DART: Detecting Unseen Malware Variants Using Adaptation Regularization Transfer Learning

Hao Li[§], Zhenxiang Chen[§*], Riccardo Spolaor[†], Qiben Yan[‡], Chuan Zhao[§], Bo Yang[§]
§University of Jinan, Jinan, China. Email: czx@ujn.edu.cn
[†]Department of Computer Science, University of Oxford, UK.
[‡]Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA.

***Abstract*—Network traffic analysis has been widely used for detecting malware at a large-scale network. Nevertheless, the emerging malware variants and zero-day exploits keep posing significant challenges to malware detection systems. In this paper, we propose DART, a framework for detecting malicious network traffic based on Adaptation Regularization Transfer Learning (ARTL), which effectively copes with the unseen malware variants problem. Specifically, DART trains the adaptive classifier by simultaneously optimizing three factors: (i) the structural risk functions; (ii) the joint distribution between the known malware and unseen malware variants domains; and (iii) the manifold consistency underlying marginal distribution. In addition, DART also works with encrypted network traffic since it does not leverage information related to the packet content. We assess the effectiveness and efficiency of our proposal with a thorough set of experiments. DART achieves over 90% F-measure and 91% recall, outperforming conventional traffic classification methods and other state-of-the-art intrusion detection systems.**

***Index Terms*—ARTL, malware variant detection, traffic classification, transfer learning.**

I. Introduction

Network traffic classification plays an important role in malware detection systems and advanced firewalls. Traffic classification techniques are considered as a key component for understanding and identifying malicious network activities [1–3]. In recent years, a great effort has been devoted to applying machine learning techniques for network traffic classification based on proxy logs information. Information-based approaches rely on network traffic content-based features, such as Uniform Resource Locator (URL) path. Zero-day exploit is a specific type of malware that has only recently been discovered. Such exploits can also affect operating systems by leveraging systems vulnerability. In addition, malware variants are not developed from scratch and instead adapted from pre-existing threats with leveraging transformation approaches. Due to frequent behavior changes in unseen malware, a detector trained on outdated malicious traffic samples becomes ineffective. Moreover, static and dynamic-based approaches [4, 5] are not helpful in detecting zero-day exploits or malware variants.

Over the years, more and more malicious network behavior has evaded detection by taking advantage of the variability of the malware. On one hand, supervised learning methods have been used to detect malicious traffic [6–9], in which a classifier is trained from labeled samples of each predefined traffic class. Unfortunately, such classifier would not be able to effectively detect novel malware variants since examples of such malware are not available for model training. On the other hand, unsupervised learning methods [7] can automatically gather a group of unlabeled training flows and applies the clustering results to the training flow classifier. However, the number of clusters must be set large enough to produce effective clusters [7].

The distribution disparity issue between source and target domains can be partially solved by *domain adaptation (transfer learning)* [10]. Recently, Bartos et al. [11] proposed a network traffic representation approach across domains to detect malware variants, which can reduce distribution differences and retain the important attributes of the original data. Also, methods above-mentioned only design a transformation that can reduce the difference in the features space or marginal distributions across domains, but it does not consider the conditional distributions between the source and target domains. It is worth noting that, for malware classification, only minimizing the difference between marginal distributions of distinct domains is not enough for knowledge transfer, since the discriminative directions that separate the positive and negative samples may remain different.

In this paper, we study a novel approach to detect unseen malware variants based on adaptive regularization transfer learning (ARTL). In particular, our proposed approach relies on ARTL to map malicious samples from a source category to a target category by using unlabeled traffic information from the target category. By doing so, our proposed framework is able to successfully detect known and unseen malware variant samples of the target category. Our approach not only aim to reduce the marginal distributions between the source and the target domains but also their conditional ones.

Contribution – The major contributions of this paper are summarized as follows:

- We design and implement DART, a network traffic analysis framework that combines a feature selection method for a robust flow representation and an ARTL based algorithm to detect the malicious variants.
- We identify the optimal parameters and features to achieve effective detection performance. To evaluate our proposal, we carry out a comprehensive set of experiments and sensitivity analysis on real-world network traffic dataset.

978-1-5386-8088-9/19/$31.00 ©2019 IEEE

- We show that DART framework correctly classifies malware variants that were not included in the training data, outperforming a state-of-the-art detection approach [12].

