Using Convolutional Neural Networks to Network Intrusion Detection for Cyber Threats

Wen-Hui Lin[1*], Hsiao-Chung Lin, Ping Wang, Bao-Hua Wu, Jeng-Ying Tsai

[1]Department of Information Management,
Kun Shan University, Tainan, Taiwan
[1]linwh@mail.ksu.edu.tw,
fordlin@mail.ksu.edu.tw,
pingwang@mail.ksu.edu.tw
weq498aa@gmail.com,
Jamestsay1207@gmail.com

Abstract

In practice, Defenders need a more efficient network detection approach which has the advantages of quick-responding learning capability of new network behavioural features for network intrusion detection purpose. In many applications the capability of Deep Learning techniques has been confirmed to outperform classic approaches. Accordingly, this study focused on network intrusion detection using convolutional neural networks (CNNs) based on LeNet-5 to classify the network threats. The experiment results show that the prediction accuracy of intrusion detection goes up to 99.65% with samples more than 10,000. The overall accuracy rate is 97.53%.

Keywords: Intrusion detection, Deep Learning, Convolutional neural networks, Behavior features, LeNet-5

Introduction

Most existing approaches for detecting cyber-attacks involve cyber-threat analyses to match the potential attack profiles by filtering malicious connections to assist defenders in analyzing the attack scenarios. Generally, maliciously attempt to compromise network breaches using phishing websites or operation system updates with legal network protocols including http, ICMP, and SSL, to pass through the firewall, virus detection engine and download the malicious applications by the users into the hosts which were controlled by remote controllers. In practice, it is difficult to discriminate legitimate or malicious connections using protocol analysis. In further, information gain approach is widely accepted as a qualitative approach to detect traffic anomaly in information flows, and distinguish the connection was benign or malicious for unknown threats, but it cannot precisely classify threats in response to cyber-attacks. It raises the problem of false positive rate in detecting the malicious connections using available flow analysis analyses.

To improve the classification accuracy of threat detection and reduce its false positive rate for NIDS, this study developed an improved behaviour-based classifier learning model for network anomaly detection by training an CNNs (Convolutional Neural Networks) to extract the enhanced behaviour features and identify the class of threats with Softmax function by using collected statistical data. In practice, information flows of network intrusion were obtained from the network nodes in NCHC (National Center for High-performance Computing), and these flows were categorized by clustering analyses in order to categorize the behavioural features from the different IPs for comparisons. And then the feature vectors are transformed into the feature matrices to form the images as the inputs of the CNN to accurately categorize cyber threats according to collected behavioural features derived from packet analyses of network traffic, enabling the defence system to quickly respond to high-risk threats.

The rest of this article is organized as follows. Section II reviews previous studies on CNNs. Section III presents the proposed approach for network intrusion detection model with CNN architecture in an online information security management system. The experimental results are presented in Section VI. Finally, Section V concludes the work.

Relate Work

The convolutional neural network (CNN) is a type of deep, feed-forward artificial neural networks which learn features that are concatenated with the original feature vectors and used for classification [1]. To improve classification accuracy in real-time detection, the CNN was regularly selected using the combinational convolution and pooling operations by several fully connected or sparsely connected layers followed by a final classification layer to determine the abstract weights of an input data. Then the Softmax function with fully connected layers, which was used for categorizing new classes. In particular, CNNs use relatively little pre-processing compared to other classification algorithms. In this paper, we explore a deep learning architecture with CNNs model for network intrusion detection based on abnormal behaviours. Generally, a CNN consists of an image input and a classification output layer, as well as multiple hidden layers. The hidden layers are either convolutional, rectified linear unit (ReLU), pooling, or fully connected which are illustrated as follows. [1-2].

To improve the classification accuracy of threat detection and reduce its false positive rate for NIDS, this study developed an improved behaviour-based classifier learning model for network anomaly detection by training an CNNs with TensorFlow developed by Google to extract the enhanced behaviour features and identify the class of threats with Softmax function by using collected statistical data。

ISBN 978-1-5386-4342-6

Network Intrusion Detection Model with Deep Neural Networks

In [3], the ID3 decision tree theory is used as a scheme of feature reduction for speeding up the learning of normal and intrusive pattern for NIDS. The ID3 algorithm begins with the original set *S* as the root node. In each iteration of the solution procedure, the algorithm will repeat through every unused attribute of the set *S* and calculates the information gain IG(*S*) of that attribute. It can effectively assist the defenders determine the required attribute which has the largest information gain value.

