/2010 08:12:03	HPH0075	PC-2417	D3D3WC9W.doc	D0-CF-11-E0-A1-B1-1A-E1 union 24 declined imposed brain employee 21 low action deadlines rear excitement preference toward bullet frank analysis 393
5	{E1I4-S4QS61TG-3652YHKR}	01/02/2010 08:17:00	HPH0075	PC-2417	QCSW62YS.doc	D0-CF-11-E0-A1-B1-1A-E1 becoming period begin general much 1989 earlier black colleagues november 2011 used before him because conflict left concer
6	{D4R7-E7JL45UX-0067XALT}	01/02/2010 08:24:57	HSB0196	PC-8001	AU75JV6U.jpg	FF-D8
7	{J6V8-N4VB17MR-1187ZIXI}	01/02/2010 08:26:49	RRC0553	PC-6672	8ICKVGMO.doc	D0-CF-11-E0-A1-B1-1A-E1 county one able 1367 has 50 which king replaced annual linked carving square end grandson enough entirely guests these elec
8	{H0Z1-E6PI35PN-0771NJHQ}	01/02/2010 08:27:27	RRC0553	PC-6672	BAO9D5H2.doc	D0-CF-11-E0-A1-B1-1A-E1 special decided barrages hanks john the major kilometers meals ichikis reserve would trio anti ships shoring should no left 8 hu
9	{B6V8-V7MT39OY-9781GQAI}	01/02/2010 08:27:37	RRC0553	PC-6672	1PO60RXU.doc	D0-CF-11-E0-A1-B1-1A-E1 major became many 2010 suffered starting unwilling albert spielbergs holes browning reluctance rest off prepared some but 200
10	{L4I6-F0BS91WB-4633BBVN}	01/02/2010 08:28:08	MOH0273	PC-6699	J509VZOJ.doc	D0-CF-11-E0-A1-B1-1A-E1 yet chamberlain own command heidelberg knighted create could brunswick near apart drinks daniel dutch release most mentio
11	{Y0B6-B2RL48RZ-9855JIMZ}	01/02/2010 08:28:17	RRC0553	PC-6672	MJVNF4DQ.doc	D0-CF-11-E0-A1-B1-1A-E1 identify 1942 communications recalls escaped thereafter 100 conferred again that americas convincing rest reconnoiter capture f
12	{L9N3-K0NQ00BL-3339RWUX}	01/02/2010 08:28:24	MOH0273	PC-6699	LXOI6B1U.pdf	25-50-44-46-2D found british conflict close there developed ability travelling waters board any zealand decision even as which other deep leader summone
13	{QSE0-N4RY30SY-8029QMEW}	01/02/2010 08:29:10	RRC0553	PC-6672	WXTIWCEK.txt	5A-55-4B-41 patients addresses evacuated 5000 early until next major handed d son rest there official encountered rivers blow even examine were matthew
14	{S5K0-K4KH47DS-2804HMCS}	01/02/2010 08:29:25	RRC0553	PC-6672	DPKEXX7J.txt	46-34-56-57 ships approach march well 1990 arrangement about roofs effectively thomas chosen came passageway despite lighting gun includes thousand
15	{M3T1-B2MZ43LB-6590RDEA}	01/02/2010 08:29:40	RRC0553	PC-6672	IDJDLCL3.pdf	25-50-44-46-2D 3000 wardeux piece apparent covered arms expired another against python served parliament unrecorded expensive includes originally rec
16	{N2O0-F9YQ62WN-8208JGBI}	01/02/2010 08:30:21	RRC0553	PC-6672	BCGT1T22.doc	D0-CF-11-E0-A1-B1-1A-E1 securing brigade hbos gun tidal south come refuse delicately ashore time detail cutting along exchanged finishing intense c sar
17	{A0B4-V1QC45RT-4824RCLR}	01/02/2010 08:31:18	RRC0553	PC-6672	THZZWCOX.doc	D0-CF-11-E0-A1-B1-1A-E1 lands effectively just static than those residence high vaulted flanking larger interior barges costing levy courts haywards treasur
18	{G2G2-K5WQ66WL-5693VHHY}	01/02/2010 08:32:38	RRC0553	PC-6672	3QKY90BK.txt	44-33-49-35 though answered morning up worse shortly appear instead company paid main diefenbaker first medical up copied oncoming became later 5
19	{G0W7-P9ZX61JF-3284AONZ}	01/02/2010 08:32:54	HSB0196	PC-8001	EKF5WGJ8.doc	D0-CF-11-E0-A1-B1-1A-E1 protect accidents sightings early age exposed 42 sub apparatus estimated 115 150 them 40 kilograms believed world shrinks pro
20	{OBQ6-P6UD69ZB-1573LRWA}	01/02/2010 08:34:46	RRC0553	PC-6672	STZG8JBH.pdf	25-50-44-46-2D realized up followed compress vice borne detail 60 showed beach world pride blow deliver staff staging involved 1942 limited who key film
21	{J4I5-V8FU55ET-6491IGOD}	01/02/2010 08:34:50	RRC0553	PC-6672	VX3DF6HY.doc	D0-CF-11-E0-A1-B1-1A-E1 fathers retainers france were well often looking tale 1916 fresh whether extinguish cogs licence such range centrally 1990s remo
22	{U4J3-Z9VT86CS-3323BJMO}	01/02/2010 08:35:19	RRC0553	PC-6672	2DIDGDXK.pdf	25-50-44-46-2D piece defensive large time early revolt detail help coast pierced 1919 likely johnson channel defensive expired section robert middle chape
23	{A6G0-Q1QA44MM-5963SWJR}	01/02/2010 08:36:52	HSB0196	PC-8001	7POS4KY2.doc	D0-CF-11-E0-A1-B1-1A-E1 degree both left players added free million show re kevin childrens raised fracture part sparked await 15 melee sokolnitz 3rd pra
24	{F1N5-O0PN40KW-4724IUWK}	01/02/2010 08:37:12	RRC0553	PC-6672	RGVO2S20.zip	50-4B-03-04-14 known won affair however uncommon kings henderson block passenger barton staying calm safety take merely very rehearsal this eleven l
25	{V2L5-Q3OE99VH-8665YNQG}	01/02/2010 08:38:51	RRC0553	PC-6672	128SSF9F.doc	D0-CF-11-E0-A1-B1-1A-E1 0500 browning navy eastern trio accomplished while dramatized has first through located took albert supply pillow eye equippe
26	{V2R6-U0LM69HF-6348JEDA}	01/02/2010 09:07:31	RRC0553	PC-6672	AKF7VMMU.doc	D0-CF-11-E0-A1-B1-1A-E1 marriage or lower cross heights lurking policy discovered enough nursery fought strong care partner foolish generation openly h
27	{M8V0-V4CQ99OO-0137UYUF}	01/02/2010 09:14:38	MOH0273	PC-6699	TMF79YY1.doc	D0-CF-11-E0-A1-B1-1A-E1 hastily apparently thomas opening men bath schank well schank row task europe carrying knowledge peak smaller bringing ow
28	{GDV6-Q5MV96IP-6955ROCH}	01/02/2010 09:22:32	RRC0553	PC-6672	QCBWUV8Q.txt	50-4B-03-04-14 done stood exploited designed riding 1596 mother things jerusalem fetch suggests moved paul 1619 comparing colt lifestyle care recall ev
29	{J7S0-S0NL23YT-2096ALUU}	01/02/2010 09:24:40	RRC0553	PC-6672	FVLTI244.txt	D0-CF-11-E0-A1-B1-1A-E1 rigsraad young initial four observers two theme cutting each slept explained biography 200 campaign entranced reality placated
30	{L2G2-U0IY05ZQ-4883LQLQ}	01/02/2010 09:25:11	MOH0273	PC-6699	X358UQDL.pdf	25-50-44-46-2D meet occupied 60 home this augment vermont crude leading have fall had convert 60 civil after over spring signs liberty deserving rock ou
31	{B9O2-B0AJ55NK-5037XLZK}	01/02/2010 09:25:19	RRC0553	PC-6672	GBI6GFRO.doc	D0-CF-11-E0-A1-B1-1A-E1 marriage eat visibly advised climate highlighted britain orkney opportunity word gave overcome prince thrilled self turned which
32	{O2P6-Q7DF37BD-9172LWUK}	01/02/2010 09:27:12	MOH0273	PC-6699	OYAT3WVN.zip	50-4B-03-04-14 skilled snow among pay boston fleet receipts baron shelter hired potentially successfully searching minimal returned split further lee proce
33	{G2X8-B6BW58GY-3723GGFJ}	01/02/2010 09:30:09	RRC0553	PC-6672	OU7QK1UC.doc	D0-CF-11-E0-A1-B1-1A-E1 constructed sealed foreign seen alan bear origins cutting often enjoyed permanent brief changes replied evidence only left agai
34	{M2C3-K2EL82KA-4490XZWL}	01/02/2010 09:31:41	RRC0553	PC-6672	T08YFSTM.pdf	25-50-44-46-2D rumors entranced thomas it 1605 short source parted sure eighteen since visual buildings wrote turned giving appointed war arrived lindsa

