ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 06, June 2025

Review Paper: A Detailed Review of Federated Learning in Cybersecurity with a Focus on Sandbox Integration

[1] Anushka Kahate, [2] Ruchali Babulkar, [3] Ruchita Chakole, [4] Sharayu Deote

[1] [2] [3] [4] Cummins College of Engineering for women Nagpur, Maharashtra, India
Corresponding Author Email: [1] anushkakahate29@gmail.com

Abstract— ***The threat of cyber-attacks, especially malware, is rapidly evolving and requires complex solutions that protect individual information from unauthorized access while providing high protection against malicious software. Federated Learning (FL) is a novel form of machine learning that enables model updates to be transferred among various clients without considering original data to be sent to a central hub. Several studies have investigated FL in cybersecurity; however, previous models present challenges associated with poisoning attacks, data heterogeneity, and no integration of sandbox for malware analysis. Based on this review thus critically discussed the current limitations on FL research in cybersecurity and possible solutions. Finally, they discuss the idea of combining Docker-based sandboxes with FL to solve these challenges, and they advocate for a feature-robust, privacy-preserving malware detection framework.***

Index Terms— ***Federated Learning, Malware Detection, Security Sandbox, Model Poisoning, Data Privacy, Docker, Threat Detection, Cybersecurity.***

I. INTRODUCTION

This is a constantly growing area of challenge given the advancements of cyber-attacks especially the malware type. ES These traditional approaches of detection like signature-based detection system and centralized data processing fails to cope up with such threats at the same time maintaining privacy. Centralized feature calls for sharing of raw data which is likely to result in privacy violation especially within banking, health, and security sectors.

A. Federated Learning (FL)

FL has come up as one of the probable solutions that would enable decentralized model training. One form of the distributed models involves each client training a local model on its data then transferring the model gradients or weights to a central unit. The updates are collected by the central server to enhance a global model all while the actual data is never visible. This process ensures privacy but introduces new challenges, such as:

- **Model Poisoning**: However, the local updates performed by clients can be malicious in the sense they will provide compromised information that will affect the global model.
- **Data Heterogeneity**: Data cross over between clients might not be Independent and Identically Distributed (Non-IID), thus affecting performance.
- **Lack of Sandbox Integration**: Existing models for malaria detection includes those based in FL are not equipped with sandboxes, which are fundamentally required for dynamic analysis of instances of malware.

Hence, this review paper seeks to evaluate the current studies on FL in cybersecurity and offer a solution of incorporating Docker-based sandboxes within the FL framework with the enhancement of the security and efficiency of the system.

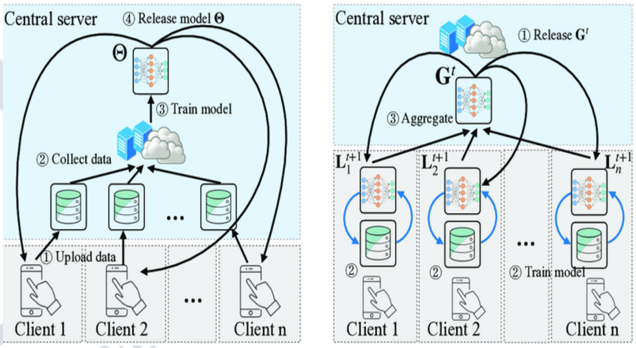


Figure 1. Centralized vs. Federated Learning – with reference to the flow chart of centralized learning as well as federated learning framework. In the federated learning system, (a) clients upload local dataset to a trust central server, (b) while, clients keep their private data locally.

II. REVIEW OF FEDERATED LEARNING IN CYBERSECURITY

In more detail, this paper aims to provide a review of the federated learning in cybersecurity proposed frameworks and approaches.