A behaviour-classification approaches for network intrusion detection model with enhanced features is presented. Detailed workflow from suspicious network flows for behaviour classification using the revised LeNet-5 model is shown in Fig. 1; Fig. 1 illustrates the revised LeNet-5 model incorporating three subphases in the behaviour-classification process for NIDS: i) the feature extraction phase, ii) the model training phase, and iii) the model verification phase. [3]

A. Feature extraction phase.

In feature extraction phase, the training sample data were obtained from two data source: i) the archive dataset for behavioural features download from KDD'Cup99 (http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html), Basically, KDD' Cup 99 dataset contains total 41 features where 34 features are numeric and 7 features are symbolic or discrete. [4] ii) suspicious network flows captured by NHSNC to extract the new behavioural features.

Step 1.1 Model learning using KDD'Cup99 dataset
First, a series of experiments were performed to investigate the of the CNN-based classifier effectiveness using a benchmark KDD'Cup99 dataset, where the learning results were regarded as a basis of model parameters including weight matrix, batch_size, batchs_per_epoch, epochs and accuracy of classification.

Step 1.2 New behaviour features extracted from recent network flows

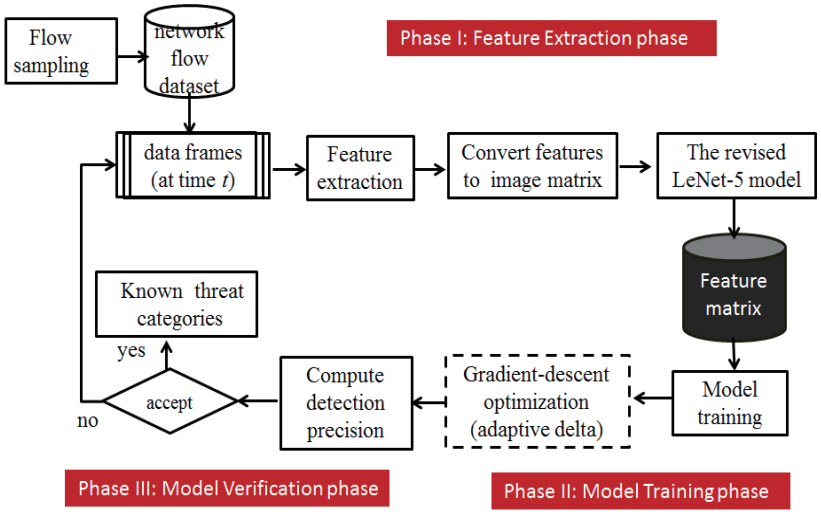

Figure 1. Basic concept of network intrusion detection by using LeNet-5 model.

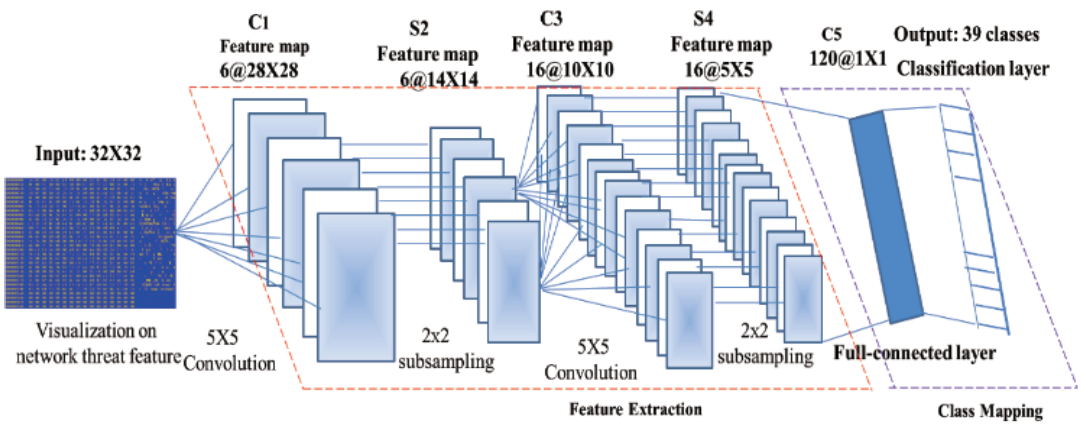

Figure 2. The revised LeNet-5 model for network threat classification

ISBN 978-1-5386-4342-6

B. Model learning phase.

In model learning phase, the present study revised the LeNet-5 model designed by LeCun et al. in 1998 and illustrated as shown in Fig.2 and Table 1. The proposed model also incorporates the gradient-descent optimization algorithm (i.e., adaptive delta algorithm) to fine-tune the model parameters for features reinforce learning using error derivatives of back-propagation with the learning rate for all layers. That is, the classification is used to determine the learning error of multiple layer neural nets and then adjust the weights of neural nets to minimize it in the learning process of CNNs.

