International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Transfer Learning Based Prototype for Zero-Day Attack Detection

Nerella Sameera, M. Shashi

Abstract: ***Intrusion detection System (IDS) is an evolving research area in cyber security, which aims to detect cyber-intrusions. Machine Learning, especially deep learning classifiers, offers promising solutions for signature-based intrusion detection provided there are abundant labeled examples. However, effectiveness of deep learning is hindered for zero-day attack detection due to lack of labelled examples; anomaly-based detection approaches often result in high FPR. Transfer Learning (TL) offers methodologies for building classifiers in a target domain containing minimal or no labelled data, leveraging the knowledge extracted from related source domain(s). When applied to zero-day attack detection, Transfer Learning models known attack data as source domain and descriptions of zero-day attacks as target domain with possible differences either in the feature space or in the proportions of attack to normal instances or both. The authors built a TL-based prototype using NSL-KDD dataset for experimentation on unification of feature space for detecting unlabeled R2L samples representing zero-day attacks from normal instances using labelled DoS samples. The proposed TL based classifier achieved 89.79% accuracy and 0.15% FPR which is higher than the state-of-the-art methods.***

Index Terms: Intrusion, Source Domain, Target Domain, Transfer Learning, Transformation, Zero-day attack

I. INTRODUCTION

Advancements and rapid technical developments in information-age have promoted computerization of day to day activity in all walks of life invariably leading to words vulnerabilities in cyber security issues. Intrusion Detection Systems (IDS) "unpublished" [1] plays a key role in protecting cyber space by detecting malicious activities. IDS is an evolving research area in the field of cyber security, which aims to detect cyber-intrusions. Machine learning offers many promising solutions for signature-based intrusion detection; especially deep learning architectures have proven capabilities to extract even the most complex patterns provided that there is sufficiently large collection of labeled examples. However, due to lack of such a large collection of labelled examples for zero-day attacks, effectiveness of deep learning for detection of zero-day attacks is hindered. Transfer learning (TL) [2] offers promising solutions to handle this problem. TL is a recent advancement of machine learning that involves creation of high-performance learners for a target domain with or without labeled examples leveraging training experience in related source domain. Researchers proved that performance of the models built using TL is on par with models built by traditional machine learning algorithms even if the TL was provided with no labeled or with only one to ten percent of the labelled training examples. TL bridges the gap between the source and target domain specifically in three aspects namely feature space distribution differences, instance distribution density differences and differences in label spaces. When applied to zero-day attack detection, TL handles the differences in feature spaces, marginal probability distributions with class labels among the attacks whose signatures are already captured, and the new zero-day attacks whose signatures are to be captured with minimal or no labelled examples describing them. So, with the help of TL, model constructed on related source domain can be refined for detection of labels for the new domain, even though it (new domain) may have a different scenario, thereby increasing the detection accuracy of emerging intrusions with reduced FPR. This paper focuses on constructing an efficient intrusion detection system by making use of Transfer Learning through knowledge sharing and model refinement. The authors have experimented on the unification of feature space in order to handle the differences in feature spaces for detecting unlabeled R2L samples representing zero-day attacks from normal instances using labelled DoS samples from NSL-KDD dataset. The remaining part of the paper is organized as follows: Section II presents background and basic concepts of TL, Section III discusses about the related work, Section IV presents the proposed TL methodology and result analysis. Finally, conclusion and future research is mentioned in Section V.

II. TRANSFER LEARNING CONCEPTS

Labeled data in certain domains is scarce due to expensive data collection process, difficulty in manual labeling process and cold start problem. Due to inadequate supply of labeled instances in such domains traditional machine learning algorithms are not successful to build classifier models. In such situations transfer learning techniques are called for to learn a classifier in a target domain with limited or no labeled examples making use of a large collection of labeled examples available in a related source domain. Transfer Learning provides the ability to leverage the knowledge and skills gained from a prevalent source domain while modeling a general task for learning a novel task in a related specific/target domain. TL is formally defined [3] as given below:

Given a source domain D_S and learning task T_S, a target domain D_T and learning task T_T, transfer learning aims to improve the learning of the target predictive function $f_T(.)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$.

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
Nerella Sameera*, Computer Science & Systems Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam (AP), India.
M. Shashi, Computer Science & Systems Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam (AP), India.

Both target and source domains/ tasks are represented with the common formalism denoted as: $D = \{\chi, P(X)\}$ and $T = \{Y, f(\cdot)\}$ respectively where χ is the feature space, P(X) is the marginal probability distribution of objects, X and Y is the label space of a domain.

There are two dichotomies of transfer learning; while the first one is based on the availability of the labelled data the other one is based on the differences in feature spaces. Within the first category there are three types of TL [3]: Inductive TL, Transductive TL and Unsupervised TL. If limited labeled data is available in target domain then it is Inductive TL irrespective of the availability of labeled data at source domain. If there is no labeled data at target domain whereas labeled data is available for source then it is Transductive TL. If there is no labeled data at both source and target domain then it comes under Unsupervised TL. Based on the feature space there are again two types of TL [2]: If source and target domain have the same feature space then it is called homogeneous TL otherwise it is called heterogeneous TL.

