Cyber Intrusion Detection by Using Deep Neural Networks with Attack-sharing Loss

IEEE DataCom '19

Boxiang Dong[1] Hui (Wendy) Wang[2] Aparna S. Varde[1]
Dawei Li[1] Bharath K. Samanthula[1]
Weifeng Sun[3] Liang Zhao[3]

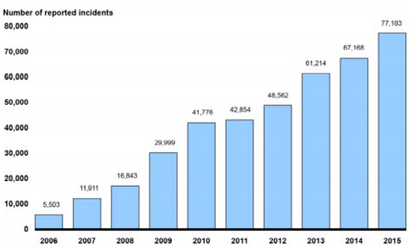
[1]Montclair State University
Montclair, NJ, USA

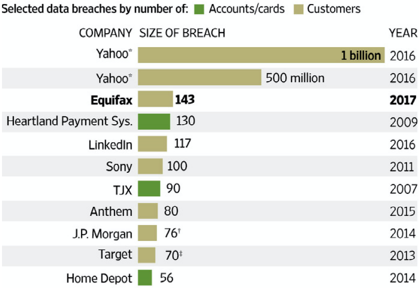
[2]Stevens Institute of Technology
Hoboken, NJ, USA

[3]Dalian University of Technology
Dalian, China

November 19, 2019

Cyber Attacks





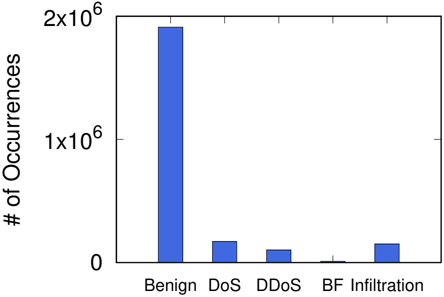
- The number of reported cyber incidents increased by 1,300% in the past 10 years.
- The amount of disclosed information in these attacks are outrageous.

Intrusion Detection Systems

- Intrusion Detection System
  - Signatured based — Need continuous inputs from security experts
  - Machine learning based — Fail to capture complex attack patterns
    - graph mining
    - SVM
    - decision tree
  - Deep learning based — Only focus on constructing features
    - autoencoder
    - CNN

Our Objective Employ deep learning to discover inherent features and learn complex classification function.

Challenges



Diversity of attacks There are quite a few types of attacks, which exhibit different behavior patterns.

Imbalanced class distribution
- A majority of the network connections are benign.
- Different types of intrusion attacks are unevenly distributed in practice.

Our Contributions

We build a new intrusion detection and classification framework named *DeepIDEA*, (a Deep Neural Network-based Intrusion Detector with Attack-sharing Loss).

- DeepIDEA takes full advantage of deep learning to extract features and cultivate classification boundary.
- DeepIDEA incorporates a new loss function (named *attack-sharing loss*) to cope with the imbalanced class distribution.
- Experiments on three benchmark datasets demonstrate the superiority of DeepIDEA.

Outline

Related Work

Intrusion detection based on deep learning

- Self-taught learning [JNSA16]
- Few-shot learning [CHK+17]
- Auto-encoder [MDES18]

Anomaly detection based on deep learning

- LSTM [ZXM+16, DLZS17]
- CNN [KTP18]

Preliminaries - Intrusion Attacks

In this paper, we focus on detecting the following five prevailing attacks.

Brute-force Gain illegal access to a site or server.

Botnet Exploit zombie devices to carry out malicious activities.

Probing Scan a victim device to determine the vulnerabilities.

Dos/DDoS Overload a target machine and prevent it from serving legitimate users.

Infiltration Leverage a software vulnerability and execute backdoor attacks.

Preliminaries - Imbalanced Classification

- The labels in intrusion detection datasets follow a long tail distribution.
- The imbalanced data forces the classification model to be biased toward the majority classes
- It renders poor accuracy on detecting intrusion attacks.

Over-sampling duplicate under-represented classes.
- overfitting
- long training time

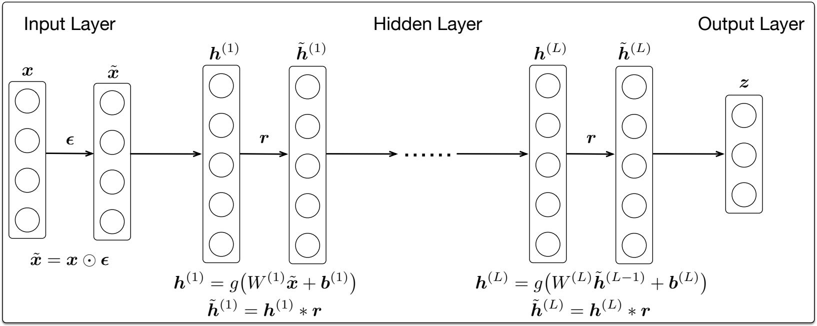
Under-sampling eliminates samples in over-sized classes.
- inferior accuracy

Cost-sensitive learning associate high weight with under-represented classes.
- non-convergence in training

Our Solution - DeepIDEA

DeepIDEA employs a fully-connected neural network to classify network connections.