Organization – The remainder of this paper is organized as follows. In Section II, we review the state of the art of malware detection approaches. We formulate the problem we are aiming to address in Section III and we present the DART framework in Section IV. Then, we experimentally evaluate our proposal in Section V. Finally, we draw some conclusions in Section VI.

II. Related Work

The analysis and detection of malware variants have been a trending topic in recent years. Several methods have been proposed to detect the increasing number of zero-day exploits and unseen polymorphic malware. In this section, we retrospect the most relevant work to the domain of our proposal.

Network-based methods are applied to detect malware variants or zero-day malware. Ambusaidi et al. [6] propose a mutual information based algorithm that analytically selects the optimal feature for build an Intrusion Detection System (IDS), but has a poor detection rate in Remote to User (R2L) and User-to-Root (U2R) attacks. ZASMIN by Kim et al. [13] is an early stage identification system for network attack identification. To identify unknown network attacks, the system performs suspicious traffic monitoring, attack verification, multi-form malware signature generation, and identification. Moreover, some modules of such system require a specific hardware-based accelerator, which makes it hard to implement. Comar et al. [14] use the third and fourth layers of network flows feature to detect malware which combines supervised and unsupervised learning techniques. The authors train a classifier to detect known malicious behaviors and use unsupervised learning to detect unseen malicious behaviors. While all the above approaches represent the state-of-the-art of malware detection, they do not take into account that network threats rapidly evolve which makes these methods less effective over time.

To cope with these shortcomings, Bartos et al. [11] construct invariant representation of network traffic which is inspired by transfer learning approaches for detecting unseen malware variants. Nonetheless, they only focus on designing a transformation to reduce the differences in the feature spaces across domains, without taking into account the conditional distribution problem. In other words, they do not consider the conditional distributions between the training and testing domains under the conditional shift and only leverage content-based information of the trace. In contrast, we propose DART, which only relies only on statistical feature of network traffic and it aims that reducing the difference between the source and target domains in terms of marginal and conditional distributions.

III. Problem Formulation

For network traffic classification, label for malicious or benign samples is hard to obtain. In addition, the distribution of traffic usually changes over time. For example, we can consider the malicious behavior of financial Trojan campaigns which were popular on July 2017 that evolved into a Ransomware attacks in August 2017 [15]. Hence, the marginal and conditional distribution of the training data are often different from the testing data, which complicates training process of the classifier. Table I summarizes the notations frequently used in this paper.

TABLE I
Notations and descriptions

Notation	Description
n	number of flows statistical feature
m	number of flows in the bag
J	joint distribution
P	probability distribution
Q	conditional probability distributions
ϕ	transformation function
num_s, num_t	number of representation samples in source/target domain
f	prediction function
b	number of bins
ρ	number of nearest neighbors
σ	shrinkage regularization
λ	MMD regularization
K	kernel matrix
w	classifier parameter
H	reproducing kernel Hilbert space
M	MMD matrix

N-dimensional feature vector $x \in \mathbb{R}^n$ is a network flow representation, and flow correlation is included in the traffic classification process to improve the accuracy of identification [7]. For the convenience of traffic classification, we use *"Bag of Flows" (BoF)* [7] rather than a signal flow trace to model flow correlation. A BoF can be described by $\mathrm{X} = \{x_1, x_2, ..., x_m\} \in \mathbb{R}^{m \times n}$, where x_i represents the ith flow in the bag. Each flow in the bag has the same label. A label y_i can be assigned to any bag created from the label set (malicious or benign). The bags may have a different number of flows.

In general, if two malicious network traffic domains D_s (source domain) and D_t (target domain) are different, they may have different feature spaces or marginal probability. To solve this domain adaptation problem aforementioned, we need to apply knowledge discovered from the training (source) domain into testing (target) domain. Hence, a transformation Φ is designed to transform the BoF features to a new representation ($\Phi(X) = \widetilde{\mathrm{X}}$), in which $P_s(\Phi(X_s)) \approx P_t(\Phi(X_t))$ or $P_{\mathrm{s}}(\widetilde{X}_s) \approx P_s(\widetilde{X}_t)$. We introduce a robust representation method in Section IV-B, in which convariance shift is applied to address the domain adaptation problem using data in different feature spaces.

Furthermore, malicious traffic detection task for source domain and target domain are different and they have different label spaces Y or conditional probability distributions $f(\widetilde{X}) = Q(y|\widetilde{X})$, i.e., the financial Trojan virus and Ransomware present different malicious behaviors. Hence we can formulate this problem as $Q(y_s|\widetilde{X}_s) \neq Q(y_t|\widetilde{X}_t)$. Our

proposed framework aims specifically at solving such transfer learning problem.