A. Federated Learning for Cybersecurity:

An Overview FL enables to jointly train models of machine learning for multiple devices (clients) while transmitting raw data to a concerned central server. This kind of decentralized approach is especially helpful where the privacy of an application is concerned, such as in cybersecurity applications. Nevertheless, cybersecurity threats as well as malware detection need dynamic analysis whereas the analyzed FL-based systems do not posse this type of analysis yet.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 06, June 2025

Research Paper Summaries:

- **"Federated Learning for Globally Coordinated Threat Detection"** (arXiv:2205.11459v3): This paper develops FL as a model for threat detection that integrates global coordination. It concentrates on bulk simultaneous detection of threats activity without disclose identity of its clients. It does respond to some extent to model poisoning attacks; it is not efficient enough for handling heterogeneous data at the client side.
- **"Federated Learning: Challenges, Methods, and Future Directions"** (arXiv:1908.07873): This work "Fl Opportunities, Challenges, Methods, and Future Directions," (1908.07873): Specifies the difficulties that affect FL systems namely data privacy, efficiency in communication, and robustness of the model. The authors suggest different approaches to address these problems, nevertheless, the absence of integration with sandbox decreases applicability in malware identification.
- **"Advances and Open Problems in Federated Learning"** (arXiv:1912.04977): The following paper provides an overview of the state-of-the-art developments and the challenges in FL ((DOI: 10.1109/ TCLT52584.2020.0900497). Whereas it gives a good account of the application of FL in different sectors, its major highlighted issues of model poisoning and communication latency imply lacunas pertaining to the protection of FL-based CYBERSECURE systems.

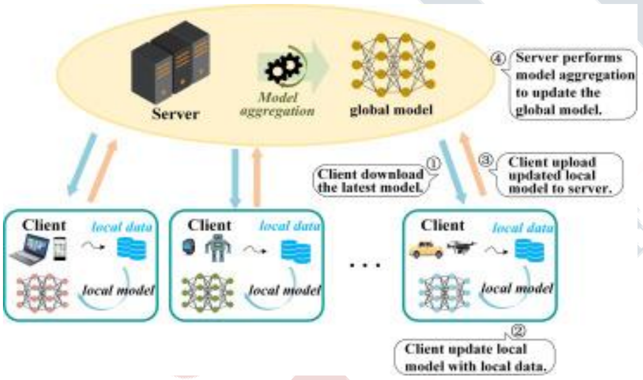


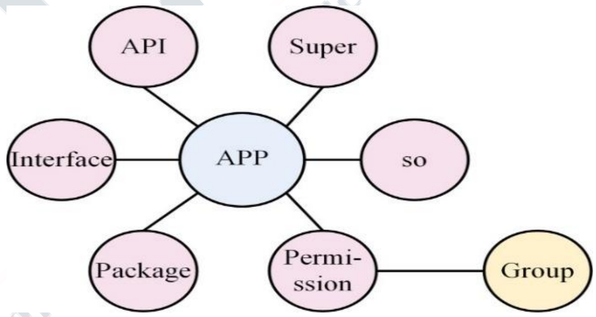
Figure 2. Overview of the general FL framework. During the nth round of communication, each user downloads the new global model from the server to start with ① and uses its own local dataset for iterative training ② to create new global models, the server ③ which performs model collection ④ and then performs training.

B. Privacy-Preserving Malware Detection Using FL

A good anti-malware system has to maintain user anonymity while providing great effectiveness. Some authors have used FL to detect malware particularly in contexts such as mobile devices and the IoT … FL contains private data on the client side and processes raw values but lacks a sandbox for dynamic analysis to detect complex malware programs.

Research Paper Summaries:

- **"Less is More: A Privacy-Respecting Android Malware Classifier Using Federated Learning"** (arXiv:2007.08319): This is another paper that develops an FL-based malware classifier that no raw data is shared between devices. While analysing the benefits of decentralized training, the paper reports the absence of an overall integrated sandbox environment for studying malware activity. Moreover, for data model updates at each client, it fails to give good model performance owing to variations in data distribution across clients.
- **"Distributed Detection of Malicious Android Apps While Preserving Privacy Using Federated Learning"** (https://doi.org/10.3390/s23042198): This paper provides example of educational data mining where an FL-based system for detection of malicious Android apps is depicted. They reveal the advantages of FL in maintaining privacy while classifying malware, more issues raised concerns the challenges of accurate model forecast due to different data owned by different clients.