- L hidden layers with ReLU units and dropout;
- One output layer with softmax activation function.



Our Solution - DeepIDEA

A classic loss function for classification models is cross-entropy loss, J_{CE}, s.t.

$$\begin{aligned} J_{CE}(\boldsymbol{\theta}) &= \mathbb{E}_{(\boldsymbol{x}^{(i)}, y^{(i)}) \sim \hat{p}_{data}} L(f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \\ &= -\mathbb{E}_{(\boldsymbol{x}^{(i)}, y^{(i)}) \sim \hat{p}_{data}} \log p(y^{(i)} | \boldsymbol{x}^{(i)}; \boldsymbol{\theta}) \\ &= -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{c} \mathbf{I}(y^{(i)}, j) \log p_j^{(i)}, \end{aligned}$$

$$\mathbf{I}(a, b) = \begin{cases} 1 & \text{if a=b} \\ 0 & \text{otherwise.} \end{cases}$$

Our Solution - DeepIDEA

A classic loss function for classification models is cross-entropy loss, J_{CE}, s.t.

$$
\begin{aligned}
J_{CE}(\boldsymbol{\theta}) &= \mathbb{E}_{(\boldsymbol{x}^{(i)}, y^{(i)}) \sim \hat{p}_{data}} L(f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \\
&= -\mathbb{E}_{(\boldsymbol{x}^{(i)}, y^{(i)}) \sim \hat{p}_{data}} \log p(y^{(i)} | \boldsymbol{x}^{(i)}; \boldsymbol{\theta}) \\
&= -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{c} \mathbf{I}(y^{(i)}, j) \log p_j^{(i)},
\end{aligned}
$$

- However, the underlying assumption of J_{CE} is that all instances have the same importance.
- In case of imbalanced class distribution, it lets the classifier concentrate on the majority class.
- As a consequence, the neural network tends to simply classify every instance as benign.

Our Solution - DeepIDEA

Two types of classification error

- **Intrusion mis-classification** An intrusion attack is mis-classified as benign event;
- **Attack mis-classification** An intrustion attack of type A (e.g., DoS attack) is mis-classified as an intrusion attack of type B (e.g., probing attack).

Our intuition Intrusion mis-classification should be penalized more than the attack mis-classification, as it enables the cyber incidents to by-pass the security check and cause potentially critical damage.

Our Solution - DeepIDEA

We design attack-sharing loss, J_{AS}.
For any instance $(\boldsymbol{x}^{(i)}, y^{(i)})$, let $y^{(i)}$ be 1 if it is benign; let $y^{(i)} \in \{2, \ldots, c\}$ otherwise.

$$J_{AS} = \underbrace{\boxed{-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{c}\mathbf{I}(y^{(i)}, j)\log p_j^{(i)}}}_{\text{cross-entropy loss}} - \lambda \underbrace{\boxed{\left(\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{I}(y^{(i)}, 1)\log p_1^{(i)} + \sum_{j=2}^{c}\mathbf{I}(y^{(i)}, j)\log(1 - p_1^{(i)})\right)\right)}}_{\text{additional penalty for class mis-classification}},$$

where $\lambda > 0$ is a hyper-parameter that controls the degree of additional penalty.

Our Solution - DeepIDEA

Advantage of attack-sharing loss

- Eliminates the bias towards the majority/benign class by moving the decision boundary towards the attack classes; and
- Respects the penalty discrepancy of different types of mis-classification.

Experiments - Dataset

Three Benchmark Datasets

- *KDD99* dataset
- *CICIDS17* dataset [1]
- *CICIDS18* dataset [2]

Class Imbalance Measure Ω_{imb}

$$\Omega_{imb} = \frac{\sum_{i=1}^{c} n_{max} - n_i}{n}$$

Dataset	# of Features	Training Size	Testing Size	# of Classes	Ω_{imb}
KDD99	41	4,898,431	311,029	5	2.96
CICIDS17	81	2,343,634	482,926	5	3.08
CICIDS18	77	5,080,071	1,063,342	4	2.31

[1] `https://www.unb.ca/cic/datasets/ids-2017.html`

[2] `https://www.unb.ca/cic/datasets/ids-2018.html`

Experiments - Dataset

Table: Class distribution in CICIDS17 dataset

Label	Training		Testing	
	Number	Fraction	Number	Fraction
Benign	1,911,674	81.57%	361,399	74.84%
DoS	170,508	7.27%	82,151	17.01%
DDoS	101,024	4.31%	27,003	5.59%
Brute-Force	10,494	0.45%	3,341	0.69%
Infiltration	149,934	6.40%	9,032	1.87%
Total	2,343,634	100%	482,926	100%