III. RELATED WORK

The potential of Transfer Learning was already proved in the areas like Computer Vision (CV) [4] and Natural Language Processing (NLP). ELMo [5], ULMFiT [6], OpenAI [7], BERT [8], etc. are the recent developments in NLP using TL. Modern researchers are exploring the applicability of Transfer Learning in the field of IDS, which are mentioned below: Zhaoet al. in [9] comes with a new TL method called HeTL which is an improvement on HeMap [10]. Here the problem is modeled as a binary classification problem. Initially both source and target data are transformed into a common latent space via spectral transformation that preserves the original structure of the data, while at the same time, maximizing the similarity between the two, and then classification is performed on top of these latent features. The authors derived three datasets from NSL-KDD dataset and performed experimentation by making one of the datasets as source dataset and another one of remaining two as target dataset. Work done by Vercruyssenet al. in [11] is based on two assumptions that anomalies are infrequent and exabits unexpected behavior. Instance-based transfer learning is used here which reweights instances from source domain to match with target domain for performing time series anomaly detection. The problem is modeled as a binary classification problem. Two decision functions named “density-based transfer decision function” and “cluster-based transfer decision function” are proposed to make a decision about transfer of an instance. Zahra Taghiyarrenaniet al. in [12] proposes new TL method that maps the source and target datasets into a common feature space by following manifold alignment technique which uses four matrices namely similarity matrix, dissimilarity matrix, source structural matrix and target structural matrix. Once the problem is modeled as a binary classification problem SVMs are used to classify the target instances into attack verses normal classes. KDD 99 and Kyoto2006+ are the datasets used for performing experimentation.

Ahmadi et al. in [13] made an attempt to detect DOS attacks on cloud by leveraging attack knowledge from a non-cloud environment and making use of common features from both source and target domains. NSL-KDD and CIDD data sets are used as cloud and non-cloud datasets respectively. TL method used here is Relation based TL which transfers knowledge based on manually identified pre-defined relation between source and target domains. Naive Bayes classifiers used for attack detection.

IV. TL-BASED PROTOTYPE FOR INTRUSION DETECTION

Zero-day attack data is unlabeled or scarcely labeled and hence detection of zero-day attacks calls for transfer learning where in classifiers are built in unlabeled target domains making use of a related source domain with sufficiently large number of labelled instances. Hence the proposed methodology falls into the category of transductive TL. Since the motive behind creation of zero-day attacks is to make it undetectable from the known attack signatures, zero-day attacks have a different set of relevant features from those of the known attacks. Hence the proposed methodology also falls into the category of heterogenous TL.When the problem is modeled as a TL problem, the zero-day attack data belongs to target domain while the labeled examples of known attacks constitute related source domain and there will be difference in the feature space of the source and target domains hence heterogenous transductive TL is applied for this problem.

The proposed TL-based prototype for zero-day attack detection aims at detecting the unlabeled attacks (zero-day attacks) and models the problem as the binary classification problem. Methodology of the proposed approach is presented in the Fig.1.As shown in the figure, source and target domains which are having heterogenous feature space are initially combined. Later these combined domains are undergone through orthogonal transformation by applying Principle Component Analysis (PCA) using eigen values and eigen vectors. With this both the domains will come into the same feature space, which is the required condition for applying classifier. Source train data and target test data should be extracted from this combined transformed data. These extracted train and test sets are submitted to the classification which will output labels (attack/normal) for target data instances.

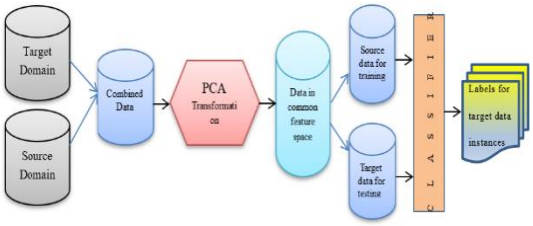

Fig 1: Block Diagram of the Proposed TL Method

A. Data set preparation

The proposed method uses the NSL_KDD dataset [14] which is the bench mark dataset proposed by many researchers for implementing IDS. This dataset contains attack instances belonging to four groups of attacks namely DoS, Prob, R2L and U2R. Two sub datasets of DoS and R2L attacks are derived from the NSL_KDD. The sub datasets are derived in such a way that DoS dataset contains 50% of the normal instances and all DoS attack instances, R2L dataset contains 20% of remaining normal instances and all R2L attack instances. Over sampling is performed on the R2L dataset to make equal size w.r.t DoS dataset.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

B. Experimentation and Result Analysis

For the purpose of experimentation and model evaluation a minimal collection of labelled data is required form the test set. Classifier is built based on the training set containing a sufficiently large collection of labeled examples from a related source domain. The authors have experimented the proposed TL method by considering R2L dataset as target data set and DoS dataset as source dataset. The details are given below:

- D_T =R2L+ Normal
- D_S=DoS+ Normal
- Target label space: {R2L, Normal}
- Source label space: {DoS, Normal}

