mathematics

MDPI

Article

Privacy Preservation Using Machine Learning in the Internet of Things

Sherif El-Gendy [1], Mahmoud Said Elsayed [2,*], Anca Jurcut [2] and Marianne A. Azer [1,3]

[1] School of Information Technology and Computer Science, Nile University, Cairo 12677, Egypt; s.elgendy@nu.edu.eg (S.E.-G.); mazer@nu.edu.eg (M.A.A.)
[2] School of Computer Science, University College Dublin, D04 C1P1 Dublin, Ireland; anca.jurcut@ucd.ie
[3] National Telecommunication Institute, Cairo 12677, Egypt
* Correspondence: eng.mahmoud101@gmail.com

Abstract: The internet of things (IoT) has prepared the way for a highly linked world, in which everything is interconnected, and information exchange has become more easily accessible via the internet, making it feasible for various applications that enrich the quality of human life. Despite such a potential vision, users' privacy on these IoT devices is a significant concern. IoT devices are subject to threats from hackers and malware due to the explosive expansion of IoT and its use in commerce and critical infrastructures. Malware poses a severe danger to the availability and reliability of IoT devices. If left uncontrolled, it can have profound implications, as IoT devices and smart services can collect personally identifiable information (PII) without the user's knowledge or consent. These devices often transfer their data into the cloud, where they are stored and processed to provide the end users with specific services. However, many IoT devices do not meet the same security criteria as non-IoT devices; most used schemes do not provide privacy and anonymity to legitimate users. Because there are so many IoT devices, so much malware is produced every day, and IoT nodes have so little CPU power, so antivirus cannot shield these networks from infection. Because of this, establishing a secure and private environment can greatly benefit from having a system for detecting malware in IoT devices. In this paper, we will analyze studies that have used ML as an approach to solve IoT privacy challenges, and also investigate the advantages and drawbacks of leveraging data in ML-based IoT privacy approaches. Our focus is on using ML models for detecting malware in IoT devices, specifically spyware, ransomware, and Trojan horse malware. We propose using ML techniques as a solution for privacy attack detection and test pattern generation in the IoT. The ML model can be trained to predict behavioral architecture. We discuss our experiments and evaluation using the "MalMemAnalysis" datasets, which focus on simulating real-world privacy-related obfuscated malware. We simulate several ML algorithms to prove their capabilities in detecting malicious attacks against privacy. The experimental analysis showcases the high accuracy and effectiveness of the proposed approach in detecting obfuscated and concealed malware, outperforming state-of-the-art methods by 99.50%, and would be helpful in safeguarding an IoT network from malware. Experimental analysis and results are provided in detail.

Keywords: internet of things; IoT privacy; machine learning; privacy; malware detection; obfuscated malware; supervised learning algorithms

MSC: 65P40

Citation: El-Gendy, S.; Elsayed, M.S.; Jurcut, A.; Azer, M.A. Privacy Preservation Using Machine Learning in the Internet of Things. *Mathematics* **2023**, *11*, 3477. https://doi.org/10.3390/math11163477

Academic Editor: Helga Silaghi and Claudiu Raul Costea

Received: 27 June 2023
Revised: 29 July 2023
Accepted: 9 August 2023
Published: 11 August 2023

1. Introduction

Smart devices have proliferated over the past decade, and the internet of things (IoT) has grown in popularity. This is because the IoT plays an important role in a large proportion of most people's daily routines and life [1]. It is a service that enables transmissions between people and objects. Machine learning (ML) technologies have been driving the

development of smart cities and enhancing our daily lives by using the vast amounts of data generated from IoT devices. [2–4]. Transportation, healthcare systems, home automation and environmental control [5] are just a few of the numerous domains in which IoT applications can be invaluable. Moreover, the International Data Corporation (IDC) forecasts that the number of connected devices will reach 41.6 billion in 2025 [6]. IoT will contribute significantly to a significant increase in the volume of data produced as a result of the rapid development in the number of IoT devices; it is anticipated that the amount of data generated globally will reach 180 zettabytes by 2025 [7].