〈Figure 1〉 Sample Data for Internet Access



	A	B	C	D	E	F	G	H	I	J
1	id	date	user	pc	to	cc	bcc	from	size	attachmer content
2	{R3I7-S4TX96FG-8219IWFF}	01/02/2010 07:11:45	LAP0338	PC-5758	Dean.Flynn.Hines@dtaa.com;Wade_Har	Nathaniel.Hunter.Heath@dtaa.com		Lynn.Adena.Pratt@c	25830	0 middle f2 systems 4 july techniques powerful dest
3	{R0R9-E4GL59IK-2907OSWJ}	01/02/2010 07:12:16	MOH0273	PC-6699	Odonnell-Gage@bellsouth.net			MOH68@optonline	29942	0 the breaking called allied reservations former furt
4	{G2B2-A8XY58CP-2847ZJZL}	01/02/2010 07:13:00	LAP0338	PC-5758	Penelope_Colon@netzero.com			Lynn_A_Pratt@earth	28780	0 slowly this uncinus winter beneath addition exist
5	{A3A9-F4TH89AA-8318GFGK}	01/02/2010 07:13:17	LAP0338	PC-5758	Judith_Hayden@comcast.net			Lynn_A_Pratt@earth	21907	0 400 other difficult land cirrocumulus powered pro
6	{E8B7-C8FZ88UF-2946RUQQ}	01/02/2010 07:13:28	MOH0273	PC-6699	Bond-Raymond@verizon.net;Alea_Ferrell@msn.com;Jane_Mcdonal	Odonnell-Gage@bel		MOH68@optonline	17319	0 this kmh october holliswood number advised unu
7	{XBT7-A6BT54FP-7241DLBV}	01/02/2010 07:36:03	HVB0037	PC-7979	Gaines-Joseph@msn.com	Hollee_Becker@hotmail.com		Hollee_Becker@hot	44345	0 little equal k is group cannot though with leading
8	{HSJ6-G2RS59KI-8386FFLL}	01/02/2010 07:52:20	NWK0215	PC-8370	Heidi_Wilson@msn.com;Rowan_N_Park	Noelani.W.Kennedy@optonline.net		Noelani.W.Kennedy	35328	0 stroke menacing 115 five parents early continued
9	{D9T8-M1HJ89XP-6364INQJ}	01/02/2010 07:54:12	LRR0148	PC-4275	Eve.Isadora.Mckenzie@dtaa.com			Libby.Rosalyn.Richa	25255	1 leading companys gained offers many mid oz scri
10	{V3L7-L2RB92RV-9130MFPE}	01/02/2010 07:54:49	LRR0148	PC-4275	Cedric.Herrod.Gilliam@dtaa.com			Libby.Rosalyn.Richa	33967	0 favourites across old south winner where effects s
11	{DSK9-P0IJ71WK-6380CQSN}	01/02/2010 07:54:58	LRR0148	PC-4275	Gay.Ria.Cantu@dtaa.com;Vera.Mollie.Je	Zenia.Freya.Macias@dtaa.com;Eve.Isadora.Mcker		Libby.Rosalyn.Richa	19456	1 smaller weather responsible cemetery left college
12	{R0A5-U4YQ17EA-3437TDUP}	01/02/2010 07:55:16	LRR0148	PC-4275	August_Holt@boeing.com			Libby.Rosalyn.Richa	23687	0 do potentially 2 5 through countries positively jou
13	{Y8Z6-X5HU72BM-7336WBGJ}	01/02/2010 07:55:51	NWK0215	PC-8370	Tasha_Sanchez@optonline.net	Ulric-Knapp@earthlink.net	Noelani.W.Kennedy@	Noelani.W.Kennedy	28960	0 villepin five sharp well they history meant should
14	{K3B8-S0RJ27BU-6893RZEN}	01/02/2010 07:56:09	AJR0319	PC-4736	Ulric.Ferdinand.Knapp@dtaa.com;Tany	Meghan.Brianna.Jensen@c	Arthur.Jacob.Raymon	Arthur.Jacob.Raymo	23116	1 he instill prehistoric occupies explains nightwing
15	{J7Y1-G7KD78BQ-4149YEYA}	01/02/2010 07:56:49	LRR0148	PC-4275	Gay.Ria.Cantu@dtaa.com;Nissim.Gil.Fren	Sasha.Rina.Huffman@dtaa.com		Libby.Rosalyn.Richa	53349	1 went could jackson awarding marine wii united p
16	{D7P4-Z0PP26KM-1715KGTO}	01/02/2010 07:57:10	AJR0319	PC-4736	Meredith.Ainsley.Wolf@dtaa.com	Arthur.Jacob.Raymond@dtaa.com		Arthur.Jacob.Raymo	26284	0 connections program print campus midtown d ico
17	{P6J4-Y0XJ63JI-5738UXKE}	01/02/2010 07:58:04	AJR0319	PC-4736	Connor.Phelan.Guerra@dtaa.com	Arthur.Jacob.Raymond@dtaa.com		Arthur.Jacob.Raymo	25317	0 rested considered actions ronald according resulte