Graph 1: Heterogeneity of Data Across Clients

C. Challenges in Federated Learning-Based Malware Detection

There are several challenges in applying FL to malware detection, which include:

1. **Model Poisoning Attacks**: Evil clients can input incorrect updates to the model which will affect the global model negatively. That is why securing the aggregation process is crucial to guaranteeing resilience.
2. **Data Heterogeneity**: The data at the clients themselves is non-IID and therefore when training a global model, it may not perform well when applied to individual clients.
3. **Lack of Real-Time Malware Analysis**: However, traditional FL systems fail to include dynamic sandbox environment that is necessary to orchestrate the examination of behaviour of the suspicious files in a real-time manner.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 06, June 2025

4. **Research Paper Summaries:**

- **"Threats, Attacks, and Defences to Federated Learning: Issues, Taxonomy, and Perspectives"** (https://doi.org/10.1186/s42400-021-00105-6): This paper elaborates on the security threats to FL with an emphasis on Model Poisoning and Inference Attacks. It also enunciates types of attacks and structural defences. However, it does not incorporate solutions for dynamic risk assessment in real time using sandboxing editions.
- **"A Federated Learning Multi-Task Scheduling Mechanism Based on Trusted Computing Sandbox"** (https://doi.org/10.3390/s23042093): This work also introduces a multi-task scheduling strategy for the FL, which relies on trusted computing sandboxes. But though it includes sandbox environments, it is still theoretical architecture and doesn't provide a practical solution for handling dynamically appeared malware at clients.

III. PROPOSED CONCEPT

In registering for decentralized learning, participants enable the creation of restricted, private safety domains. Therefore, drawing from the limitations recognized in the prior studies, we contribute a new paradigm which is the blend of Docker-based security sandboxes together with Federated Learning to improve the identification of malware with privacy preservation.

A. Concept Overview

The authors in propose the usage of Docker based sandboxes for creating the isolated environment for malware analysis and the application of Federated Learning for secure and private model training. The Docker sandbox enables files to run in an isolated manner; FL ensures that private data are run only on every client's system. Clients establish their local models based on the data computed on the sandbox and send to the central server only the deltas of those models.

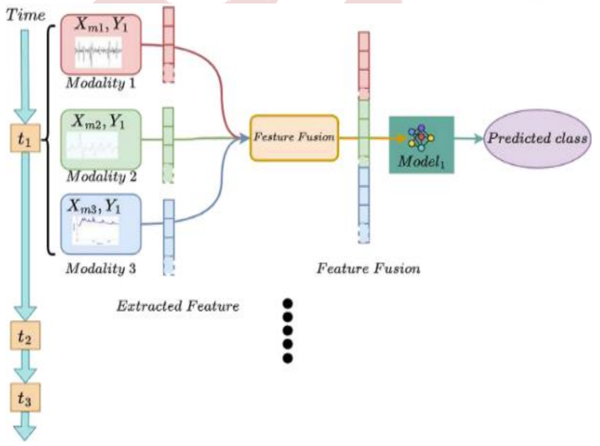


Figure 3. Proposed Framework – Federated Learning with Docker Sandbox Integration.

B. Federated Learning Sandbox Architecture

a. Docker-Based Sandbox

Files which are thought to be malware are placed in Docker containers. Every client has a containerized environment for malware testing within which the samples are run with no threat to system integrity. Such containers may report file system changes, network activity, and system calls which accompany a file, offering beneficial attributes for malware categorization.

- **Isolation and Security**: The isolation skills of Docker make it impossible for malicious files to affect the host system.
- **Portability**: Containers can be adopted to various client settings making it easy to scale up easily.

b. Federated Learning Model

FL model is trained from data obtain in the sandboxed environment on each of the clients. The process includes the following steps:

- **Local Training**: With this kind of architecture, each clients data stored locally is fed into an ML model that tries to detect and learn the behaviour of Malware.
- **Model Updates**: Instead of sending raw data, clients send model weights and gradients computing at the other's end to be averaged.
- **Global Model Aggregation**: It again sends updates from many clients to update the global malware detection model in the server side. It is then returned to clients for enhancing local performance according to proposals made by the author of this writing.

