2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3503-4779-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICONSTEM56934.2023.10142810

A Multi-Criteria Intelligence AID Methodology and IoT Based Data Protection Using Machine Learning

Dr.K.P Manikandan[1], Anusha.P[2], Arockia Jaya.J[3], Sumit Pundir[4], K.Rammohan[5] and Dr.Priyabrata Adhikary[6]

[1]Assistant Professor, Department of Computer Science Engineering, Madanapalle Institute of Technology & Science, Kadiri Road, Angallu, Madanapalle, Andhrapradesh-517325,
[2]Assistant Professor, Department of Electronics and Communication Engineering, R.M.K. Engineering College, Chennai, TamilNadu, India
[3]Associate Professor, Department of Computer Science and Engineering, Idhaya Engineering College for Women, Kallakurichi, TamilNadu
[4]Associate Professor, Department of Computer Science Engineering, Graphic Era Deemed to be University, Dehradun, Uttrakhand, India.
[5]Assistant Professor, Department of Computer Science and Engineering, St.Martin's Engineering College, Secunderabad, Telangana, India
[6]Professor, Department of Mechanical Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

E-mail : manikandankp@mits.ac.in anushamevlsi@gmail.com sr.jayaiecw@gmail.com Sumitpundir1983@gmail.com krammohancse@smec.ac.in priyabrata24@gmail.com

Abstract- — IoT-powered devices have become one of the key companions of humans in recent times. Almost every sector of the modern market is heavily reliant on IoT-powered devices such as smartphones, computers and other smart gadgets. This has resulted in the massive availability of digital data containing the personal information of the users. This has caused a massive issue regarding data privacy on IoT-powered devices in recent times. This has resulted in the application of ML-based technologies for fostering better data protection. The concerned study has focused on the role of the different ML-powered technologies such as federated learning and differential learning models. Thai entre study has also focused on numerical analysis of the issues associated with the mentioned models during data protection. It has also provided survey data regarding different FL-based algorithms such as LM, NN, NM, DT and CM. This study has analyzed the issues and strengths of ML-based data protection on local and global data sets.

Keywords: federated learning, differential learning, machine learning, distributed learning, cryptographic methods.

I.INTRODUCTION

Technology has become one of the key factors in the modern-day world with its presence in almost every sector. People are surrounded by technology at every step and entire daily life operations are becoming digital day by day. This has made a huge presence of digital data of people on internet media. This has become one of the key threats to people in recent times. It has been found that incidents of cyberattacks and cyberbullying are increasing day by day. One of the major examples of data breaching has been seen during the **Facebookdata privacy scandal** way back in 2018 [1]. This has made the issue of data protection and privacy much more serious. IoT has become one of the key technologies in the modern world which has billions of connected devices across the world. Progress of IoT is also happening at a rapid pace and it is expected to have **27 billion** IoT devices by 2025. This massive use and growth of IoT devices have resulted in large-scale user-generated data over the internet. It has been reported that the amount of user-generated data will stand between **4 zettabytes to 140 zettabytes** between2020-2025 [2]. A significant amount of personal data of users is collected through IoT devices which can make the user's privacy very vulnerable. It has resulted in reviewing the privacy policies and strategies associated with digital data systems. Concerning this, the evolution of new technologies like **machine learning** has become very crucial in recent times. Machine learning can be one such technology which can help in fostering the data protection regime in IoT-powered devices and systems. However, ML can also play some different roles to create problems regarding privacy protection in IoT-powered devices. This has made this entire study identify the opportunities of ML for data

protection regimes in IoT devices. However, besides that, the entire study has also highlighted the problems and privacy risks with different ML models and systems. This entire study has attempted to provide a comprehensive analysis and investigation of the efficiency of ML in data protection through different issues related to IoT-powered tools. This entire study has also come up with a detailed methodology for efficient data collection regarding the mentioned issue. This entire study has come up with numerical formulas for analyzing the issues and problems in the ML and FL-based data protection system. It has provided survey data based on different algorithms used in the SGD model of the FL learning system which is one of the latest advancements in traditional ML techniques. This analysis has helped the study to also formulate future steps for better data protection in future.