Despite such a promising vision, consumer privacy on IoT devices is a huge worry. Although these data show that IoT has tremendous future prospects, several problems must be solved in order for this technology to be more trustworthy and usable. These difficulties include identity management [8], interoperability [9,10], standardization [11], and IoT greening [1]. Other major issues for IoT include privacy and security [12,13]. These devices consist of sensors that can collect data, process them using built-in circuitry, and send them to a remote location. These data are transmitted to the cloud space, where they are subsequently stored and analyzed to provide the individuals with specialized services.

In contrast, most used schemes do not provide privacy and anonymity to legitimate users [14]. Cascading failures also is one of the key issues affecting the reliability of edge-assisted IoTs, and it falls under the category of security issues in the IoTs and can potentially lead to privacy leaks [15]. As a result of their interconnections, these devices can exchange information through the internet as shown in Figure 1. It follows that individual users' data are collected. In some scenarios, they may contain personal and private information about the user, such as usernames and passwords for online accounts, email contacts and phone numbers, contracts and other essential documents, payment card and other financial information, sensitive photos, and other data.

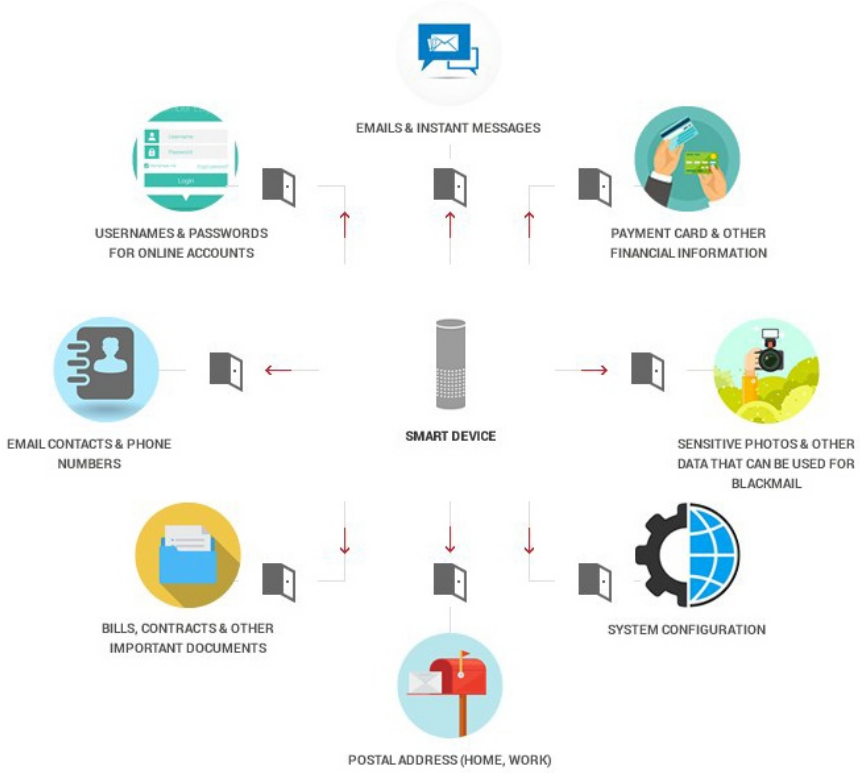

Figure 1. Types of information that cyber criminals can gain through different IoT privacy and security attacks [16].

With these collected data and in light of the seamless connectivity and the continuous interactions among the IoT devices worldwide, there is an urgent need for IoT solutions

that secure the highest degree of security for private information. Data breaches that compromise sensitive information may result in permanent harm. The examples are many. Criminals may utilize a breach of personal information to commit fraud and extortion. A serious threat to national security might come from a data leak in government networks. As an example, a sensitive device's information and password would be sent in the clear in the IoT context. Upon loading the device's web interface and using Wireshark to analyze packets [17], it was discovered that device details were transferred via HTTP in the clear. This provides the current firmware version, its most recent update, and the serial number [18].

Once the user registers into the web interface, the user's information and the device's information, including the password, are transferred in clear text, which is a more dangerous conclusion. As a result, adversaries might simply read the HTTP packet and gain access to critical information. The user's password can be reused across different accounts of the same user [18].