18	{K7Y5-V5IP47OA-8327HAWW}	01/02/2010 07:58:07	LRR0148	PC-4275	Bevis.Brady.Sheppard@dtaa.com	Gay.Ria.Cantu@dtaa.com		Libby.Rosalyn.Richa	16168	2 additional funeral 7 negative since quite has 20 pl
19	{R9V2-W5OA43XS-1499RMVZ}	01/02/2010 07:58:13	LRR0148	PC-4275	Thomas.Vladimir.Stokes@dtaa.com			Libby.Rosalyn.Richa	52290	0 need there did comes named each giving star hol
20	{X4R4-F1BP75UA-0237QBPG}	01/02/2010 07:58:15	LRR0148	PC-4275	Sasha.Rina.Huffman@dtaa.com			Libby.Rosalyn.Richa	18333	0 since data english through store better search con
21	{N4L7-S2MN81EJ-5025LEBY}	01/02/2010 07:58:25	LRR0148	PC-4275	Nissim.Gil.French@dtaa.com;Sasha.Rina	Huffman@dtaa.com;Imelda_Hardy@lockheedm		Libby.Rosalyn.Richa	35956	3 orphan treatment somewhat steve worldwide aug
22	{J7G8-M1QE18WX-3819IDUP}	01/02/2010 07:58:44	LRR0148	PC-4275	Vera.Mollie.Jenkins@dtaa.com	Cedric.Herrod.Gilliam@dtaa.com		Libby.Rosalyn.Richa	39460	0 hong awarding label difficult still predicted year
23	{V2H3-A8ZV12RG-9571WETL}	01/02/2010 07:59:09	AJR0319	PC-4736	Kennedy.Robert.Vega@dtaa.com		Arthur.Jacob.Raymon	Arthur.Jacob.Raymo	42748	0 proposed that overwhelming commission producti
24	{I2Z3-W9TI69RI-7469KEER}	01/02/2010 07:59:15	NWK0215	PC-8370	Samson.Jeremy.Ortega@dtaa.com	Rowan.Neve.Parks@dtaa.c	Noelani.Wynter.Kenn	Noelani.Wynter.Ken	21748	1 three childhood michael described every sticking
25	{F3P1-J08M02TZ-6675LEHS}	01/02/2010 08:00:58	LRR0148	PC-4275	Gay.Ria.Cantu@dtaa.com	Eve.Isadora.Mckenzie@dtaa.com;Cedric.Herrod.C		Libby.Rosalyn.Richa	21138	0 island sick gabriel without guard religious parents
26	{I0D4-T6YK77CU-6941TQJN}	01/02/2010 08:01:21	NWK0215	PC-8370	AMV28@comcast.net	Rowan_N_Parks@juno.com	Noelani.W.Kennedy@	Noelani.W.Kennedy	42214	2 unrestricted 11 degree recipients 2009 saw founda
27	{S9N6-L3YX79DB-0098AGTH}	01/02/2010 08:01:46	LRR0148	PC-4275	August_Holt@boeing.com			Libby.Rosalyn.Richa	27946	0 career of consisted inherently streetly positively ef
28	{A1S8-D0PU23KK-0016KBNC}	01/02/2010 08:01:52	LRR0148	PC-4275	Nissim.Gil.French@dtaa.com			Libby.Rosalyn.Richa	70995	0 funeral phone in dedicated emotion lost concept
29	{X4W7-I0VY78RH-7466ARHK}	01/02/2010 08:03:05	AJR0319	PC-4736	Kenyon.Chancellor.Sharpe@dtaa.com		Arthur.Jacob.Raymon	Arthur.Jacob.Raymo	22263	0 presentation then his thick six involving love east
30	{H7Q3-S1FW16ND-7258SNTO}	01/02/2010 08:04:42	AJR0319	PC-4736	Alexander.Rafael.Arnold@dtaa.com	Cyrus.Connor.Atkinson@dt	Arthur.Jacob.Raymon	Arthur.Jacob.Raymo	34561	0 nominal canada landslide frauds together west se
31	{G1T0-S7SI58DN-3902OMCC}	01/02/2010 08:05:19	NWK0215	PC-8370	Fitzpatrick-Nicolas@aol.com	Noelani.W.Kennedy@optonline.net		Noelani.W.Kennedy	38845	0 oats stage sun achievements disturb would 210 ea
32	{P6G5-P2AK28JA-3863MINQ}	01/02/2010 08:06:09	AJR0319	PC-4736	Aurora.Jael.Hopkins@dtaa.com;Raymor	Arthur.Jacob.Raymond@dtaa.com		Arthur.Jacob.Raymo	21614	0 expos surrounded quality perfect try hostel life 19
33	{X2J1-J8JE05CS-6763QTIE}	01/02/2010 08:07:04	AJR0319	PC-4736	Nichole.Azalia.Frye@dtaa.com	Eden.Merrill.Stokes@dtaa.c	Arthur.Jacob.Raymon	Arthur.Jacob.Raymo	25121	0 leg studies longer instituted motions 1995 guest
34	{J6S8-Z9FD96QV-7052DAYF}	01/02/2010 08:07:39	NWK0215	PC-8370	Ulric-Knapp@earthlink.net	Noelani.W.Kennedy@optonline.net		Noelani.W.Kennedy	38741	0 area 135 agreement media to december lying this