II. METHODOLOGY

Research methodology is one of the most important segments for a study to attain its proposed outcomes. The selection of the right methodology allows a study to execute better data collection as well as better analysis of the collected data. This entire study has mainly focused on secondary data collection methods. The concerned study has collected both qualitative and quantitative secondary data from authentic sources. It has focused on survey data for identifying the impact of ML in protecting the data privacy of users. The concerned study has collected survey data for identifying the efficiency of ML in providing data safety over mobile networks. This survey data has also been used for analyzing the impact of ML on data protection in cloud computing. This entire study has also collected secondary data through surveys for identifying the issues and threats regarding data privacy in IoT. The concerned study has focused on collecting data regarding **"centralized ML-based privacy protection solutions"** which can be very effective for providing future steps in attaining better results regarding data protection. Data regarding encryption techniques and access control in **centralized encryption** has also been collected in this study. The collection of this type of data has helped the study to perform a better analysis of results regarding the concerned topic. Different techniques and models have been identified during the data collection method.

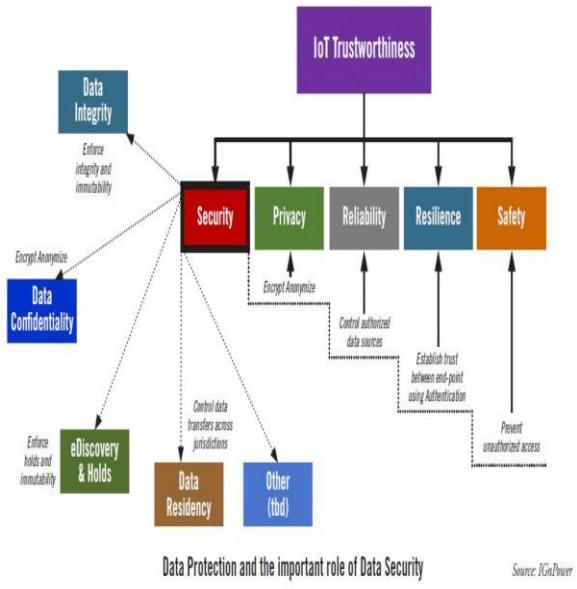

Figure 1: IoT Data Protection (Source: [3])

One such technique is **Differential privacy** which is widely used for data protection for IoT-powered tools. This type of data protection technique adds data perturbation with the original user-generated data for executing data protection [3]. It has also collected secondary survey data regarding **distributed learning.** The concerned study has used a quantitative approach. This approach has helped the concerned study to execute mathematical analysis of the collected secondary data which has resulted in better outcomes for the study.

Numerical analysis of machine learning model for data protection

$$\theta * = \arg\min\left(\sum_{i} L\left(yi, f\theta(xi)\right) + \Omega(\theta)\right),$$

The entire study has also collected numerical data regarding different ML-based data protection models. It can be found from the above formula that $(f\theta)$ is denoting the imputed data which is the **x vector** concerning the output data which is denoted by the **y vector**[4].Y is the set of different data classes and **x^d** is the vector space of **d-dimensional data sets.** This numerical formula has been developed for the **training model of** ML-based data protection. This model can be very efficient for finding the optimal parameter regarding the accurate relationship between the input and output data.

Data fusion, model building and data processing are the three key steps associated with the

centralized learning method. For executing centralized learning methods, this study collected numerical data regarding ***federated learning and gated recruiting units***. The FL-based data protection reads data sets which are stored in ***single or globally predicted*** models. This has been one of the key problems in FL-based data protection.

Therefore, Numerical formula for the globally predicted model problem of the data set

$$\arg\min_{\omega \in R^d} J(\omega) := \sum_{k=1}^{K} \frac{\sum_{i \in D_k} f_i(\omega) + \lambda h(\omega)}{D}.$$

The above formula is describing the ***global model of data aggregation.*** In the above formula, DK is the local storage of data sets in the IoT-powered devices K. The above formula is used for identifying the upgrade in the data sets which can be very helpful for fostering better data protection in future [5]. The main purpose of the mentioned numerical formula in figure 2 is to identify the output data with the loss function. This can be very effective for identifying the issues with ML and FL-powered data protection systems in future. In the above formulas, ***f(ω)*** is denoting the data set dimension in which the input data is denoted concerning output data.