IV. Design of DART Framework

In this section, we describe DART, a framework that detects malicious network traffic generated by unseen malware variants. We present the components of DART framework in Figure 1. To this end, our detection framework combines data preprocessing methods and a robust feature extraction using transfer learning. In this section, we first present data preprocessing methods for network. Then, we explain the method to obtain robust statistical feature from the network flows. Finally, we describe an ARTL based model using transfer learning algorithm and illustrate some examples about DART transformation.

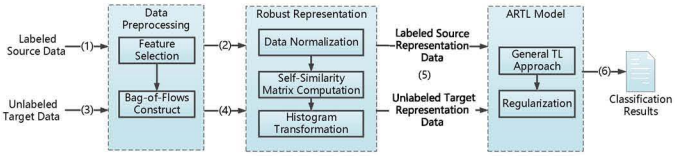


Fig. 1. DART framework.

A. Data Preprocessing

As mentioned in Section III, DART framework pre-process the traffic by generating *"Bag of Flows" (BoF)* [7] that contains the same three-tuple $< dst\_ip, dst\_port, protocol >$ to be used in the DART framework, and the trace generated from the same malware belongs to the same label. BoF enables a traffic correlation technique to make traffic classification more convenient.

Nevertheless, redundant and irrelevant features of traffic data may hamper the detection performance [6]. In this paper, we adopt a feature selection technique based on Mutual Information (MI) [6] to rank flow features based on their relevance with the corresponding class. MI measures the mutual dependence between two features based on information entropy. Formally, the mutual information of traffic features in the source domain variable (U) and target domain variable (V) can be defined as:

$$I(U;V) = \sum_{u \in U} \sum_{v \in V} p(U,V) \log \left(\frac{p(U,V)}{p(U)p(V)}\right), \quad (1)$$

where $p(U,V)$ is the joint probability function of U and V, and $p(U)$ and $p(V)$ are the marginal probability distribution functions of U and V respectively.

B. Robust Representation

A BoF is represented as X and consists of a set of m flows $\{x_1, ..., x_m\}$. s The new representation is calculated by applying a transformation approach to the BoF that consists of a series of steps to ensure its robustness in traffic classification issue.

To guarantee the robustness of the invariance representation which is transformed from BoF, we normalize feature values and create a self-similarity matrix method and histogram transformation approach [11]. First, the matrix of BoF, X, is normalized in the interval 0 to 1. Then, the resulting self-similarity matrix method will be adopted to make sure that the invariance represents a typical property (e.g., eigenvector, mean value) of the BoF. In this paper, we calculate a self-similarity matrix for each feature:

$$Z^k = \begin{pmatrix} Z^k_{11} & \cdots & Z^k_{1n} \\ \vdots & \ddots & \vdots \\ Z^k_{m1} & \cdots & Z^k_{mn} \end{pmatrix}, \quad (2)$$

where the element Z^k_{ij} is the feature values distances between x_{ik} and x_{jk} of the kth feature. All features of the set of self-similarity matrices are represented as $\widetilde{Z}$. The final step of the proposed histogram transformation method is the transition from the matrices $Xnorm$ and $\widetilde{Z}$ to normalized histograms which is an accurate representation of the distribution of the traffic data.

In the end, each network BoF in the source and target domain are represented as the robust representation $\widetilde{X}$, as follows:

$$\widetilde{X} = (\Theta(Xnorm^1), ..., \Theta(Xnorm^n), \Theta(\widetilde{Z}^1), ..., \Theta(\widetilde{Z}^n)), \quad (3)$$

where n is the number of the original flow-based features, and Θ is the BoF histogram vector computed based on predefined number of bins and bin edges which represents the probability distribution of the traffic data. We discuss and define the optimal parameters in Section VI-D.

C. ARTL Model

A robust representation $\widetilde{X}$, as a sparse matrix, represents the probability distribution of the BoF. Nevertheless, the problem of different marginal and conditional probability distributions in source and target malware domains still needs to be solved. In this section, to address the above mentioned issues, we apply the transfer learning algorithm, ARTL [10], to reduce the difference between the source and target domains in unseen malicious traffic detection framework. This model optimizes the structural risk function of the malware variants detector, joint distribution adaptation and manifold regularization between the known domain and the unseen malicious variants' domain by leveraging the kernel trick and Representer theorem [10], simultaneously.