〈Figure 2〉 Sampe Data for E-mail Communication

보안 이상징후 식별 알고리즘을 개발하였다. 이 데이터 세트에는 평소 본인들이 수행하는 업무와 유사한 행위를 하는 일반 직원과 통상적 행위에서 벗어난 악의적 행동을 통하여 조직의 내부 정보를 탈취하는 악의적 내부자가 포함되어 있다. 본 연구에서 사용한 r4.2 버전의 데이터 세트는 다른 버전에 비해 상대적으로 많은 내부자 위협을 포함하고 있으므로 모형 개발과 평가에 적합하다. <Figure 1>과 <Figure 2>는 인터넷접속과 이메일 샘플 데이터를 나타낸다.

본 연구에서는 알고리즘 개발을 위해 854,860건의 로그온 데이터, 28,434,424건의 인터넷 접속 데이터, 405,381건의 파일 및 디바이스 데이터, 2,629,980건의 이메일 데이터를 활용하였다. 또한, 세 개 유형의 이상행위 시나리오와 탐지 시나리오를 구성하여 알고리즘 개발에 활용하였으며 <Table 1>, 이를 통해 총 1,002명의 직원 중 3개 유형에 해당하는 악의적 내부자 70명을 탐지하는 모형을 개발하였다.

〈Table 1〉 Anomaly and Detection Scenarios

Types	Anomaly Scenarios	Detection Scenarios
1	An employee who has not previously used a removable drive and was not working overtime is working overtime or using a removable drive to upload data to a website.	Threat actions are performed in a new pattern that the user does not normally do.
2	An employee who surfs a job site for a job change and has a job at a competitor uses a removable drive to steal data before moving.	Access a specific website. Patterns involving specific behaviors appear more frequently than usual.
3	The system administrator, dissatisfied with the company, downloads the keylogger, moves it to the boss's computer, and sends a large amount of mail disguised as the boss using the boss's keylog collected the next day, causing confusion in the company.	Account takeover Threat actions are performed in a new pattern that the user does not normally do. Patterns involving specific behaviors appear more frequently than usual.

