Federated Anomaly Detection

Chunjiong Zhang
Dept. of AI Convergence Network
Ajou University
Suwon 16499, Korea.
cjz@ajou.ac.kr

Byeong-hee Roh*
Dept. of AI Convergence Network
Ajou University
Suwon 16499, Korea.
bhroh@ajou.ac.kr

Gaoyang Shan
Dept. of Software and Computer Engineering
Ajou University
Suwon 16499, Korea.
shanyang166@ajou.ac.kr

***Abstract*—More and more portable intelligent devices are connected to the Internet in recent years. A way to effectively use the isolated cyber data without involving privacy and realize the cyber intrusion anomaly detection on the portable intelligent devices with relatively limited hardware storage resources and computing power is worth exploring. In this paper, we propose a framework of federated anomaly detection, which enables the device effectively detect the anomaly by sharing the parameters of the federated model in a fully distributed fashion. We formulate the model training problem as a distributed robust optimization problem and subsequently devise an efficient algorithm for it. Experimental studies have also been carried out to reveal the superior performance of the proposed framework and underscore the significant benefits of federated anomaly detection.**

***Index Terms*—federated learning, intelligent devices, anomaly detection, robust optimization problem**

I. Introduction

A wide variety of portable intelligent devices facilitate the interconnection of people, enabling them to quickly share and access information from the internet. At present, most solutions have made significant progress in using machine learning for anomaly detection based on feature technology [3]. It is, however, difficult to learn all features of abnormal attacks using the machine learning model due to the fact that there is diversity of network anomaly data, and that the anomaly data is usually sparse and statistically Non-iid. In addition, portable intellgent devices do not easily implement machine learning models on devices because of their limited hardware storage resources and computing power. The data isolated and dispersed in different domains cannot be effectively utilized, especially when the network data contains overlapping but different privacy information, e.g, the geographic location, identity, etc [2].

In this paper, we propose a federated anomaly detection framework (FAD) that can effectively handle the anomalous behavior of network data in most circumstances such as the internet of things and cloud services. The federated learning (FL)-based architecture ensures information security, and protection of terminal data and personal privacy when big network data is exchanged. We conduct efficient machine learning between multiple computing nodes for sparse and isolated network data. Through federated model parameter sharing, it can avoid memory overflow problems caused by a large number of data, thereby effectively performing abnormality detection on the portable intelligent device. We also describe the model training problem as a distributed robust optimization problem that make full use of the particularity of different types of anomaly on the entire federated.

This work is supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-2018-0-01431) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

II. The Proposed Framework

A. Overview of the Framework

The goal of FAD is to achieve accurate network anomaly detection on portable intelligent devices through FL without compromising user privacy. Under FL framework, our approach involves multiple portable intelligent devices (users) and one server. The framework consists mainly of three procedures. Firstly, the cloud model on the server side is the core component of the entire federated anomaly detection. It takes the average of the model parameters uploaded by users, and calculates the duality factor of each type of attack through distributed robust optimization. The cloud model and the duality factor are then distributed to users. For the purpose of model training, a user can adjust the number of network anomaly data through the dual factor. Consequently, the user uploads the model parameters of current training and the loss function of each type of network anomaly attack to the server.

B. Framework Implementation

We assume that there is a total of K type of anomaly attack. The dataset for the k^{th} domain associated with the m^{th} users is denoted as $\mathbf{x}^k(m) = \{x_k^i(m), y_k^i(m)\}$, where $x_k^i(m)$ is the i^{th} sample from the k^{th} domain , and $y_k^i(m)$ is the associated label. We further assume there are a total of M users and the model parameters associated with the m^{th} device is $\underline{\mathrm{w}}_m$. Let $f_k(\underline{\mathrm{w}}_m; \mathbf{x}^k(m))$ denote the loss function of the k^{th} type of anomaly attack for the m^{th} users. The cluster structure is assumed in this work, i.e., users within the same cluster share similar model parameters. Let C be the total number of clusters. ϑ_c indicates the set containing the indices of all users associated with the c^{th} cluster. The model parameters associated with the c^{th} cluster is $\tilde{\underline{\mathrm{w}}}_c$. Let $\tilde{\underline{\mathrm{w}}}$ represent the collection of model parameters from all clusters, i.e., $\tilde{\underline{\mathrm{w}}} = [\tilde{\underline{\mathrm{w}}}_1, ..., \tilde{\underline{\mathrm{w}}}_C]$. We aim to minimize the following empirical loss function.