III. Result and Discussion

Machine Learning based cyber security

Machine learning has been one of the key AI-based technologies which help to attain better data protection. ML-based data protection technology helps to get a quick insight into threats regarding digital data protection. The survey data regarding different models of AI-powered data protection is given in the below images.

Publication	Summary	Scope	
		Area	Scenario
Al-Garadi [9] (2020)	Survey of ML and DL methods for IoT Security.	Cybersecurity ML	IoT
Hussain [11] (2020)	Survey of ML and DL based security solutions for IoT networks.	Cybersecurity ML	IoT
Waheed [12] (2020)	Survey of security and privacy using ML and BC in IoT.	Cybersecurity ML	IoT
Khan [13] (2020)	Survey of the technologies used to build the 5G security and privacy model.	Cybersecurity ML	5G
Dixit [14] (2021)	Survey of DL algorithms for cybersecurity applications.	Cybersecurity DL	IoT
Rodriguez [16] (2021)	Survey of DL based cybersecurity solutions for mobile networks.	Cybersecurity DL	Mobile networking
Gosselin [17] (2022)	Privacy and Security in Federated Learning: A Survey.	Cybersecurity FL	IoT
Rigaki [18] (2020)	Survey of privacy attacks against machine learning.	Privacy attacks ML	IoT
Tanuwidjaja [19] (2019)	Survey of privacy-preserving DL techniques.	Privacy-preserving DL	Mobile networking
Boulemtafes [20] (2020)	Survey of privacy-preserving techniques for DL.	Privacy-preserving DL	Mobile networking
Liu [21] (2021)	Survey that reviews interactions between privacy and machine learning.	Privacy ML	Mobile networking
Zheng [22] (2019)	Privacy-preserving ML review for Cloud Computing and IoT.	Privacy-preserving ML	Cloud Computing
Seliem [23] (2018)	Survey of privacy threats in IoT environments.	Privacy-preserving	IoT
Amiri-Zarandi [24] (2020)	Survey of ML-based solutions to protect privacy in the IoT.	Privacy-preserving ML	IoT
Kounoudes [25] (2020)	Survey of user-centric privacy protection approaches in IoT.	User-centric privacy-preserving	IoT
Zhu [26] (2021)	Review privacy-preserving ML training solutions in IoT aggregation scenarios.	Privacy-preserving ML Training	IoT
El Ouadrhiri [27] (2022)	Differential Privacy for Deep and Federated Learning: A Survey.	Differential Privacy FL	IoT

Figure 2: Survey regarding ML-based privacy protection (Source [6])

It can be analyzed from the above survey data that an ML-based data protection system has efficiency over both mobile networks and cloud computing. The above data has helped to analyze that ML-based data protection system provides solutions for three different areas such as ***malware, privacy and network.*** The above survey data has also collected data regarding the ***Federated Learning*** model of data protection. The idea of ***federated learning*** was introduced first by Google in 2017. One of the key features of FL has been that it allows data scientists to link and train **statistical models** of **decentralized devices** with the local data set [7]. This has omitted the necessity of uploading any private data to the cloud storage for training purposes of the data scientists. This has been one of the potential differences between ML-based data privacy and FL-based data privacy. It has been seen that in traditional ML-based data protection systems data sets are required to be uploaded on a single server. This has allowed FL-based data protection systems to provide better privacy protection which is also clear from the survey data. Provide better data protection compared to traditional ML-based data protection.