The ARTL model is based on a semi-supervised learning method, and it can be formalized can be written as follows:

$$\begin{aligned} f \;=\; \underset{f \in H_K}{arg\,min} \; & \sum_{i=1}^{num_s} (y_i - f(\widetilde{x}_i))^2 + \sigma ||f||^2_K \\ & + \lambda D_{f,K}(J_s, J_t) + M_{f,K}(P_s, P_t), \end{aligned} \quad (4)$$

where the first half part of Equation (4) is the basic malware variants classifier trained from the source domain and $f = w^T \phi(\widetilde{x})$ is the malware variants prediction function. H_K is a set of predictors in the network kernel space and $||f||^2_K$ is the squared norm of the malware variants prediction function in H_K. Furthermore, the second half part of the formula is

the regularization term that controls the performance of the classification model transferring between source and target domains, $D_{f,K}(J_s, J_t)$ is the joint distribution between the known source and the unseen malicious target domain and $(y_i - f(\widetilde{x}_i))^2$ is the squared loss for *Regularized Least Squares (RLS)*. $M_{f,K}$ is the consistency manifold regularization of the source and target network flows data and mainly impacted by the number of ρ-nearest neighbors which is the regularization parameter of the $M_{f,\,K}$.

D. Training and Testing Procedures

For malware variants detection problem, there exist both labeled known malicious traffic data and unlabeled traffic data, and our objective is to learn an adaptive network traffic classifier from the labeled source (known malicious trace) domain D_s to detect unseen traffic samples of the target domain D_t. Leveraging the source and target domains traffic samples, feature space transformation, joint distribution domain adaptation, and *manifold regularization* [10] methods are adopted for better function learning. In this paper, DART leverages standard classification algorithms (i.e., RLS) as the loss function, we train the optimal adaptive classifier f that can identify the unseen malicious flows (testing domain) effectively and efficiently.

E. Illustrative examples

In this section, we illustrate how the DART transformation approach works with two real-world malware variants from Denial-of-Service (DoS) and Backdoors malware families.

Figure 2 summarizes the joint distribution adaptation process of the DoS and Backdoors malware categories. DoS is an attack that aims to exhaust the network resources of a host connected to the Internet. Backdoor attack is a technique that bypasses normal authentication or encryption to access a device. Hence, DoS and Backdoors belong to different malware categories and present different malicious network behaviors. This means that their network flows have different marginal and conditional probability distributions. As shown Figure 2, we are not able to classify the Backdoors malware effectively by training a model from the labeled flows of DoS. In this case, the joint distribution adaptation and manifold regularization method map the training (DoS) and testing (Backdoors) network flow features to the reproducing kernel Hilbert space (RKHS), and minimize the MMD distribution distance which makes these distributions more closer to enhance the effectiveness of the classification model.

V. Evaluation

We evaluate our DART on a real-world dataset which contains various types and numbers of malware variants samples. In this section, we first provide the details of the experimental setup and the specification of datasets, then we discuss the feature selection and parameter optimization procedures. Finally, we present the detection performance on a set of experiments and compare with other approaches.

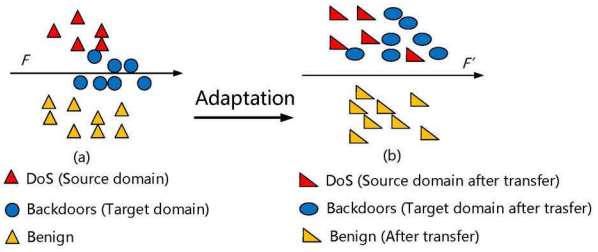


Fig. 2. DoS and Backdoors data distribution adaptation process. F: traditional machine learning classifier; F': ARTL classifier. The shape changes are caused by the mapping adaptation of ARTL.

TABLE II
UNSW-NB15 DATA SET DISTRIBUTION

Category	Samples	Percentage
Analysis	2,151	0.117%
Backdoors	1,795	0.097%
DoS	15,186	0.823%
Exploits	39,117	2.121%
Fuzzers	19,196	1.041%
Generic	207,959	11.277%
Reconnaissance	12,228	0.663%
Shellcode	1,288	0.070%
Worms	150	0.008%
Benign	1,542,015	83.621%
All	1,841,084	100%

A. Dataset & Experimental Setup

In our experiment, we use the UNSW-NB15 dataset [16]. The UNSW-NB15 dataset is composed of around 100 Gigabytes in CSV file format and it contains approximately 1.8 million flows of network packets. It includes ten different classes: one benign and nine types of malware (i.e., Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Fuzzers for anomalous activity, Shellcode and Worms). The overview of this dataset is shown in Table II. The malware family of Worms is not considered in our experiments due to its insufficient number.

