Malware Detection using Federated Learning based on HPC Events and Localized Image features

Sanket Shukla*, Gaurav Kolhe†, Houman Homayoun†,Sai Manoj P D†, Setareh Rafatirad*
*Department of Information Sciences and Technology
†Department of Electrical and Computer Engineering
George Mason University, Fairfax VA, USA 22030
Email: {sshukla4, gkolhe, hhomayou, spudukot, srafatir}@gmu.edu

***Abstract*—Malware is a global threat and it has seen a tremendous increase as well as diversity which made threat detection and analysis a pivotal challenge to address. The increasing diversity in the malware syntax and behavior is some of the basic challenges to address for efficient malware detection. Thus, an efficient detection requires knowledge of different threats across the globe. However, it is impractical to have a signature-based detection or maintain a database with all malware signatures or syntax. To address these challenges, we propose a federated learning(FL)-based framework that aids to learn the threat features and characteristics irrespective of its origin and without breaching users' data or privacy for an enhanced and robust security of billions of devices across the globe against malware. The federated learning (FL) model obtains the models from a selected set of devices to determine the model parameters required for efficient detection of heterogeneous malware types. Further, one model that encompasses of knowledge from different models obtained from different devices is emerged, which will be further broadcasted to the individual device for efficient malware detection, despite a given device has previously encountered or trained with characteristics of the malware. For the individual devices, we deploy a two-pronged malware detection technique. In first prong, we extract the microarchitectural traces obtained while executing the application to detect traditional malware and in second prong, we introduce an automated localized feature extraction technique to detect obfuscated malware. With the proposed FL framework, we achieved 91% malware detection accuracy, irrespective of training data used at device-level. Furthermore, the proposed framework achieves up to 11% higher detection accuracy compared to the existing malware detection techniques.**

I. Introduction

In the era of Internet-of-Things (IoT), millions of IoT and computing devices are connected irrespective of their spatial location. Despite advanced computing features, these devices also pose significant threats. Among multiple threats on these devices, malware is one of the pivotal threats. Malicious Software, generally known as 'malware' is a software program or application developed by an attacker to gain unintended access to the computing device in order to perform unauthorized accesses as well as malicious activities such as stealing data and sensitive information such as credentials. To alleviate the threats and meet the IoT device constraints, it is non-trivial to devise an efficient malware detection.

One major concern in detecting malware is the heterogenity in the malware behavior with respect to its syntax, style, and origin as the malware developers are spread across the globe. This makes it impractical to maintain a global database for all malware signatures and formats. The existing malware detection techniques are inefficient to global database as they are trained on limited datasets [1]. Thus, we need a framework which outperforms existing traditional and signature based malware detection techniques [2] [3].

To mitigate the dependence of node-level training on the malware detection performance and enable malware detection irrespective of its origin, we propose a federated learning (FL)-based framework in this work. The FL module obtains the malware detection model information and the features they are using to detect threats and forms a newer resilient and efficient detection model through federated averaging method [4]. Further, the updated model is pushed to all the connected devices, whose models will be updated, making them capable of detecting unseen malware (malware which is not seen during their initial training). This process is repeated iteratively by choosing different devices depending on certain characteristics such as power consumption, similar to [5]. Such framework makes the detection efficient, spatiality-agnostic as well as preserve the user data and privacy.

II. Proposed Architecture

The proposed FL-based malware detection is shown in Figure 1. Federated learning (FL) is a collaborative learning approach, where a collaborative model is formed based on the models obtained from the peers. To overcome the computational complexity induced due to backpropagation and other computations, federated averaging algorithm [4] is utilized. Further, as the communication to and from devices could be expensive, a light communication protocol is devised that performs federated optimization [5]. One major advantage of FL stems from the fact that with the model provided by the FL module, an individual device can detect the threats that were not used in its detection training dataset and is location agnostic. The learning algorithm for this setup takes place in every round, where each device independently computes an update to the current model based on its local data, and communicates this update to a central server. On this central server, devices' updates are averaged to compute a new global model. Functionally, a device that is a part of FL computing architecture, downloads a model that is meant for running on devices. It then executes the model locally on the node and improves it by learning from local data stored there. Subsequently, it summarizes the changes as a small update, typically containing the model parameters and corresponding weights for training. The update to the model is then sent to