Table 1: Comparison between different ML-based models for data protection

FL Studies	main area	data partitioning	model implementation	privacy mechanism	communication architecture	remark
FedAvg [129]	Effective Algorithms	horizontal	NN	\	centralized	SGD-based
FedSVRG [94]			LM			
FedProx [108]			LM, NN			
SCAFFOLD [90]			LM, NN			
FedNova [190]			NN			
Per-FedAvg [52]			NN			
pFedMe [46]			LM, NN			
IAPGD, AL2SGD+ [69]			LM			
IFCA [61]			LM, NN			
Agnostic FL [134]			LM, NN			
FedRobust [155]			NN			
FedDF [114]			NN			
FedBCD [120]		vertical	NN			NN-specialized
PNFM [213]		horizontal				
FedMA [189]						
SplitNN [189]		vertical				
[illegible]		[illegible]				
Tree-based FL [217]	Effective Algorithms	horizontal	DT	DP	decentralized	DT-specialized
SimFL [104]				hashing		
FedXGB [122]						
FedForest [121]						
SecureBoost [38]		vertical				
Ridge Regression FL [141]		horizontal	LM	CM		LM-specialized
PPRR [36]						
Linear Regression FL [162]		vertical				
Logistic Regression FL [72]						
Federated MTL [169]					centralized	multi-task learning
Federated Meta-Learning [33]			NN			meta-learning
Personalized FedAvg [81]						
LFRL [115]						reinforcement learning
FBO [44]			LM	\		Bayesian optimization
Federated MTL [169]	Practicality Enhancement	horizontal	NN	\	centralized	multi-task learning
Federated Meta-Learning [33]						meta-learning
Personalized FedAvg [81]						
LFRL [115]						reinforcement learning
FBO [44]			LM			Bayesian optimization
Structure Updates [95]			NN		decentralized	efficiency improvement
Multi-Objective FL [226]						
On-Device ML [79]						
Sparse Ternary Compression [164]						
DPASGD [128]						
Client-Level DP FL [60]			LM, NN	DP		privacy guarantees
FL-LSTM [130]						
Local DP FL [20]						
Secure Aggregation FL [23]			NN	CM		
Hybrid FL [181]			LM, DT, NN	CM, DP		
Backdoor FL [16, 174, 188]			NN			robustness and attacks
Adversarial Lens [19]						
Distributed Backdoor [203]						
Image Reconstruction [58]						
RSA [100]			LM			
Model Poison [53]			LM, NN			
q-FedAvg [110]			LM, NN	\	centralized	fairness
BlockFL [93]			LM			incentives
Reputation FL [87]						
FedCS [143]			NN			edge computing
DRL-MEC [194]						
Resource-Constrained MEC [192]			LM, NN			
FedGKT [73]			NN			
FedCF [14]	Applications		LM			collaborative filter
FedMF [29]						matrix factorization
FedRecSys [177]			LM, NN	CM		recommender system
FL Keyboard [71]			NN			natural language processing
Fraud detection [222]			NN			credit card transaction
FedML [74]	Benchmarks	horizontal &vertical	LM, NN	\	centralized &decentralized	general purpose benchmarks
FedEval [30]			NN		centralized	
OARF [77]			NN	CM,DP	centralized	
Edge AIBench [70]			\		\	
PerfEval [142]		horizontal	NN		centralized	targeted benchmarks
FedReID [227]						
semi-supervised benchmark [216]				\		

(Source: [8])

The above table is showing a detailed comparison between different models of ML-based data protection. As mentioned earlier FL-based data protection is one such ML model of data protection which does not require any new data sets and this feature has made it much more reliable for data protection in future. The above table is showcasing that FL-based data protection is based on SGD [9]. The above table is also showcasing different SGD-based algorithms such as **LL, NM and DT** which are being used for showcasing neural networks and linear models of data privacy in IoT-powered tools and devices. In the above table, **CM** has been used for indicating ***cryptographic methods*** **and DP** has been used for indicating ***differential privacy*** during data protection in IoT devices [10]. It can be stated by analyzing the above table that the algorithms named **LM, and NN** are efficient for centralized learning models during ML and FL-based data protection. It has been also analyzed from the above table that algorithms like **CM and DP** are efficient for both the centralized and decentralized data protection regime in IoT-powered devices.

However, the mentioned algorithms can be very very problematic during the upgradation of local data set models as well as the global models. These issues can be addressed by framing a ***personalized federated learning algorithm*** [11]**.** This can be very helpful for allowing the data scientist to read the personalized models of the local data set which can result in better data privacy in IoT devices in future.

IV. Conclusion and Future Direction

This entire study has identified different models and algorithms which can be used for ML-based data protection in IoT devices. It can be concluded from the study that traditional ML-based techniques can be better replaced with FL-based models. This advanced ML powered model can provide data privacy without asking for a new data set or user generated information. However, through the mathematical formula it has been identified that SGD and FL-based models have some issues regarding up gradation of global data sets. This can

be improved in future with more advanced ML algorithms. This entire study has also provided a list of algorithms which can help in advance of cryptographic methods for data protection. This study may also help in betterment of differential learning method through CM and DP algorithms for data protection in future.