Malware samples referred to as "positive bags", where one positive bag is a set of flows from the same source to the same destination. The bags that are not labeled as malicious are considered as benign. Each bag should contain at least three flows to be able to compute a meaningful histogram representation. Training and testing data are composed of completely different malware families, in other words, for each iteration (one for each family), we pick a single family to be part of the training set, and all the other families to be part of the testing set. For example, we use Exploits as the source domain and other seven malware categories as the target domains. This experimental design aims to simulate the occurrence of a new unseen threat created by malware developers to evade the detection of state-of-the-art methods.

B. Feature Selection

Comprehensive experiments are carried out for feature selection which is stated in Section IV-A, and 16 most relevant features are ultimately selected by mutual information-based

feature selection algorithm. List of the selected featured are presented in Table III. We underline that these 16 features are all flow-based statistical features (e.g., mean of the flow packet size transmitted by the source) and none of them is related to flow contents. In contrast to the proposal in [11], DART guarantees high detection efficiency without any additional information but the statistical features extracted from network traffic (e.g., URL, URL query names, hostname). This property makes DART packet content-agnostic, as a result, it is highly desirable to be deployed in a Security Operation Center (SOC) since it works without undermining users' privacy.

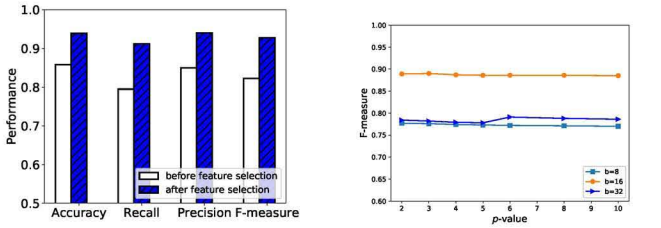
TABLE III
FEATURES SELECTED FROM THE NUSW-NB15 DATASET

Name	Feature Description	Relevance
sintpkt	Source interpacket arrival time (mSec)	24.590%
synack	TCP connection setup time, the time between the SYN and the SYN_ACK packets	23.810%
ackdat	TCP connection setup time, the time between the SYN_ACK and the ACK packets	23.306%
tcprtt	The sum of 'synack' and 'ackdat' of the TCP	19.957%
sload	Source bits per second	19.484%
dpkts	Destination to source packet count	19.085%
dur	Record total duration	18.216%
dload	Destination bits per second	18.041%
dmeansz	Mean of the flow packet size transmitted by the dst	16.716%
dintpkt	Destination interpacket arrival time (mSec)	16.064%
smeansz	Mean of the flow packet size transmitted by the src	15.310%
sbytes	Source to destination transaction bytes	15.018%
dbytes	Destination to source transaction bytes	14.449%
sttl	Source to destination time to live value	13.920%
dttl	Destination to source time to live value	12.678%
ct_state_ttl	No. for each state according to specific range of values for source/destination time to live	12.359%

C. Feature Selection & Impact of the Parameters

In contrast to the artificial feature selection approach, the feature selection algorithm based on mutual information can better deal with the characteristics of nonlinear dependent data. Figure 3(a) displays the overall detection performance of DART before and after the feature selection. The experimental result shows that a feature selection algorithm based on mutual information has positive impacts to the DART framework detection performance.

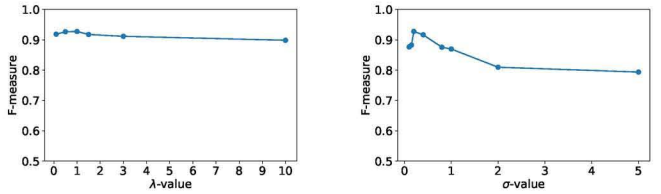
Furthermore, DART framework involves several important parameters that need to be optimized, specifically, number of bins and nearest neighbors, shrinkage regularization σ and MMD regularization λ. In Figure 3(b), we use the F-measure as the evaluation metric and report the comparative experimental results to find out the optimal configurations of bin number b and nearest neighbor number ρ. From Figure 3(b), we can see that when bin number b=16, the classifier performs better than other considered values (i.e., b=8 and b=32). For b=16, we can notice that varying the nearest neighbor parameter p does not affect the performance. For the sake of lowering computation complexity, the number of nearest neighbors is set to p=3.