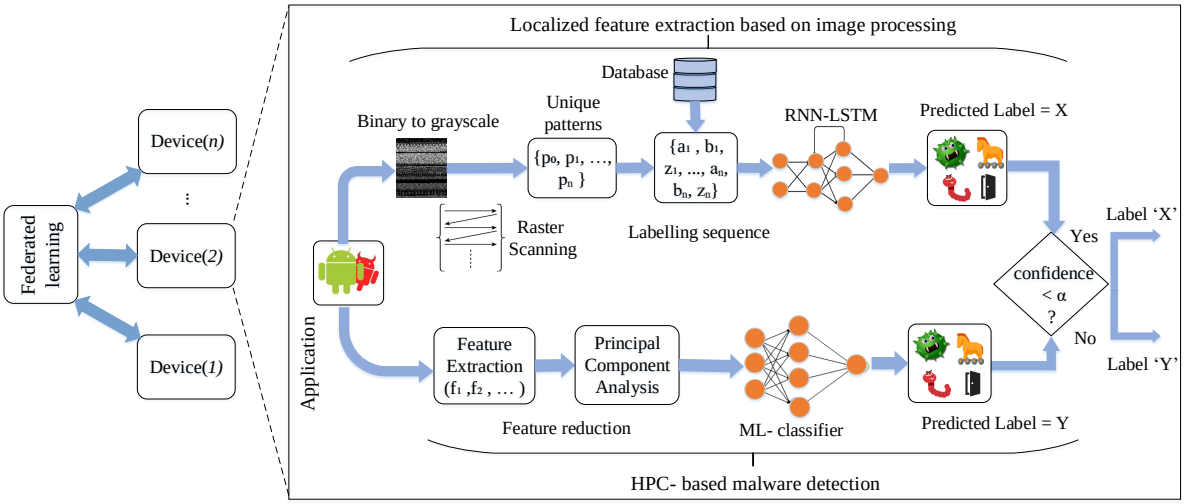

Fig. 1. Proposed malware detection framework using federated learning

the cloud or central server using encrypted communication, for example, homomorphic encryption (HE). This update is then averaged with other user updates to improve the shared model. Most importantly, all the training data remains on user's device, and no individual updates are identifiably stored in the cloud. Our FL global model consists of number of local models trained for different classes of malware.

In the proposed framework, we assume that each node-level malware detector consists of two pronges: HPC-based and localized feature extraction-based analysis, as in Figure 1. In the HPC-based technique, the microarchitectural event traces are captured through hardware performance counters (HPCs) for analysis. To determine the most prominent microarchitectural events, principal component analysis (PCA) is utilized. Further, the chosen prominent features/events are captured and fed to machine learning classifiers for threat detection [6]. Despite being lightweight, this approach is inefficient for detection malware crafted through some of the advanced obfuscation techniques. To address such concerns, a localized feature extraction based approach is deployed. Here, the application binary is converted into a gray-scale image for localized feature extraction. Further, the unique patterns in the image are obtained through techniques such as cosine similarity. Once the image patterns are recognized for a given binary file, the whole image binary is converted into a sequence of patterns. These sequence of patterns is fed to a long short-term memory (LSTM) recurrent neural network (RNN) to classify and detect the incoming stealthy malware binary by predicting the corresponding class label. Depending on the confidence of the prediction from both the techniques, the label is chosen at the node-level. The built model at node-level will be communicated to the FL module and will be further updated based on the input from the FL module. Thus, the proposed FL-based approach can aid in detecting threats that are never seen nor trained at device level and is widely apt for increasing malware threats across the globe.

III. Experimental Result

The proposed FL framework is implemented on an Intel core i7-8750H CPU with 16GB RAM with nearly 4 different individual computing nodes connected as in Figure 1 with overall 13000 malware and stealthy malware samples used for training and inference. It should to be noted that different nodes are trained with different classes of malware. Figure 2 illustrates performance comparison of different models used for malware detection task. It can be observed that FL model outperforms the malware detection task with average detection accuracy of 91%, while the average malware detection accuracy significantly drops when using other ML models.

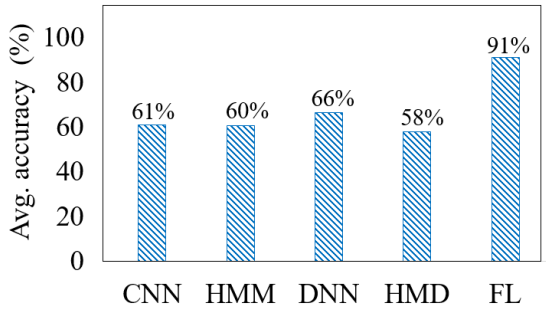

Fig. 2. Performance comparison of models for malware detection

IV. Conclusions

In this work, we propose a training and location agnostic FL framework for malware and stealthy malware detection. The proposed approach can efficiently detect the malware with an average accuracy of 91%. Our future research will be exploring the FPGA implementation, hardware analysis and GPU based processing for the implemented framework.

References

[1] L. Nataraj and et al., "Malware images: Visualization and automatic classification," in *Int. Symposium on Visualization for Cyber Security*.

[2] Q. Chen and et al., "Automated behavioral analysis of malware: A case study of wannacry ransomware," in *IEEE Int. Conf. on Machine Learning and Applications*.

[3] J. Su and et al., "Lightweight classification of iot malware based on image recognition," in *IEEE Annual Computer Software and Applications Conf.*

[4] J. Konečný and et al., "Federated optimization: Distributed machine learning for on-device intelligence."

[5] S. Caldas and et al., "Expanding the reach of federated learning by reducing client resource requirements."

[6] H. Sayadi and et al., "Ensemble learning for effective run-time hardware-based malware detection: A comprehensive analysis and classification," *Design Automation Conf.(DAC)*, 2018.