C. Addressing Key Challenges

a. Model Poisoning

In this work, the proposed system also has protective measures against model poisoning attacks. Several secure aggregation methods are used to ensure the authenticity of the model updates to merge into the global models.

b. Data Heterogeneity

Therefore, to overcome the issue of non-IID data across clients, the system uses multi-task learning than clients train models for their environments, for instance, IoT malware and Android malware environments. This approach guarantees that the global model captures different conducting behavior of malware.

c. Real-Time Malware Detection through Sandbox

With Docker based sandboxes, real time malware analysis can be conducted. Files are loaded at run-time into an enclave and executed to generate behavior to train the local models. This method offers a direct operate threat detection approach, which is missing in traditional FL models.

IFERP Explore Your Research Journey...
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 12, Issue 06, June 2025

D. Implementation of the Security Sandbox

a. Docker Setup

The approach involves each client running Docker and setting up a sandbox container containing basic tools for detection of malware. Because Docker is light, it can be effectively deployed on mobiles and IoT devices that are likely to be resource limited.

b. Federated Learning Workflow

Clients perform the following steps:

- **File Testing**: Files are conducted and uploaded inside the Docker sandbox.
- **Local Model Training**: Machine learning model is trained at local environment using the collected logs of the sandbox.
- **Model Sharing**: Users only submit new model parameters to the central server instead of raw client data being used.
- **Global Model Aggregation**: The updates collected by the central server can enhance the Global malware detection model for all the individual servers.

c. Malware Detection and Testing

After the global model gets updated, the clients carry out the classification of files, as either possessing malicious or benign content. The system makes it possible to test during operation, and therefore increases the effectiveness of detecting malware.

IV. CONCLUSION

This review identifies key limitations in existing FL-based malware detection systems and proposes an innovative solution: a union of native presumed-safety Docker-based security sandboxes with federated neuromorphic learning. This approach contradicts the issues of model poisoning, consistency in the analyzed data, besides, dynamic analysis of malware. More work in the future will be target at enhancing the communication protocols, the accuracy of the model and extending the framework to embed other categories of computer cyber threats.

REFERENCES

[1] Hongbin Liu 1, Han Zhou 2, Hao Chen 3, Yong Yan 3, Jianping Huang 3, Ao Xiong 2, Shaojie Yang 2, Jiewei Chen 2, Shaoyong Guo 2, "A Federated Learning Multi-Task Scheduling Mechanism Based on Trusted Computing Sandbox," Journal Sensors, in Volume 23, Issue 4, of 2023, PP. 1-6, DOI: https://doi.org/10.3390/s23042093.

[2] Talha Ongun∗, Simona Boboila, Alina Oprea, Tina Eliassi-Rad, "CELEST: Federated Learning for Globally Coordinated Threat Detection," arXiv preprint in 2022, PP. 1-6, DOI: arXiv:2205.11459v3.

[3] Suchul Lee, "Distributed Detection of Malicious Android Apps While Preserving Privacy Using Federated Learning," journal Sensors in 2023, PP. 1-6, DOI: https://doi.org/10.3390/s23042198.

[4] Rafa Gálvez, Veelasha Moonsamy, and Claudia Diaz, "Less is More: A privacy-respecting Android malware classifier using federated learning," ArXiv preprint from 2020, PP. 1-6, DOI: https://arxiv.org/pdf/2007.08319.

[5] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, Virginia Smith, "Federated Learning: Challenges, Methods, and Future Directions," ArXiv preprint in 2019, PP. 1-6, DOI: https://arxiv.org/abs/1908.07873.

[6] Pengrui Liu, Xiangrui Xu, Wei Wang, "Threats, attacks and defenses to federated learning: issues, taxonomy and perspectives," Journal Cybersecurity in 2021, PP. 1-6, DOI: https://doi.org/10.1186/s42400-021-00105-6.