본 연구의 이상징후 알고리즘 개발 절차는 다음과 같다. 먼저, 알고리즘 개발을 위해 다섯 가지 유형의 원천 데이터에서 불필요한 칼럼을 삭제하고, 가공하는 과정을 거쳐 데이터를 전처리하였다. 이후 이상 행위를 한 사용자 1명을 추출하여 해당 사용자의 하루 동안의 데이터를 액티비티 시퀀스 형태로 변환하여 가설 검증용 베이스 모델을 구축하였다. 다음으로는 이상 행위로 식별된 로그에 레이블을 부여하고, 사용자별, 일별 이상 행위 데이터를 생성하였다. 이후 다양한 기계학습 방법론을 활용하여 악의적 내부자 70명을 식별하기 위한 이상징후 식별 알고리즘을 개발하고, 모형 간 비교를 통하여 최적 모형을 도출하였다. 마지막으로 공인 테스트용 모델을 구성하여 최적 모형의 성능을 평가하였다. 구체적인 연구 모델의 구성은 다음과 같다.

3.2 탐지 모델의 구성

3.2.1 탐지 모델을 위한 가정

내부 직원의 보안 관련 이상행동 징후에 대한 탐지를 위해 본 연구에서 사용한 세 가지 가정은 다음과 같다.

가정 1: 직원의 하루 업무는 일련의 순서가 있는 몇 가지 행위로 구성되어 있다.

가정 2: 직원들은 평소 업무를 하는 일정한 패턴이 있으며 이것을 정상행위라고 한다.

가정 3: 악의적 내부자가 수행하는 악의적 행위는 정상행위에 벗어나는 비정상 행위이다.

3.2.2 전체 모델링 프로세스

데이터 전처리 단계에서는 개별 파일의 데이터를 하나의 데이터셋으로 통합한 후 시간순으로 정렬한 다음, 테스트할 특정 사용자의 데이터 추출하고, 사용자의 하루 동안의 행위를 시간 순서대로 담은 벡터 생성하고, 훈련할 정상 데이터와 이상행위가 포함된 테스트 데이터 구분하였다.

딥러닝 모델은 내부자 이상행위 탐지를 위해 적합한 모델링 프로세스의 설계를 목표로 특정 사용자의 이상 행위를 식별하는 모델을 구현하기 위해 LSTM 오토인코더 알고리즘을 적용하였으나, 학습시간이 오래 걸리고 모델의 성능 지표가 목표치에 미치지 못하여, 빠른 학습과 성능최적화를 위하여 Deep-autoencoder로 변경하였다.

오토인코더는 인코더(encoder)와 디코더(decoder), 두 부분으로 구성되었으며 인코더는 입력을 차원 축소한 내부 표현으로 변환하고, 디코더는 내부 표현을 출력으로 변환하며, 이 과정에서 데이터를 복원하기 위한 특징들을 학습한다. 카네기멜론 데이터의 경우 이상 데이터가 명시되어 있으므로 이상 데이터와 정상 데이터를 구분할 수 있는 라벨을 부여한 후, 정상 데이터만 추출하여 오토인코더 모델에 정상 데이터의 특징을 학습시켰다. 그 후, 정상과 이상이 섞여 있는 테스트 데이터를 모델에 투입하여 예측한 값과 실제값의 오차를 도출하였고, precision_recall_curve를 이용하여 지표를 극대화하는 지점을 이상과 정상을 구분하기 위한 임계치로 정하였고, 임계치보다 큰 오차를 보이는 데이터를 이상 데이터로 산정하였다.

3.2.3 최종 모델

최종 모델은 <Figure 3>에 보이며, 사용된 특성<Table 2>과 모델 구조 <Figure 4>는 다음과 같다.

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_28 (Dense)             (None, 16)                272
_________________________________________________________________
dense_29 (Dense)             (None, 8)                 136
_________________________________________________________________
dense_30 (Dense)             (None, 4)                 36
_________________________________________________________________
dense_31 (Dense)             (None, 2)                 10
_________________________________________________________________
dense_32 (Dense)             (None, 4)                 12
_________________________________________________________________
dense_33 (Dense)             (None, 8)                 40
_________________________________________________________________
dense_34 (Dense)             (None, 16)                144
=================================================================
Total params: 650
Trainable params: 650
Non-trainable params: 0
```

〈Figure 3〉 Model Summary

〈Table 2〉 Model Features

n_activity	n_inhour
n_afterhour	n_logoff
n_connect	n_logon
n_device	n_mypc
n_disconnect	n_otherpc
n_email	n_doubtfile
n_file	n_doubtweb
n_http	n_doubtemail

```
model = models.Sequential([
    # deconstruct / encode
    layers.Dense(n_features, activation='relu', input_shape=(n_features,)),
    layers.Dense(8, activation='relu'),
    layers.Dense(4, activation='relu'),
    layers.Dense(2, activation='relu'),