Table 1. The revised LeNet-5 model architecture

Layer	**Function description**
convolution layer C_1	Apply 5x5 size filter to a 32 x 32 x 3 input images derived from image matrix of behavior features and generates 6 feature maps of size 28x28.
pooling layer S_2	Pooling layer S2 on the 6@28x28 feature maps resulting in 6@14x14 pooled feature maps
convolution layer C_3	It applies 5x5 size filter a 14 x 14 x 6 input images. This results in 16 feature maps of size 10x10
pooling layer S_4	The subsampling layer S4 takes input feature maps of size 10 × 10 and generates 16 feature maps of size 5x5.
fully-connected layer(FCL) C_5	Apply 5x5x1920 filters to convert sixteen feature maps of size 5x5 into a fully-connected vector of size 120×1.
Classification layer	Output: the numbers of the classes is 7 for Case I . Output: the numbers of the classes is 40 for Case II .

C. Model validation phase.

We adopt a cross-validation scheme to evaluate the predicted accuracy of CNN model for overcoming over-training problem by using various *n*-folds of the cross-validation scheme; for example, $k = 8$ means that 70% of the dataset collected was used in the training experiment, and the remaining 10% of the dataset was used for alternative testing repeated 8 times. In model validation phase, the system provides the benefit of quick-responding for threat classification thru the use of weights of neural nets using the trained CNNs in model learning phase.

Experimental Results

In this section, the performance of the proposed CNN-based intrusion detection model is demonstrated by means of two cyber security examples for intrusion detection.

Step 1. Feature extraction phase

To examine the model efficiency, the first example incorporates a revised LeNet-5 model with KDD'Cup99 dataset where LeNet-5 model is composed of one or more convolutional layers with FCLs, similar to typical artificial neural networks where model categorizes input images into 40 classes.

In the experiment, feature extraction involved a pre-process including two steps: i) feature reduction, and ii) convert features to an image matrix.

Step 1.1 Feature reduction

The features of the NSL-KDD 1999 Dataset were outrank according to the score assigned by information gain (IG) measure. To improve classification speed, a set of reduced feature was regularly selected using information gain (IG) scheme where the number of feature selected (32 at IG ≥0.119). A set of reduced features selected from top 32 ranked features using IG approach are listed in Table 2. [3]

Table 2 Ranked features using information gain

Rank	Feature	Rank	Feature
1	srv_serror_rate	17	dst_host_srv_diff_host_rate
2	serror_rate	18	root_shell
3	flag	19	wrong_fragment
4	logged_in	20	dst_host_diff_srv_rate
5	dst_host_srv_serror_rate	21	dst_host_srv_count
6	diff_srv_rate	22	rerror_rate
7	dst_host_serror_rate	23	count
8	dst_bytes	24	urgent
9	hot	25	protocol_type
10	dst_host_same_srv_rate	26	dst_host_srv_rerror_rate
11	src_bytes	27	dst_host_count
12	same_srv_rate	28	dst_host_same_src_port_rate
13	srv_diff_host_rate	29	num_file_creations
14	service	30	num_shells
15	num_failed_logins	31	num_compromised
16	is_guest_login	32	num_root

Step 1.2 Convert features to an image matrix

This step is to preprocess for the experiment data, including (i) the symbol conversion of the network packets and (ii) normalization for numeric data, and (iii) convert features to an image matrix Finally, input images were normalised to a size of 32 × 32 pixels, were used in the following experiments. This study randomly selected 80% as a training dataset, the rest 20% as a testing, where both data sets wholly contain 39 sub-categories of attack samples to avoid classification bias.

Step 2. Model learning phase

Two cases of different purposes were conducted to verify the effectiveness of CNN-based classifier as follows.

Case I. Classification of major types of intrusion threats using large amounts of samples

First, conduct an experiment for classifying major intrusion threats by selecting the attack type which has large amounts of samples, (i.e., more than 100,00 samples of captured data) in training process, the statistical results of threat types for KDD cup'99 for the following training and test are illustrated in Table 3. In table 3, only six major types of threats were screened to be classified in the experiment.

