	ぞるざや へるれんちめぐ
氏　　名	**Zolzaya Kherlenchimeg**
本籍（国籍）	モンゴル
学位の種類	博士(工学)
学位記番号	工博　第３１７号
学位授与年月日	令和２年３月２３日
学位授与の要件	学位規則第５条第１項該当　課程博士
研究科及び専攻	工学研究科デザイン・メディア工学専攻
学位論文　題目	**A Study of Effective Framework Combining Sparse Autoencoder Based Feature Transfer Learning and Long Short-Term Memory for Network Intrusion Detection System** **（ネットワーク侵入検知システムのためのスパースオートエンコーダベースの特徴転移学習と長・短期記憶を組み合わせた効果的なフレームワークの研究）**
学位審査委員	主査　教　授　　今野　晃市
	副査　教　授　　藤本　忠博
	副査　准教授　　明石　卓也

論　文　内　容　の　要　旨

The rapid growth of technology uses all over the world, our daily lives and activities for the better in many ways. However, this exponential growth of interconnections has led to also concern network security issues. A vulnerability and potential malicious threat might be due to a bug in applications and ill-managed networks. Therefore, we must address these critical issues of network security such as detect suspicious activities, a countermeasure against intruders and unauthorized access to the existing data. In the last decades, the Intrusion Detection System (IDS) plays a vital role in detecting network attacks. The IDS is a process of monitoring and analyzing the events occurring in a computer system and network to detect signs of security problems.

In general, IDS categorized into misuse-based detection and anomaly-based detection. A misuse-based IDS also known as a signature-based IDS that measures its similarity between input and signatures of known attacks. Therefore, the known attacks can be detected immediately and reliably with a lower false-positive rate. While the misuse-based detection method has disadvantages that it cannot detect unknown attacks and novel attacks. The

anomaly-based detection technique is the process of comparing activity the enterprise considers normal against observed activity to identify significant deviations. The advantage of anomaly-based detection techniques is suitable to predict and adapt to unknown attacks. This kind of detection method uses a machine learning approach to create a predictive model by simulating regular activity and known activity, then compare new behaviors with the existing model. However, an anomaly-based IDS usually produces a high percentage of false alarms rate and a low rate of detection rate, it might be effect the efficiency of real-world applications. In an anomaly-based IDS, there are different levels at which an IDS can monitor activities in a network. It faces a large number of features representation of monitored network traffic. The high dimensionalities of the network traffic give to raise the hypothesis search space and also lead to large classification errors. Therefore, to address those problems, this study focused to improve the accuracy of detecting unknown attacks through the developing an effective framework.

In this study, we propose a network IDS framework for intrusion detection. The proposed framework consists of two stages. The first stage is a feature extraction stage which has two steps: unsupervised pre-training and supervised fine-tuning. The first step is an unsupervised pre-training step that learns the typical patterns of the network traffic using a single-layer Sparse Autoencoder (SAE) which is an effective learning algorithm for reconstructing a new feature presentation of the data through the nonlinear mapping. Consequently, the second step is a supervised fine-tuning step that can extracts the primary features of the network traffic using the preceding optimal parameters in supervised manner while gradually reduce the data dimension. The SAE model determines an approximation to the identity function, so as to output data that is similar to their input data. In other words, the function involves finding the optimal network parameters weight, biases by minimizing the discrepancy between input and its reconstruction data. However, the degree of input features increases the model becomes more complex and has to fit all data. Therefore, to prevent the problem of overfitting, we use the L2 regularization method by augmenting the cost function with the sum of the squared magnitude of all weights in the network. As well as, we regularize the feature extractor model by using a Kullback-Leibler (KL) divergence as a sparsity penalty term which constrains the neurons to be inactive most of the time.

Accordingly, we train a single-layer feature learning SAE model on training

set only in unsupervised manner using 5-fold cross-validation, while optimize hyperparameter values of the network. It involves finding the optimal network parameters weight, biases and hyperparameters of cost function. After the network learned optimal values for weights and biases, save the network parameters. Once selecting proper hyperparameter, re-train the feature extractor SAE model using these optimal hyperparameter on the training set with a label. Finally, the feature extractor SAE model can extract the new feature representations which represent the source data. In the next stage, train a Long Short-Term Memory model to identify the network traffic as being either normal or attack using the extracted new feature representations dataset. We apply the 10-fold cross-validation method to validate the results on the LSTM model to prevent overfitting issue. In final, we evaluate the effectiveness of the proposed IDS framework on the public NSL-KDD benchmark dataset. The experimental result shows that the proposed framework performs better than previous studies which proves the effectiveness of our framework. Furthermore, the result confirmed that our feature extractor SAE model significantly effected to improve the performance of this work.