2833-292X/24/$31.00 ©2024 IEEE
DOI 10.1109/DSN-S60304.2024.00041

$$\begin{aligned} \min \quad & \left\{ \sum_{m=1}^{M} \gamma_m + \lambda_1 \sum_{c=1}^{C} \sum_{m \in \vartheta_c} \|\tilde{\underline{\mathrm{w}}}_c - \underline{\mathrm{w}}_m\| \right\} \\ \text{s.t.} \quad & \gamma_m \geq f_k(\underline{\mathrm{w}}_m), \forall k, \\ \text{var} \quad & \underline{\mathrm{w}}, \gamma_m, \tilde{\underline{\mathrm{w}}}. \end{aligned} \tag{1}$$

The first component in the objective function in (1) is the maximum loss of the machine learning model across all type of anomaly attack for the m^{th} users. The second component depicts the similarities among machine learning models across different users. Assuming the optimal solution to the above problem is $\underline{\mathrm{w}}^*, \gamma_m^*$. We have

$$\gamma_m^* = \max_k f_k(\underline{\mathrm{w}}_m^*) = \max_{[p_1,...,p_k]:\sum_k p_k=1} p_k f_k(\underline{\mathrm{w}}_m^*). \tag{2}$$

Let $\Phi = \{[p_1, p_2, .., p_K] : \sum_k p_k = 1\}$. The optimization problem given in (1) can be reformulated as follows.

$$\min_{\underline{\mathrm{w}}, \tilde{\underline{\mathrm{w}}}} \left\{ \sum_{m=1}^{M} \underline{\mathrm{p}}_m \max_{\underline{\mathrm{p}}_m \in \Phi} \underline{\mathrm{p}}_m^T \underline{\mathrm{f}}(\underline{\mathrm{w}}_m) + \lambda_1 \sum_{c=1}^{C} \sum_{m \in \vartheta_c} \|\tilde{\underline{\mathrm{w}}}_c - \underline{\mathrm{w}}_m\| \right\} \tag{3}$$

where $\underline{\mathrm{p}}_m$ can be deemed to be the adversarial distribution of different type abnormal [1]. Φ denotes the feasible set of $\underline{\mathrm{p}}_m$. $\underline{\mathrm{f}}(\underline{\mathrm{w}}_m)$ denote a vector of loss functions, given by $\underline{\mathrm{f}}(\underline{\mathrm{w}}_m) = [f_1(\underline{\mathrm{w}}_m), f_2(\underline{\mathrm{w}}_m), ..., f_k(\underline{\mathrm{w}}_m)]^T$. λ_1 are weighting factors to tradeoff these two components in the objective function. Let c represent the cluster index of the m^{th} users. Let $\beta_1 = \lambda_1 - 1$ and $\underline{\mathrm{g}}_m(\underline{\mathrm{w}}_m) = \underline{\mathrm{f}}_m(\underline{\mathrm{w}}_m) + \beta_1 \|\tilde{\underline{\mathrm{w}}}_c - \underline{\mathrm{w}}_m\|$. Since $||P||_1 = 1$, the optimization problem given in (3) can be compactly represented as follows.

$$\min_{\underline{\mathrm{w}}, \tilde{\underline{\mathrm{w}}}} \left\{ \sum_{m=1}^{M} \underline{\mathrm{p}}_m \max_{\underline{\mathrm{p}}_m \in \Phi} \underline{\mathrm{p}}_m^T \underline{\mathrm{g}}_m(\underline{\mathrm{w}}_m) + \sum_{c=1}^{C} \sum_{m \in \vartheta_c} \|\tilde{\underline{\mathrm{w}}}_c - \underline{\mathrm{w}}_m\| \right\}. \tag{4}$$

The above observation motivates us to devise an efficient optimization algorithm to solve (4) in a distributed manner. As describe overview of the framework that the process will be iterated until convergence. Let η_p and η_w represent the learning rates. More details of this algorithm are given in Algorithm II-B.