(a) Impact of feature selection to the detection performance.

(b) Impact of b and ρ values to the detection performance.

Fig. 3. Impact of feature selection and b and ρ parameters.



(a) Impact of λ values.

(b) Impact of σ values to the detection performance.

Fig. 4. Impact of the λ and σ parameter to the detection performance.

Figures 4(a) and 4(b) show the evaluation of the performance (in terms of F-measure) for shrinkage regularization parameter σ and MMD regularization λ parameter, respectively. We can notice that the detection performance of DART can be affected by the value given to these parameters. From the experimental results, we can assess that the optimal values of parameter σ and λ are 0.2 and 1, respectively. In summary, we choose b=16, ρ=3, σ=0.2 and λ=1 to be the optimal parameters for the DART framework.

D. Detection Performance

This section shows the benefit of the DART for two-class classification problem compared with the traditional approaches or algorithms and other state-of-the-art methods for malware variants detection.

1) Comparison with traditional approaches: We first compare the performance of our method against traditional machine learning approaches for detecting malware variants. In particular, we consider support vector machine (SVM), logistic regression (LR) and semi-supervised machine learning (SSL) methods.

Figure 5 illustrates the comparison between DART detection framework and the other machine learning approaches in terms of F-measure. We can see that the proposed model outperforms the model trained by traditional machine learning algorithms in almost all of the categories except Generic family. If no samples are available for training, it is difficult for traditional approaches to detect unseen malicious family, as no discriminative patterns can be discovered by machine learning algorithms. Hence, if malware variants evade detection by changing their network behaviors, the traditional machine learning algorithms will fail to detect the malware variants effectively due to distribution disparity between training data and

TABLE IV
DETECTION PERFORMANCE OF DART AND COMPARISON WITH GAA-ADS [12] ON THE NUSW-NB15 DATASET

	UNSW-NB15 Dataset		
Methods	Recall	Accuracy	FPR
GAA-ADS(K=2)	0.754	0.776	0.082
GAA-ADS(K=4)	0.852	0.86	0.063
GAA-ADS(K=6)	0.871	0.882	0.061
GAA-ADS(K=8)	0.912	0.927	0.059
GAA-ADS(K=10)	0.913	0.928	0.051
DART	**0.912**	**0.939**	**0.028**

testing data. DART takes full account of the different marginal and conditional distributions between source malware domain and target unseen malware variant domain, which significantly improves the detection performance in terms of F-measure metrics.

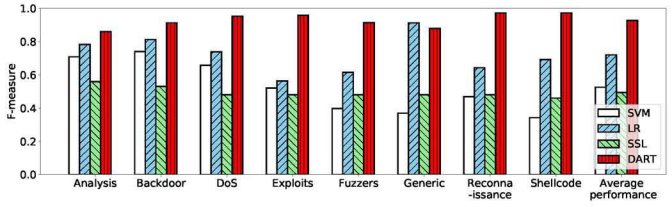


Fig. 5. F-measure rate of DART and comparison with other traditional machine learning approaches.

2) Comparison with other state-of-the-art methods: Although DART framework shows better performance compared to traditional machine learning approaches, we also compare our proposal against another intrusion detection system technique, GAA-ADS method [12], which is a state-of-the-art intrusion detection system. The performance of the GAA-ADS approach and DART framework in terms of the overall recall value, accuracy and false positive rate (FPR), are presented in Table IV, which shows that our method achieves a much higher accuracy than GAA-ADS with a much lower FPR with the same dataset, i.e., UNSW-N15. DART can achieve over 93.9% accuracy and 91.2% recall value with less than 2.8% FPR value.