This dissertation was aimed to develop an effective framework combining a single-layer Sparse Autoencoder (SAE) based feature transfer learning and Long Short-Term Memory (LSTM). Initially, the feature extractor SAE model which proposed to extract the most relevant features for use in representing the data. In the following, the LSTM method proposed for classifying network traffic either a normal or an attack. The result of the proposed framework detected network attack with high accuracy and it outperformed other similar studies.

論 文 審 査 結 果 の 要 旨

本論文は，ネットワーク侵入検知システム（NIDS）のためのフレームワークを提案している．本論文で提案しているフレームワークは，未知のネットワーク攻撃にも対応可能なアノマリ型の NIDS であり，スパースオートエンコーダ（SAE）と Long Short-Term Memory（LSTM）という２つの異なるニューラルネットワークの組み合わせによって新たに構成された，機械学習アルゴリズムが用いられている．

ネットワークからの攻撃は日々増加しており，NIDS の役割はますます重要なものとなっている．NIDS にはその検知方法の違いから，大きく分けてシグネチャ型とアノマリ型という２種類がある．前者のシグネチャ型は，データベースに登録された不正パターンと一致する攻撃を検知するものであるが，未知の攻撃を検知することは難しい．それに対してアノマリ型は，正常なパターンとネットワーク上のトラフ

ィックを比較し，そのパターンから外れたものを攻撃として検知するもので，未知の攻撃を検知可能である．しかし，アノマリ型はシグネチャ型と比較して検知性能に課題があることが，研究の背景として述べられている．

本論文では，転移学習を使用したSAEによりネットワークトラフィックを次元圧縮することで重要な特徴を抽出し,それをLSTMに学習させ攻撃を検知するフレームワークが提案されている．この手法は，従来と比較して検知性能を向上させることが可能であり，手法の新規性が認められる．また，実験により高い検知性能が得られていることが示されており，手法の有効性も検証されている．

本論文の構成は以下の通りである．

第１章は序論であり，研究の背景，課題，目的が述べられている．

第２章では，NIDSの一般的な説明とともに，本研究と関連する先行研究について述べられている．具体的にはアノマリ型NIDSでの正常パターン学習に，ニューラルネットワーク，k 近傍法，サポートベクターマシン等を用いたものや，深層学習による手法が紹介されている．深層学習では，自己教示学習ベースのオートエンコーダ，回帰型ニューラルネットワーク（RNN），畳み込みニューラルネットワークを用いた研究等が示されている．

第３章では，第４章で提案するフレームワークで使う深層学習手法を含む，SAE，RNN，LSTM等が解説されている．次にNIDSとして，SAEにより次元圧縮したデータを，RNN により学習することで攻撃を検知するという新たな手法が提案されている．提案された手法の有効性を確認するために，NIDSのベンチマークとして広く用いられている NSL-KDD データセットを用いた実験を行っている．実験では正確度80.0%となり，第２章で述べた先行研究と同等の性能が得られている．

第４章では，第３章で提案した手法を改良し，検出性能向上を目指した新しいNIDSフレームワークを提案している．提案されたフレームワークは２つのステージで構成されている．1st ステージではSAEによるネットワークトラフィックデータの次元圧縮が行われる．ここで，正常・不正といったラベルのないデータを使ったSAEの教師なし学習の結果を，教師あり学習を行うためのSAEに転移学習することで学習効果を高めている．この点は特に独創的な部分として評価できる．次の 2nd ステージでは，1st ステージで次元圧縮されたデータをLSTMに学習させ，トラフィックデータを正常と不正に分類するネットワークを得ることで攻撃を検知している．提案したフレームワークについての実験では正確度84.8%となり，第２章で述べた先行研究の正確度よりも高い検知性能が得られたことが述べられている．これは，提案フレームワークの有効性を示すものである．

第５章は結論であり，本論文をまとめるとともに，今後の課題が述べられている．

以上，本論文は，転移学習を用いたSAEによる次元圧縮を行うことでトラフィックデータの特徴を抽出し,その圧縮されたデータでLSTMを学習することで攻撃を検知可能とする，新しいNIDSフレームワークを提案したものである．提案フレームワークは既存手法と比較して検知性能が高く，ネットワークセキュリティの分野の発

展に寄与するところが大である.

よって，本論文は博士（工学）の学位論文として合格と認める.

原著論文名（１編を記載）

A Deep Learning Approach Based on Sparse Autoencoder with Long Short-Term Memory for Network Intrusion Detection, Zolzaya Kherlenchimeg, Naoshi Nakaya, IEEJ Transactions on Electronics, Information and Systems, (Vol.140, No.6), 令和２年６月発行予定