REFERENCES

[1] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F. and Lin, Z., 2021. When machine learning meets privacy: A survey and outlook. ACM Computing Surveys (CSUR), 54(2), pp.1-36.

[2] Rodríguez, E., Otero, B. and Canal, R., 2023. A Survey of Machine and Deep Learning Methods for Privacy Protection in the Internet of Things. Sensors, 23(3), p.1252.

[3] Zhao, P., Zhang, G., Wan, S., Liu, G. and Umer, T., 2020. A survey of local differential privacy for securing internet of vehicles. The Journal of Supercomputing, 76, pp.8391-8412.

[4] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F. and Lin, Z., 2021. When machine learning meets privacy: A survey and outlook. ACM Computing Surveys (CSUR), 54(2), pp.1-36.

[5] Liu, Y., James, J.Q., Kang, J., Niyato, D. and Zhang, S., 2020. Privacy-preserving traffic flow prediction: A federated learning approach. IEEE Internet of Things Journal, 7(8), pp.7751-7763.

[6] Rodríguez, E., Otero, B. and Canal, R., 2023. A Survey of Machine and Deep Learning Methods for Privacy Protection in the Internet of Things. Sensors, 23(3), p.1252.

[7] ODSC, 2023. ""How You Can Use Federated Learning for Security & Privacy". Available at: https://odsc.medium.com/how-you-can-use-federated-learning-for-security-privacy-ee0c99cf54b3 [Accessed on 14/03/2023

[8] Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X. and He, B., 2021. A survey on federated learning systems: vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering.

[9] Chopra, P., Gollamandala, V.S., Ahmed, A.N., Babu, S.B.G., Kaur, C., Achyutha Prasad, N. and Nuagah, S.J., 2022. Automated Registration of Multiangle SAR Images Using Artificial Intelligence. Mobile Information Systems, 2022.

[10] Chopra, P., Gollamandala, V.S., Ahmed, A.N., Babu, S.T., Kaur, C., Prasad, N.A. and Nuagah, S.J., 2022. Research Article Automated Registration of Multiangle SAR Images Using Artificial Intelligence.

[11] Babu, S.T. and Rao, C.S., 2020, July. Statistical features based optimized technique for copy move forgery detection. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

[12] Kalimuthukumar Sakthivel, Rajesh Krishnasamy, Kannapiran Balasubramanian, Vijayakumar Krishnakumar , Manikandan Ganesan " A revolutionary Partial Resonant Inverter and doubler rectifier with MPPT based on Sliding Mode Controller for harvesting Solar photovoltaic sources", Sustainable Computing: Informatics and Systems, Volume 36 December 2022 100811. https://doi.org/10.1016/j.suscom.2022.100811

[13] Kalimuthukumar Sakthivl,Rajesh Krishnasamy, Kannapiran Balasubramanian, Vijayakumar Krishnakumar and Manikandan Ganesan. "Averaged state space modelling and the applicability of the series Compensated Buck-Boost converter for harvesting solar Photo Voltaic energy", Sustainable Energy Technologies and Assessments, Volume 53, Part C, October 2022, 102611

[14] T. Vino,S.S.Sivaraju,R. V.V. Krishna,T. Karthikeyan,Yogesh kumar Sharma, K.G.S.Venkatesan, G.Manikandan, R.Selvameena,and Mebratu Markos "Multicluster Analysis and Design of Hybrid Wireless Sensor Networks Using Solar Energy", Hindawi, International Journal of Photoenergy, Volume 2022, Article ID 1164613, 8 pages, https://doi.org/10.1155/2022/1164613.

[15] Manikandan, G. and Anand, M. "SEC-TAED Based Error Detection and Correction technique for data transmission systems", Indonesian Journal of Electrical Engineering and Computer Science, Vol.10,No.2, May 2018, pp.696-703, ISSN:2502-4752

[16] S B G Tilak Babu and Ch Srinivasa Rao, "An optimized technique for copy–move forgery localization using statistical features", ICT Express, Volume 8, Issue 2, Pages 244-249, 2022.

[17] Ashokkumar, N., and A. Kavitha. "Transition level energy consumption of NoC (network-on-chip) using data encoding techniques." 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2015.