[7] Thanh Bui, "Analysis of Docker Security," ArXiv preprint in 2015, PP. 1-6, DOI: https://arxiv.org/abs/1501.02967.

[8] Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong, "Federated Machine Learning: Concept and Applications," ArXiv preprint in 2019, PP. 1-6, DOI: https://arxiv.org/abs/1902.04885.

[9] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G.L. D'Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaid Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, Sen Zhao, "Advances and Open Problems in Federated Learning," ArXiv preprint published in 2019, PP. 1-6, DOI: https://arxiv.org/abs/1912.04977.

[10] Mohamed Amine Ferrag, Fatima Alwahedi, Ammar Battah, Bilel Cherif, Abdechakour Mechri, Norbert Tihanyi, "Generative AI and Large Language Models for Cyber Security: All Insights You Need," 2024 ArXiv preprint, PP. 1-6, DOI: https://arxiv.org/abs/2405.12750.

[11] Antonio João Gonçalves de Azambuja, Christian Plesker, Klaus Schützer, Reiner Anderl, Benjamin Schleich, Vilson Rosa Almeida, "Artificial Intelligence-Based Cyber Security in the Context of Industry 4.0—A Survey," Journal Electronics in 2023, PP. 1-6, DOI: https://www.mdpi.com/2079-9292/12/8/1920.

[12] Gartner Research. Market Share: PCs, Ultramobiles and Mobile Phones, All Countries, 4Q21 Update. 2022.

Available online: https://www.gartner.com/en/documents/4011646 (accessed on 12 February 2023).

[13] Kaspersky. IT Threat Evolution in -Q2 2022. Mobile Statistics. 2022. Available online: https://securelist.com/it-threat-evolutionin-q2-2022-mobile-statistics/107123/ (accessed on 12 February 2023).

[14] Vinod, P.; Zemmari, A.; Conti, M. A machine learning based approach to detect malicious android apps using discriminant system calls. Future Gener. Comput. Syst. 2019, 94, 333–350. [CrossRef]

[15] Lee, S.; Kim, S.; Lee, S.; Choi, J.; Yoon, H.; Lee, D.; Lee, J.R. LARGen: Automatic Signature Generation for Malwares Using Latent Dirichlet Allocation. IEEE Trans. Dependable Secur. Comput. 2018, 15, 771–783. [CrossRef]

[16] Drainakis, G.; Katsaros, K.V.; Pantazopoulos, P.; Sourlas, V.; Amditis, A. Federated vs. Centralized Machine Learning under Privacy-elastic Users: A Comparative Analysis. In Proceedings of the 2020 IEEE 19th International Symposium on Network Computing and Applications (NCA), Cambridge, MA, USA, 24–27 November 2020; pp. 1–8. [CrossRef]

[17] Preuveneers, D.; Rimmer, V.; Tsingenopoulos, I.; Spooren, J.; Joosen,W.; Ilie-Zudor, E. Chained Anomaly Detection Models for Federated Learning: An Intrusion Detection Case Study. Appl. Sci. 2018, 8, 2663. [CrossRef]

[18] Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks, 2014. arXiv 2014, arXiv:1406.2661.

[19] Kang, M.; Kim, H.; Lee, S.; Han, S. Resilience against Adversarial Examples: Data-Augmentation Exploiting Generative Adversarial Networks. KSII Trans. Internet Inf. Syst. 2021, 15, 4105–4121. [CrossRef]

[20] Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and open problems in federated learning. Found. Trends® Mach. Learn. 2021, 14, 1–210. [CrossRef]

[21] Smith, V.; Chiang, C.K.; Sanjabi, M.; Talwalkar, A.S. Federated multi-task learning. Adv. Neural Inf. Process. Syst. 2017, 30, 4427–4437.

[22] Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra, V. Federated learning with non-iid data. arXiv 2018, arXiv:1806.00582.

[23] Criado, M.F.; Casado, F.E.; Iglesias, R.; Regueiro, C.V.; Barro, S. Non-IID data and Continual Learning processes in Federated Learning: A long road ahead. Inf. Fusion 2022, 88, 263–280. [CrossRef]

[24] Li, X.; Huang, K.; Yang, W.; Wang, S.; Zhang, Z. On the Convergence of FedAvg on Non-IID Data, 2019. arXiv 2019, arXiv:1907.02189.