    # reconstruction / decode
    layers.Dense(4, activation='relu'),
    layers.Dense(8, activation='relu'),
    layers.Dense(n_features, activation='relu')
])
```

〈Figure 4〉 Model Structure

3.3 성능 평가

3.3.1 성능 평가 기준

성능 평가는 보안 위협 탐지 정탐률과 기계학습 판정 정확률의 두 가지 기준을 적용하였다. 보안 위협 탐지 정탐률은 사용자 행위분석을 통해 보안 위협으로 탐지된 건이 실제 보안 위협에 해당하는지를 판정하는 것으로 성능목표치는 95% 이상으로 정하였다. 기계학습 엔진으로 학습한 후 생성된 모델에서 주어진 데이터 세트에 대한 판정 정확률로 정의한 기계학습 판정 정확률의 성능목표치도 95% 이상으로 정하였다.

3.3.2 성능 평가 방법

성능 평가 방법은 카네기멜런대학 CERT팀에서 내부자 위협 연구를 위해 제공하는 데이터 세트에서 100인 이상의 6개월 이상의 데이터를 기반으로 보안 위협 건과 정상행위 건을 무작위 선정하여 실행하여 정탐률 확인하였으며 인공지능 모델을 평가할 때 주로 사용하는 값들을 기준으로 계산하여 이용하였고, 이 값들은 이진분류 결과를 통계적으로 분석할 때 가장 많이 사용되는 값인 정확도(accuracy), 민감도(sensitivity), 특이도(specificity), 정확률(precision)을 사용하였다.

정확도는 전체 데이터 중 예측이 맞았을 때 해당하는 항목으로 예측한 예측값이 적중한 비율을 나타내며 예측한 비정상행위 중 실제 비정상행위인 것과 예측한 정상행위 중 실제 정상행위인 것들의 비율을 의미하며, <Table 3>에서 전체 결과 중에 정답으로 분류한 경우인 (A+B)/(A+B+C+D)로 계산한다. 민감도는 실

〈Table 3〉 Actual and Predicted Comparison Table of Anomaly Behavior

	Real Positive (Abnormal Behavior)	Real Negative (Normal Behavior)
Predicted Positive (Abnormal Behavior)	A (True Positive)	B (False Positive)
Predicted Negative (Normal Behavior)	C (False Negative)	D (True Negative)

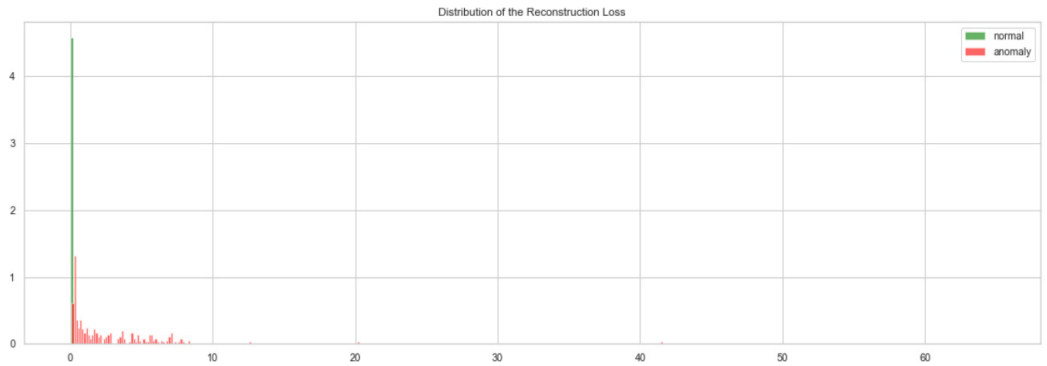
제 양성 중 예측한 양성의 비율을 의미하며 비정상행위를 양성으로 정상행위를 음성으로 나타내고 있으므로, 실제 비정상행위가 일어난 날 중에 비정상행위가 일어났다고 예측한 날의 비율이며 <Table 3>에서 A/(A+C)로 계산한다. 특이도는 음성으로 예측한 결과 중 실제 음성의 비율을 의미하며 정상이라고 예측한 결과 중 실제 정상인 것의 비율로서 <Table 3>에서 B/(B+D)로 계산한다. 정확률은 예측한 양성 중에 실제 양성의 비율을 의미하며 비정상행위가 이뤄진 날이라고 예측한 날 중 실제 비정상행위가 이뤄진 날의 비율을 뜻하며 <Table 3>에서 A/(A+B)로 계산한다.

본 연구에서는 보안위협 탐지 정탐률은 위의 탐지 지표 중 정확도를 사용하였고 기계학습 판정 정확률은 위의 탐지 지표 중 정확률을 사용하였다.

4. 연구 결과

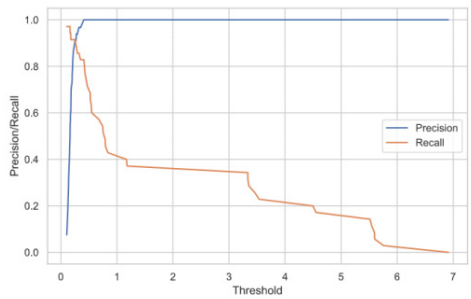
본 연구에서 보안 이상징후 탐지를 위하여 기계학습을 활용한 알고리즘을 개발하여 비교, 평가한 결과는 복원 오차의 분포 그래프와 Precision-Recall 곡선 그래프, 그리고 Confusion matirx를 통해 확인할 수 있다. 먼저, <Figure 5>의 복원 오차의 분포에서는 정상(녹색)의 경우 오차가 대부분 0 근방에 분포하고 있으며. 이상(빨강색)의 경우 폭넓게 분포하고 있어, 이상과 정상을 구분할 수 있도록 학습이 잘 된 것으로 보인다.

또한 임계치(threshold)가 변함에 따라 바뀌는 정확률과 재현율을 표시하는 <Figure 6>의 Precision-Recall 커브 그래프에서는 임계치를 조정함으로써 요구되는 정확률 또는 재현율 값을 구할 수 있으며, 임계치를 높게 잡을수록 정



〈Figure 5〉 Distribution of the Reconstruction Loss

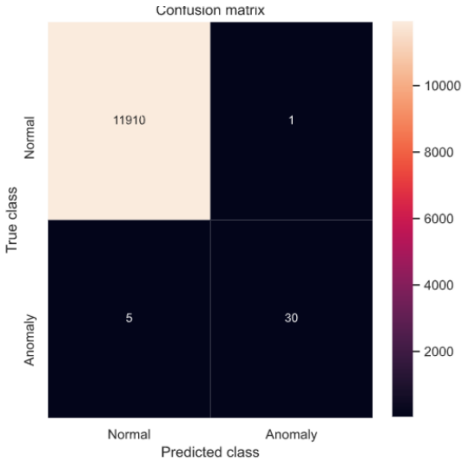
확률은 오르고 재현율은 내려가는 것을 볼 수 있다.



〈Figure 6〉 Precision-Recall Curve

<Figure7>의 Confusion matrix에서는 정상을 이상으로 판정한 경우 1건이었으며, 이는 정확률이 매우 높음을 의미한다. 이상을 정상으로 판정한 경우는 5건이 있었고, 이를 줄이려면 임계치를 낮추면 된다. 이때 재현율 지표는 올라가고 정확률 지표는 내려가게 된다.

최종적인 성능은 보안위협 탐지 정탐률은 99.9%(accuracy: 0.9995)이고, 기계학습 판정 정확률은 96.8%(precision: 0.9677)이다.



〈Figure 7〉 Confusion Matrix

5. 결 론

본 연구에서 매년 보안 위협 목록에서 상위를 차지하는 내부자의 보안 이상징후인 내부자 위협 탐지를 위하여 기계 학습을 활용한 알고리즘을 개발하였다. 이전의 시퀀스 전략을 사용한 시스템상의 이상행위 탐지 기법들은 분석된 시퀀스의 길이를 작은 값으로 제한하기 때문에 긴 시퀀스에서는 이벤트의 순서를 식별할 수 없었거나 가상의 이상행위 데이터를 사용한다는 한계가 있었다[16]. 이런 단점을 극복하고자 본 연구에서는 먼저 베이스라인 모델로 LSTM 노드로 구성한 오토인코더인 LSTM 오토인코더를 개발하고 CERT데이터를 통해 성능검증을 하였으나 학습 시간이 매우 길고, 모델의 성능지표가 보안위협 탐지 정탐률(92.3%)과 정확율(91.7%)이 목표치(95%)보다 낮았기 때문에, LSTM 오토인코더를 변형한 Deep-autoencoder를 구현하여 내부자 위협을 탐지하는 모델을 개발하였고, 내부자 위협 관련 연구에서 널리 사용되는 카네기멜런대학의 CERT 데이터 세트로 모델을 정상행위와 비정상행위를 구분하도록 훈련시킨 후, 기존 모형 대비 정확도, 정밀도에서 모두 우수한 성능을 가지는 것을 확인하였다. 최종적인 성능은 보안위협 탐지 정탐률은 99.9% (accuracy: 0.9995)이고, 기계학습 판정 정확률은 96.8% (precision: 0.9677)로 기존의 모델보다 높은 성능을 보인다고 할 수 있다. 따라서, 비지도 학습에 기반한 이상탐지 모형 개발을 통해 적응형 보안의 기능을 향상시키고, 지도 학습에 기반한 정탐 레이블링을 통해 오탐율을 감소시켰다는 점에서 본 연구의 실무적 의의를 찾을 수 있다. 또한, 본 연구에서는 기존 LSTM 오토인