ISBN 978-1-5386-4342-6

Table 3. A statistical analysis for threat types in KDD cup'99 dataset (>10,000 records)

Attacks	Kddcup.data.corrected	No. of records	Training data	Test data
smurf	2807886	15000	12000	3000
neptune	1072017	15000	12000	3000
normal	972781	15000	12000	3000
satan	15892	15892	12713	3178
ipsweep	12481	12481	9984	2496
portsweep	10413	10413	8330	2082

In the following, the CNN was trained to detect network intrusion using the trained weights of networks. The model prediction accuracy is 99.65 %.

Table 4. Accuracy associated using different n-fold of cross-validation scheme (Case I)

n-fold	Accuracy (%)
k=4	99.60%
k=8	99.65%
k=10	99.63%

Case II. Classification for 39 sub-categories of attacks

In the experiment, CNNs are used for classifying data into 39 sub-categories according to distinct behaviours. In KDD cup'99 dataset, the number of sample for a specific threat type is generally less than 500, it must be evenly and randomly replicated to 500 samples to avoid too few data that produces the learning bias compared to large samples of the attack type. Table 5 shows that the corresponding accuracy (%) by using the cross-validation method (k = 4,8,10). For Case II, the average accuracy is 95.41 %.

Table 5. Accuracy associated using different n-fold of cross-validation scheme (Case II)

n-fold	Accuracy (%)
k=4	95.07%
k=8	95.51%
k=10	95.40%

Step 3. Model validation phase

Experimental results show that the classification error decreased as the size of the testing dataset increased. For different sizing of data samples (large amount and few amount), the prediction accuracy of threat classification increases with an increasing *N* value. The prediction accuracy of intrusion detection increased to 99.65% when $N \geq 10{,}000$. The overall accuracy rate was 97.53%. In summary, the prediction accuracy of threat detection was higher with the CNN-based classifier than with the existing LeNet-5 model (approximately 95%) of the schemes proposed in [5-9] and malicious code identification [10] in existing in deep learning schemes.

Conclusion

This paper presents an intrusion detection model based on a CNNs-based classifier for enhancing the precision of model. Importantly, the proposed approach revises the LaNet-5 model with the adaptive delta optimization algorithm to fine-tune the model parameters and minimize a classification error by using error derivatives of back-propagation and quick response to intrusion detection using Tensor flow. The proposed method improves the accuracy of intrusion detection for threat classification by using enhanced behaviour features from trained CNNs. Overall, the proposed approach can enhance the precision for network intrusion detection.

Acknowledgement

This work was supported jointly by the Ministry of Science and Technology of Taiwan under Grant Nos. MOST 106-3114-E- 492 -001 and MOST 106-2410–H -168-002.

References

[1] Wikipedia, Convolutional neural networks, available at https://en.wikipedia.org/wiki/ convolutional neural networks

[2] P. Wang, W.H. Lin, K.M. Chao, C.C. Lo. A face-recognition approach using deep reinforcement learning approach for user authentication, 2017 IEEE 14th International Conference on e-Business Engineering (ICEBE), 2017.

[3] P. Wang, K. M. Chao, H. C. Lin. An Efficient Flow Control Approach for SDN-based Network Threat Detection and Migration Using Support Vector Machine. The 13th IEEE International Conference on e-Business Engineering (ICEBE 2016), Macau University, 4-6 Nov. 2016.

[4] University of California, Irvine (UCI), KDD Cup 1999 Data, available at https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[5] L. Khan, M. Awad, B. Thuraisingham, A new intrusion detection system using support vector machines and hierarchical clustering, The VLDB Journal, 16(4), 507–521, 2007.

[6] L. Li, Z.P. Gao, W.Y. Ding, Fuzzy multi-class support vector machine based on binary tree in network intrusion detection, 2010 International Conference on Electrical and Control Engineering (ICECE), 25-27 June 2010.

[7] X. Guan, H. Guo, L. Chen, Network intrusion detection based on agent and SVM, .The 2nd IEEE International Conference on Information Management and Engineering (ICIME), pp. 16-18 April 2010.

[8] N. Kausar, B. B. Samir, S. B. Sulaiman, I. Ahmad, M. Hussain, An approach towards intrusion detection using PCA feature subsets and SVM, 2012 International Conference on Computer & Information Science (ICCIS), 12-14 June 2012.

[9] S. Singh, J. P. Singh, G. Shrivastva, A hybrid artificial immune system for IDS based on SVM and belief function, Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE 2013.

[10] Y. Wang, W. D. Cai, P. C. Wei, A deep learning approach for detecting malicious javascript code, Security & Communication Networks, 51(8), 28656-28667, 2016.

ISBN 978-1-5386-4342-6