VI. CONCLUSION

This paper proposed DART, a novel framework that addresses the challenge of detecting malware variants or unseen malware. The core of DART framework is an Adaptive Regularization Transfer Learning module that allows projecting the knowledge from a learning domain to another. After data preprocessing of network trace and feature selection, we first group a set of network traffic into bags and then construct robust representation by using self-similarity matrix and histogram representation approaches. Lastly, we introduce the adaptation regularization transfer learning to address cross-domain learning problem in the form of unseen malicious traffic classification. Extensive experiments on NUSW-NB15 datasets validate that the proposed approach can achieve over 90% of F-measure and 93% of recall in detecting unseen malware, significantly outperforming the conventional machine learning-based intrusion detection systems.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grants No. 61672262, No. 61472164, No.61702218 and No. 61572230, the Shandong Provincial Key R&D Program under Grant No. 2016GGX101001 and No.2018CXGC0706. This work is also supported in part by NSF grant CNS-1566388.

REFERENCES

[1] Luca Invernizzi, Stanislav Miskovic, Ruben Torres, Sabyasachi Saha, Sung Ju Lee, Marco Mellia, Christopher Kruegel, and Giovanni Vigna. Nazca: Detecting malware distribution in large-scale networks. In *Network and Distributed System Security Symposium*, 2014.

[2] Shanshan Wang, Qiben Yan, Zhenxiang Chen, Bo Yang, Chuan Zhao, and Mauro Conti. Detecting android malware leveraging text semantics of network flows. *IEEE Transactions on Information Forensics & Security*, PP(99):1–1, 2017.

[3] Shanshan Wang, Qiben Yan, Zhenxiang Chen, Lin Wang, Riccardo Spolaor, Bo Yang, and Mauro Conti. Lexical mining of malicious urls for classifying android malware. In *International Conference on Security and Privacy in Communication Systems*, pages 248–263. Springer, 2018.

[4] Mingshen Sun, Xiaolei Li, John C. S. Lui, Richard T. B. Ma, and Zhenkai Liang. Monet: A user-oriented behavior-based malware variants detection system for android. *IEEE Transactions on Information Forensics & Security*, PP(99):1–1, 2016.

[5] Xiao Liang, Yanda Li, Xueli Huang, and Xiaojiang Du. Cloud-based malware detection game for mobile devices with offloading. *IEEE Transactions on Mobile Computing*, 16(10):2742–2750, 2017.

[6] Mohammed A. Ambusaidi, Xiangjian He, Priyadarsi Nanda, and Zhiyuan Tan. Building an intrusion detection system using a filter-based feature selection algorithm. *IEEE Transactions on Computers*, 65(10):2986–2998, 2016.

[7] Jun Zhang, Yang Xiang, Yu Wang, Wanlei Zhou, Yong Xiang, and Yong Guan. Network traffic classification using correlation information. *IEEE Transactions on Parallel & Distributed Systems*, 24(1):104–117, 2013.

[8] Michal Piskozub, Riccardo Spolaor, and Ivan Martinovic. Malalert: Detecting malware in large-scale network traffic using statistical features. *ACM SIGMETRICS Perform. Eval. Rev.*, 2019.

[9] Bushra A AlAhmadi and Ivan Martinovic. Malclassifier: Malware family classification using network flow sequence behaviour. In *Proc. of IEEE eCrime*, 2018.

[10] Mingsheng Long, Jianmin Wang, Guiguang Ding, Sinno Jialin Pan, and Philip S. Yu. Adaptation regularization: A general framework for transfer learning. *IEEE Transactions on Knowledge & Data Engineering*, 26(5):1076–1089, 2014.

[11] Karel Bartos, Michal Sofka, and Vojtech Franc. Optimized invariant representation of network traffic for detecting unseen malware variants. In *Usenix Security Symposium*, 2016.

[12] Nour Moustafa, Jill Slay, and Gideon Creech. Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks. *IEEE Transactions on Big Data*, PP(99):1–1, 2017.

[13] Ikkyun Kim, Daewon Kim, Byunggoo Kim, Yangseo Choi, Seongyong Yoon, Jintae Oh, and Jongsoo Jang. A case study of unknown attack detection against zero-day worm in the honeynet environment. 03:1715–1720, 2009.

[14] Prakash Mandayam Comar, Lei Liu, Sabyasachi Saha, Pang-Ning Tan, and Antonio Nucci. Combining supervised and unsupervised learning for zero-day malware detection. In *Proceedings of the IEEE INFOCOM 2013*, pages 2022–2030, 2013.

[15] Symantec. Internet security threat report. https://www.symantec.com/content/dam/symantec/docs/reports/istr-23-2018-en.pdf, 2017.

[16] Nour Moustafa and Jill Slay. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In *Military Communications and Information Systems Conference*, pages 1–6, 2015.