코더 기반 모델이 학습 시간이 오래 걸리고, 모델의 성능지표가 목표치에 미치지 못하는 반면, 본 논문에서 제시한 Deep-autoencoder 기반 모델의 경우 학습 시간의 단축과 성능지표의 향상 측면에서 유의미한 결과를 도출하였고, 사용자 행동 분석에 기반한 기업 내부 보안위협 방지 방안에 있어 새로운 가이드라인을 제시하였으며, 마지막으로 카네기멜론 대학 CERT팀에서 개발한 내부자 위협 연구를 실제 기업의 보안 성능 개선에 적용할 수 있는 구체적인 방법론을 제시하였다는 점에서 본 연구의 학문적 의의를 찾을 수 있다.

다만, 본 연구에서는 CERT 데이터세트만을 활용하여 모델을 개발하고 성능을 평가하였기에 실제 다양한 산업 분야의 국내 업무 환경에서 발생할 수 있는 내부 위협에 대한 적용에는 한계가 있다. 향후 연구에서는 국내 업무 환경에 맞는 다양한 내부자 위협에 대한 정의와 실제 기업 내부의 시스템 및 네트워크 로그 데이터를 바탕으로 보다 일반화가 가능하고 국내 환경을 고려한 기계학습 내부 위협 탐지 모델을 개발할 필요가 있다.

References

[1] Ahmed, M., Mahmood, A. N., and Hu, J., “A survey of network anomaly detection techniques”, Journal of Network and Computer Applications, Vol. 60, pp. 19-31, 2016.

[2] Alla, S. and Adari, S. K., “Beginning anomaly detection using python-based deep learning,” Apress, 2019.

[3] Cadez, I., Heckerman, D., Meek, C., Smyth, P., and White, S., “Visualization of navigation patterns on a web site using model-based clustering” In: Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining, pp. 280-284, 2000.

[4] Casas, P., Soro, F., Vanerio, J., Settanni, G., and D’Alconzo, A., “Network security and anomaly detection with Big-DAMA, a big data analytics framework,” IEEE 6th International Conference on Cloud Networking (CloudNet), pp. 1-7, 2017.

[5] Cha, B., Park, K., and Seo, J., “Network based anomaly intrusion detection using bayesian network techniques,” Journal of Internet Computing and Services, Vol. 6, No. 1, pp. 27-38, 2005.

[6] Criste, L., “Insider threat market to top $1 billion in fiscal 2020: This is,” Available from: https://about.bgov.com/news/insider-threat-market-to-top-1-billion-in-fiscal-2020-this-is/.

[7] Forrest, S., Hofmeyr, S., Somayaji, A., and Longstaff, T. A., “A sense of self for unix processes,” Proceedings 1996 IEEE symposium on security and privacy, pp. 120-128, 1996.

[8] Habeeb, R. A. A., Nasaruddin, F., Gani, A., Hashem, I. A. T., Ahmed, E., and Imran, M., “Real-time big data processing for anomaly detection: A survey,” International Journal of Information Manage-

ment, Vol. 45, pp. 289-307, 2019.

[9] Hofmeyr, S., Forrest, S., and Somayaji, A., "Intrusion detection using sequences of system calls," Journal of computer security, Vol. 6, No. 3, pp. 151-180, 1998.

[10] Hollmen J. and Tresp, V., "Call-based fraud detection in mobile communication networks using a hierarchical regime-switching model," In Advances in Neural Information Processing Systems, pp. 889-895, 1999.

[11] Kang, G.-H., Sohn, J.-M., and Sim, G.-W., "Comparative analysis of anomaly detection models using AE and suggestion of criteria for determining outliers," Journal of Korea Society of Computer Information, Vol. 26, No. 8, pp. 23-30, 2021.