III. Experiments

In the experiment, the KDDcup99 dataset is employed. The data contains four types of anomaly attacks, namely Dos, R2L, U2R, and Probe. Implementation of FAD is an effective way to solve abnormal problem. Autoencoder has been employed as the base model in this experiment. 2 different machine learning are considered baseline, Isolation Forest (IF) to identify single-class anomaly attacks. Multiple types of abnormal attacks (MCSU) is considered in the second approach in which we build machine learning models based on the datasets from multi-category abnormal attacks for a single device without federated. Table I summarizes the F-score of these three approaches. It is farely clear that the proposed FAD can achieve robust performance in the anomaly detection of different types of attacks, with more expected performance than other methods.

Algorithm 1 FAD Framework

1: Initialize $\left\{\underline{\mathrm{w}}_{m,0}\right\}$
2: **for iterations** $s = 0, 1, ...S$ **do**
3: Initialize $\underline{\mathrm{p}}_{m,1} = [\frac{1}{K}, ..., \frac{1}{K}]$
4: **for all users** $m \in \{1, 2, ..., M\}$ **do in parallel**
5: **for iterations** $t = 1, ..., T$ **do**
6: Compute the average loss function $\tilde{g}_{mk}(\underline{\mathrm{w}}_{m,t})$ for the k^{th} type abnormal on the m^{th} device.
7: $\underline{\mathrm{w}}_{m,t+1} = \underline{\mathrm{w}}_{m,t} - \eta_w \sum_k p_{mk} \nabla \tilde{g}_{mk}(\underline{\mathrm{w}}_{m,t})$
8: **end for**
9: **end for**
10: return all $\tilde{g}_{mk}(\underline{\mathrm{w}}_{m,T})$ and $\underline{\mathrm{w}}_{m,T}$ from m^{th} node to server
11: servers compute $p_{m,t+1}^k = \frac{p_{m,t}^k exp\left(\eta_p \tilde{g}_{mk}^t(\underline{\mathrm{w}}_{m,t})\right)}{\sum_k p_{m,t}^k exp\left(\eta_p \tilde{g}_{mk}^t \underline{\mathrm{w}}_{m,t}\right))}$, $\tilde{\underline{\mathrm{w}}}_c = \frac{1}{|\vartheta_c|} \sum_{m \in \vartheta_c} \underline{\mathrm{w}}_{m,T}, \forall c$ and $\underline{\mathrm{w}}_{m,0} = \tilde{\underline{\mathrm{w}}}_c, \forall m \in \vartheta_c$ and send back to nodes
12: **end for**
13: return: $\left\{\underline{\mathrm{w}}_{m,0}\right\}$

TABLE I
KDDcup99 F-score performance

	Dos	*R2L*	*U2R*	*Probe*
IF	0.8669	0.4969	0.4587	0.9031
MCSU	0.9207	0.6763	0.6916	0.9014
FAD	**0.9620**	**0.9320**	**0.9008**	**0.9332**

IV. Conclusion

In this paper, we have proposed a framework of federated anomaly detection for drawing inference from dispersed datasets. The proposed framework harnesses the power of both cyber anomaly detection and federated learning while fully preserving the data privacy. An efficient optimization algorithm has also been developed to solve the obtained robust optimization problem in a fully distributed fashion.Experiments have been conducted to reveal the superior performance of the proposed FAD approach.

References

[1] W. Liang, Y. Li, J. Xu, Z. Qin, D. Zhang, and K.-C. Li, "Qos prediction and adversarial attack protection for distributed services under dlaas," *IEEE Transactions on Computers*, 2023.

[2] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, "Federated-learning-based anomaly detection for iot security attacks," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2545–2554, 2021.

[3] Q. Qian, S. Zhu, J. Tang, R. Jin, B. Sun, and H. Li, "Robust optimization over multiple domains," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 4739–4746.