As we are aiming to detect zero-day attacks, the class labels of the R2L dataset are removed and can be kept aside for evaluating the performance of the TL method. With this, the target dataset is unlabeled and the source data set is labeled. On completion of the preparation of datasets, both DoS and R2L data sets are combined into a single data unit. PCA is applied on the combined data unit by taking ten dimensions [15] in the transformed space. From the transformed data DoS dataset is extracted as a training set and R2L dataset is extracted as a testing set. K-nearest neighbor (KNN) classifier is applied on these extracted training and testing instances by taking K value as 3. This results in classification of zero-day attacks (R2L) with accuracy of 89.79% and FPR of 0.15% which dominates the state-of-the-art methods like HeTL [9]. The corresponding results are shown in table 1. Here we are not making any comparisons with No_TL methods because we are assuming that the target data (R2L) is completely unlabeled which is not assumed in case of No_TL methods.

Table I: Accuracy of zero-day attack detection

Method	Accuracy %	Data sets used	Classifier
TL-based prototype approach	89.79	DOS → R2L	KNN
HeTL [9]	78	DOS → R2L	KNN

V. CONCLUSION

Cyber threat is growing in par with the advancements in digital age, that makes IDS to get a lot of attention now a days. Modern researchers are exploring the applicability of Transfer Learning for IDS to detect zero-day attacks and to minimize FPR's. Transfer Learning offers methodologies for building classifiers in a target domain containing minimal or no labelled data, leveraging the knowledge extracted from related source domain(s). In this paper, the authors have applied the concept of transfer learning to detect un-labelled R2L attacks (zero-day attacks) of NSL-KDD dataset by making use of labeled DoS attacks of NSL-KDD dataset and succeeded by getting an accuracy of 89.79%. and FPR of 0.15%. which is higher than the state-of-the-art methods like HeTL. The prototype classifier will be extended in our future research to handle the remaining aspects of TL for better accuracy.

REFERENCES

1. Nerella Sameera, M. Shashi, "Intrusion Detection Analytics: Comprehensive Survey", un published.
2. Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang, "A survey of transfer learning", Journal of Big Data 3.1 (2016): 9.
3. Pan, SinnoJialin, and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering 22, no. 10 2010, 1345-1359.
4. Gopalakrishnan, Kasthurirangan, Siddhartha K. Khaitan, Alok Choudhary, and Ankit Agrawal, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection" , Construction and Building Materials 157, 322-330,2017.
5. Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer, "Deep contextualized word representations", arXiv preprint arXiv:1802.05365, 2018.
6. Howard, Jeremy, and Sebastian Ruder, "Universal language model fine-tuning for text classification", arXiv preprint arXiv:1801.06146, 2018.
7. Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever, "Improving language understanding by generative pre-training", URL https://s3-us-west-2. amazonaws. com/openai-assets/researchcovers/languageunsupervised/language understanding paper. Pdf, 2018.
8. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805, 2018.
9. Zhao, Juan, Sachin Shetty, and Jan Wei Pan, "Feature-based transfer learning for network security", In MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM), pp. 17-22. IEEE, 2017.
10. Shi, Xiaoxiao, Qi Liu, Wei Fan, S. Yu Philip, and Ruixin Zhu, "Transfer learning on heterogenous feature spaces via spectral transformation", In 2010 IEEE international conference on data mining, pp. 1049-1054. IEEE, 2010.
11. Vercruyssen, Vincent, WannesMeert, and Jesse Davis, "Transfer Learning for Time Series Anomaly Detection", In IAL@ PKDD/ECML, pp. 27-36. 2017.
12. Zahra Taghiyarrenani, et,al. "Transfer Learning based Intrusion Detection", International Conference on Computer and Knowledge Engineering (ICCKE 2018), October, pp. 25-26, 2018.
13. Ahmadi, Roja, Robert D. Macredie, and Allan Tucker, "Intrusion Detection Using Transfer Learning in Machine Learning Classifiers Between Non-cloud and Cloud Datasets", In International Conference on Intelligent Data Engineering and Automated Learning, pp. 556-566. Springer, Cham, 2018.
14. Dhanabal, L., and S. P. Shantharajah. "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms", International Journal of Advanced Research in Computer and Communication Engineering 4, no. 6 446-452,2015.
15. Vasan KK and Surendiran B, "Dimensionality reduction using Principal Component Analysis for network intrusion detection", Perspectives in Science.1;8:510-2, 2016.

AUTHORS PROFILE

Mrs. Nerella Sameera has received Bachelor of Technology from Chirala Engineering College, affiliated to JNTU Kakinada in the year 2011 and Master of Technology from Andhra University in year 2013. She is currently pursuing Ph.D in the Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam since 2017. Her main research work focuses on Data Analytics for Cyber Security and Intrusion Detection Systems. She has 3 years of teaching experience.

Prof. M Shashi pursed B.E in Electrical and Electronics Engineering and M.E in Computer Science Engineering, and Ph.D. in Artificial Intelligence and Knowledge Engineering. She is currently working as a Professor in the Department of Computer Science and Systems Engineering. Her areas of interest are Data Analytics, Data Warehousing & Mining, AI, and Data Structures.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)