The number of IoT cyberattacks was more than doubled during the first half of 2021, according to "Kaspersky", when 1.51 billion breaches of IoT devices took place, compared to only 639 million breaches in 2020. The Telnet, which is one of the oldest protocols to provide remote access to computers and networks, became the most common unauthorized gateway for attackers into IoT devices. To gain access to IoT networks, the majority of attackers made use of the Telnet protocol, a command line interface that enables remote interaction with a device or server. Over 872 million IoT attacks, or 58% of them, used Telnet. Many of these assaults sought to steal personal information, crash DDoS systems, or mine cryptocurrencies [19].

It is now more crucial than ever to detect malware targeted at IoT devices. Solutions to secure IoT networks must provide protection against malware-based attacks [20]. With the exponential rise in the number and the type of malware, traditional ways of protections (e.g., anti-virus software) cannot defeat it, nor can they provide early detection, hence the urgent need for machine learning (ML) in malware analysis.

Indeed, ML shows significant ability to provide superior approaches for malware detection: the hardware-assisted malware detection framework based on explainable ML [21]. ML is inherently able to handle large volume of data and hence is capable of keeping pace with the fast-changing malware [22]. Because it does not need the production of signatures for each malware family to be detected, ML gains confidence in malware detection.

Implementing ML to improve IoT performance became widely popular. Within IoT, a massive number of devices are connected and voluminous data are generated. These data can be utilized to obtain invaluable information by extracting trends and behaviors and provide estimations and forecasts [14]. From this perspective, the ML-based computational models provide IoT devices with a brain to think, i.e., an embedded intelligence [23]. The use of ML improves the performance of various operations (e.g., data aggregation, access control,authentication, and regulatory compliance), eliminates raw data exchange, minimizes communication overhead and latency, and enhances data protection and security [24].

The contributions of our paper can be summarized as follows:

- Our research makes a significant contribution by thoroughly reviewing and analyzing the literature on privacy activities in the context of the IoT ecosystem, with a specific emphasis on the utilization of ML techniques.
- We present the different privacy threats and attacks that an IoT environment faces.
- We discuss the critical privacy issues and key privacy requirements of IoT.
- We explore and present a review of the commonly used solutions to resolve privacy and confidentiality challenges through ML algorithms in the IoT systems to preserve privacy.
- We validate our work with practical experiments to prove the capability of ML in detecting malicious and anomalous attacks and preserving.

- We discuss challenges and possible directions for using ML algorithms to resolve IoT and privacy challenges.

The remainder of this paper is organized as follows. Section 2 reviews existing studies and the related literature. In Section 3, we present the IoT architecture, GDPR and its implications for IoT. Section 4 summarizes and compares common IoT vulnerabilities, privacy threats and privacy attacks, respectively. Section 5 discusses IoT privacy requirements and the privacy-preserving solutions using machine learning. Section 6 describes the experiments and evaluation used in this study. Experimental results and analysis are provided in Section 7. Finally, conclusions and future work are provided in Section 8.

2. Related Literature

To investigate IoT security and privacy problems, a number of surveys have been written in the past. In [25], security vulnerabilities in several IoT applications are analyzed, while in [26], smart home security is assessed. Furthermore, existing research on possible threats in the IoT environment is examined in [27,28]. Additionally, recent research on privacy and protection from the standpoint of technology and protocols has generated some attention [28,29]. All of these research are primarily concerned with security issues and remedies.

This section analyzes the common ML models used in the context of malware detection within IoT environment and offers prior research, sorted into groups based on how they relate to the various IoT architectural tiers. The related literature uses various structures in describing the IoT systems. Some studies present three-layer [10,28] (various parties have employed service-based architectures [28]), five-layer [30], and seven-layer [30] structures. In this study, we use a three-layer design that consists of the application layer, the network layer, and the perception (physical) layer as depicted in Figure 2 [31]. The next sections examine the most common ML-based solutions for maintaining privacy across the three tiers of the IoT architecture.