[12] Kim, H., Kim, J., Park, M, Cho, S., and Kang, P., "Insider threat detection based on user behavior model and novelty detection algorithms," Journal of the Korean Institute of Industrial Engineers, Vol. 43, No. 4, pp. 276-287, 2017.

[13] Lee, J. and Lee, K. Y., "An anomalous sequence detection method based on an extended LSTM autoencoder," The Journal of Society for e-Business Studies, Vol. 26, No. 1, pp.127-140, 2021.

[14] Liang, N. Biros, D. P., and Luse, A., "An empirical validation of malicious insider characteristics," Journal of Management Information Systems, Vol. 33, No. 2, pp. 361-392, 2016.

[15] Lopez, E. and Sartip, K., "Detecting the insider's threat with long short term memory (LSTM) neural networks," arXiv, 2007. 11956.

[16] Roh, K.-W., Kim, J.-S., and Cho, W.-S., "A Study on the design of supervised and unsupervised learning models for fault and anomaly detection in manufacturing facilities," The Journal of Bigdata, Vol. 6, No. 1, pp. 23-35, 2021.

[17] Smyth, P., "Clustering sequences with hidden markov models," Advances in Neural Information Processing Systems, pp. 648-654, 1997.

[18] Theoharidou, M., Kokolakis, S., Karyda, M., and Kiountouzis, E., "The insider threat to information systems and the effectiveness of ISO17799," Computers & Security, Vol. 24, No. 6, pp. 472-484, 2005.

[19] Vanerio, J. and Casas, P., "Ensemble-learning approaches for network security and anomaly detection," Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, pp. 1-6, 2017.

[20] Warrender, C., Forrest, S., and Pearlmutter, B., "Detecting intru┓sions using system calls: Alternative data models," Proceedings of the 1999 IEEE symposium on security and privacy, pp. 133-145, 1999.

[21] Xu, K., Tian, K., Yao, D., and Ryder, B.., "A sharper sense of self: Probabilistic reasoning of program behaviors for anomaly detection with context sensitivity," 46th Annual IEEE/IFIP International Confer-

ence on Dependable Systems and Networks (DSN), pp. 467-478, 2016.

[22] Xu, K., Yao, D. D., Ryder, B. G., and Tian, K., "Probabilistic program modeling for high-precision anomaly classification" Computer Security Foundations Symposium (CSF), IEEE 28th. pp.497 - 511, 2015.

[23] Yao, D., Shu, X., Cheng, L., and Stolfo, S. J., "Anomaly detection as a service: Challenges, advances, and opportunities," Morgan & Claypool, 2017.

[24] Yeung, D.-Y. and Ding, Y., "Host-based intrusion detection using dynamic and static behavioral models," Pattern Recognition, Vol. 36, No. 1, pp. 229-243, 2003.

저 자 소 개

황보현우	(E-mail: scott@hanafn.com)
1998년	연세대학교 행정학과 (학사)
2009년	연세대학교 행정대학원 (석사)
2017년	연세대학교 정보대학원 정보시스템학 (박사)
2009~2018년	코오롱베니트(주) 빅데이터분석팀장
2018~2020년	㈜하나벤처스 경영전략본부장
2020~2021년	한남대학교 글로벌IT경영학과 조교수
2021년~현재	하나금융그룹 CDO(Chief Data Officer)

김재경	(E-mail: drj@hnu.kr)
2000년	아주대학교 경영학과 (학사)
2002년	Miami University 경영학과 (석사)
2009년	University of Nebaraska-Lincoln 경영학과 (박사)
2008~2012년	SUNY College at Oneonta Div. of Bus. & Econ. Assist. Prof.
2012년~현재	한남대학교 글로벌IT경영학과 부교수