Experiments - Baselines

SVM

KNN $k = 5$, minkowski distance

DT 10 layers at most

MLP+CE deep feedforward network with cross-entropy loss function

MLP+OS [JS02]

MLP+US [KM+97]

Cost-Sensitive cost-sensitive loss function [KHB+18]

CNN [KHB+18] 2 convolution layers, 2 maxpooling layers and 6 fully-connected layers

Experiments - Setup and Metrics

Setup

- Implemented by using Tensorflow
- 10 hidden layers, 100 units per layer
- 0.8 keep probability in dropout layers
- Batch size: 128
- Training on a NVIDIA RTX 2080 Ti GPU within 3 hours

Evaluation Metrics

- Measure precision and recall for each class
- Evaluate the average class-wise recall as the overall class-balanced accuracy (CBA) [DGZ18].

Experiments

Detection Accuracy on CICIDS17 Dataset

Classifier	Benign		DoS		DDoS		Brute-Force		Infiltration		CBA
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	
SVM	86.42	76.38	96.58	53.74	92.62	16.03	0	0	7.27	86.18	46.47
KNN	91.92	85.05	75.88	48.22	72.56	86.23	0	0	10.92	84.75	60.85
DT	66.51	100	0	0	0	0	0	0	0	0	20
MLP+CE	87.04	90.76	74.12	63.69	74.73	79.53	7.37	4.8	28.03	61.54	60.06
MLP+OS [JS02]	86.03	95.05	80.14	52.5	56.68	76.06	3.65	1.63	28.18	53.62	55.45
MLP+US [KM+97]	86.88	54.9	50.91	59.31	26.13	11.32	7.17	27.39	13.8	58.03	42.19
Cost-Sensitive [KHB+18]	61.58	61.17	17.69	28.09	0	0	0	0	0	0	17.85
CNN [CHK+17]	0	0	23.42	96.04	0	0	8.07	11.07	0	0	21.42
DeepIDEA	88.5	94.06	88.77	62.97	76.31	83.19	8.29	4.1	26.46	64.53	61.77

- DeepIDEA produces similar and satisfying precision and recall on every class, except for Brute-Force.
- DeepIDEA yields the highest CBA, meaning that it reaches the best balance among all classes.

Conclusion

In this paper, we design DeepIDEA to detect network intrusion attacks, which

- takes full advantage of deep learning for both feature extraction and attack recognition; and
- copes with the imbalanced class distribution by using attack-sharing loss function.

In the future, we aim at extending our work by

- utilizing a more advanced model such as RNN; and
- improving the performance on the extremely under-represented classes.

References I

[CHK+17] Md Moin Uddin Chowdhury, Frederick Hammond, Glenn Konowicz, Chunsheng Xin, Hongyi Wu, and Jiang Li.
A few-shot deep learning approach for improved intrusion detection.
In *IEEE Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, pages 456–462, 2017.

[DGZ18] Qi Dong, Shaogang Gong, and Xiatian Zhu.
Imbalanced deep learning by minority class incremental rectification.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[DLZS17] Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar.
Deeplog: Anomaly detection and diagnosis from system logs through deep learning.
In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pages 1285–1298, 2017.

[JNSA16] Ahmad Javaid, Quamar Niyaz, Weiqing Sun, and Mansoor Alam.
A deep learning approach for network intrusion detection system.
In *Proceedings of the EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pages 21–26, 2016.

[JS02] Nathalie Japkowicz and Shaju Stephen.
The class imbalance problem: A systematic study.
Intelligent Data Analysis, 6(5):429–449, 2002.

[KHB+18] Salman H Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A Sohel, and Roberto Togneri.
Cost-sensitive learning of deep feature representations from imbalanced data.
IEEE transactions on neural networks and learning systems, 29(8):3573–3587, 2018.

References II

[KM+97] Miroslav Kubat, Stan Matwin, et al.
Addressing the curse of imbalanced training sets: one-sided selection.
In *Icml*, volume 97, pages 179–186. Nashville, USA, 1997.

[KTP18] B Ravi Kiran, Dilip Mathew Thomas, and Ranjith Parakkal.
An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos.
Journal of Imaging, 4(2):36, 2018.

[MDES18] Yisroel Mirsky, Tomer Doitshman, Yuval Elovici, and Asaf Shabtai.
Kitsune: an ensemble of autoencoders for online network intrusion detection.
arXiv preprint arXiv:1802.09089, 2018.

[ZXM+16] Ke Zhang, Jianwu Xu, Martin Renqiang Min, Guofei Jiang, Konstantinos Pelechrinis, and Hui Zhang.
Automated it system failure prediction: A deep learning approach.
In *IEEE International Conference on Big Data*, pages 1291–1300, 2016.

Q & A

Thank you!

Questions?

dongb@montclair.edu