[25] Wang, H.; Sievert, S.; Liu, S.; Charles, Z.; Papailiopoulos, D.; Wright, S. Atomo: Communication-efficient learning via atomic sparsification. Adv. Neural Inf. Process. Syst. 2018, 31, 9872–9883.

[26] McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics, PMLR, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.

[27] Wang, Z. Deep learning-based intrusion detection with adversaries. IEEE Access 2018, 6, 38367–38384. [CrossRef]

[28] Huang, C.H.; Lee, T.H.; Chang, L.h.; Lin, J.R.; Horng, G. Adversarial attacks on SDN-based deep learning IDS system. In Proceedings of the International Conference on Mobile andWireless Technology, Hongkong, China, 25–27 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 181–191.

[29] Schultz, M.G.; Eskin, E.; Zadok, F.; Stolfo, S.J. Data mining methods for detection of new malicious executables. In Proceedings of the 2001 IEEE Symposium on Security and Privacy. S&P 2001, Oakland, CA, USA, 14–16 May 2001; IEEE: Piscataway, NJ, USA, 2000; pp. 38–49.

[30] Kong, D.; Yan, G. Discriminant malware distance learning on structural information for automated malware classification. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, FL, USA, 11–14 August 2013; pp. 1357–1365.

[31] Li, Q.; Li, X. Android malware detection based on static analysis of characteristic tree. In Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Xi'an, China, 17–19 September 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 84–91.

[32] Santos, I.; Brezo, F.; Ugarte-Pedrero, X.; Bringas, P.G. Opcode sequences as representation of executables for data-mining-based unknown malware detection. Inf. Sci. 2013, 231, 64–82. [CrossRef] Sensors 2023, 23, 2198 15 of 15

[33] Ni, S.; Qian, Q.; Zhang, R. Malware identification using visualization images and deep learning. Comput. Secur. 2018, 77, 871–885. [CrossRef]

[34] Nataraj, L.; Karthikeyan, S.; Jacob, G.; Manjunath, B.S. Malware images: Visualization and automatic classification. In Proceedings of the 8th International Symposium on Visualization for Cyber Security, Pittsburgh, PA, USA, 20 July 2011; pp. 1–7.

[35] Han, K.S.; Lim, J.H.; Kang, B.; Im, E.G. Malware analysis using visualized images and entropy graphs. Int. J. Inf. Secur. 2015, 14, 1–14. [CrossRef]

[36] Bayer, U.; Comparetti, P.M.; Hlauschek, C.; Kruegel, C.; Kirda, E. Scalable, behavior-based malware clustering. In Proceedings of the NDSS, San Diego, CA, USA, 11–16 February 2009; Volume 9, pp. 8–11.

[37] Anderson, B.; Quist, D.; Neil, J.; Storlie, C.; Lane, T. Graph-based malware detection using dynamic analysis. J. Comput. Virol. 2011, 7, 247–258. [CrossRef]

[38] Fujino, A.; Murakami, J.; Mori, T. Discovering similar malware samples using API call topics. In Proceedings of the 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 140–147.

[39] Arivazhagan, M.G.; Aggarwal, V.; Singh, A.K.; Choudhary, S. Federated learning with personalization layers. arXiv 2019, arXiv:1912.00818.

[40] Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated optimization in heterogeneous networks. Proc. Mach. Learn. Syst. 2020, 2, 429–450.

[41] Shokri, R.; Shmatikov, V. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1310–1321.

[42] Welekar, A.R., Borkar, P. and Dorle, S.S., 2012. Comparative study of IEEE 802.11, 802.15, 802.16, 802.20 standards for distributed VANET. Int J Electr Electron Eng (IJEEE), 1(3).

[43] Shende, M.R. and Welekar, A., 2016. Advanced Steganography for Hiding Data and Image using Audio-Video. International Journal on Recent and Innovation Trends in Computing and Communication, 4(1), p.112