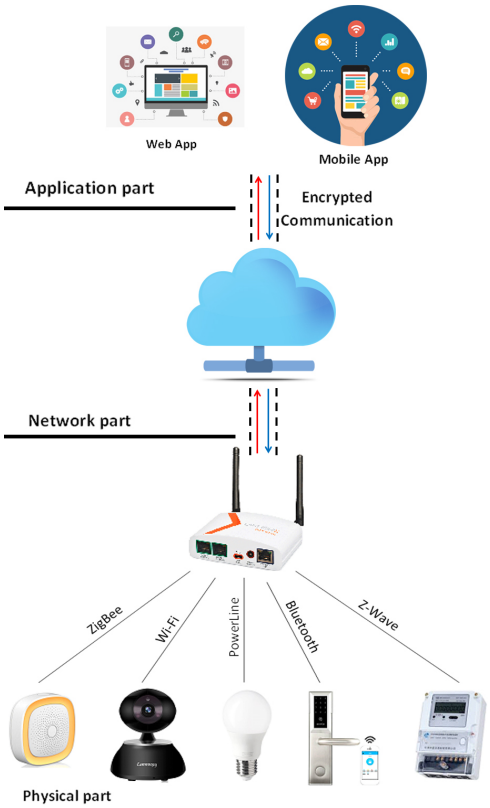

Figure 2. Three layers of IoT architecture [31].

2.1. Perception Layer

Sensors pose the majority of the privacy issues in this layer, which are frequently targeted by adversaries that steal and change data throughout the data-gathering process. One solution is to send grouped data rather than raw data. This strategy is compatible with the energy limits of IoT devices. Furthermore, because it employs representative values rather than specific data fields, it helps to reduce the possibility of data leakage [32,33]. Table 1 summarizes the common ML models in the existing studies and their relationship with the perception layer. According to the survey, the fundamental issue in these research studies is that the extra noise affects the accuracy of the ML models.

Table 1. Common ML models in previous studies and their relationship with IoT perception layer.

Authors	Summary
[34]	The authors outlined a clustering technique that uses aggressive learning neural networks to build a classification model. Sensors have the property of correlating data between geographically close nodes. As a result, clustering networks and data aggregation are critical in wireless sensor networks. They employed Kohonen SOM, which works without supervision, to translate sensor input into context. The suggested technique is CODA (cluster-based self-organizing data aggregation); after sensing the surroundings for a predetermined amount of time, the network is clustered depending on the data and the cluster is redistributed in accordance with the combined value. It works better than typical database aggregation systems for improving data quality.
[35]	In order to enable the base station to access the observations, the authors created a distributed method for doing Principal Component Analysis (PCA). The suggested technique is based on the transmission burden of the intermediate nodes. An intermediate node can provide just one packet rather than relaying all of the receiving packets by using PCA to combine the intermediate node's incoming packets into one packet. This method results in a large reduction in the amount of broadcast bytes for nodes close to the base station. Additionally, the authors created an aggregation service that uses PCA-based aggregation techniques, including the PCAg approach, to compute the reconstruction error at the base station. Using this aggregate service, PCAg may evaluate the algorithm's precision and thereafter dynamically modify its update rate. The thorough simulation results based on the performance metrics for accuracy and efficiency show that the suggested strategy performs better than other approaches of a similar kind.
[36]	In various areas, the authors improved the currently available method for differentially private k-means clustering. They developed a noninteractive technique for differentially private k-means clustering and enhanced an interactive method using a systemized error analysis. The following are the insights acquired by k-means clustering about the subject of noninteractive vs. interactive. The noninteractive EUGkM clearly has an advantage, particularly when the privacy budget is limited. Consider the additional benefit of noninteractive approaches, which allow for additional examination of the dataset. In this comparison, the authors conclude that noninteractive wins. They believe that this trade-off will hold true for a wide range of additional data analysis activities.
[37]	The authors reviewed at least six current papers on the differential privacy frontier. The work built relationships with other subjects and groups in various situations, including statistics, cryptography, complexity, geometry, mechanism design, and optimization. The abundance of new tools, the formulation of new issues, and the productive interaction with other areas all provide rich ground for effervescent growth in an intellectually stimulating and socially important pursuit. Differential privacy has received a lot of interest in recent years, coupled with the use of ML algorithms in this subject, such as clustering, logistic regression, support vector machines, and deep neural networks.
Our study	Our model employing a strategy of sending grouped data instead of raw data. This approach is energy-efficient for IoT devices and reduces the possibility of data leakage, as it utilizes representative values rather than specific data fields. By doing so, our model is better equipped to handle noisy data and improve the accuracy of predictions on the perception layer.

2.2. Network Layer

At this layer, the privacy risk is reduced, and data privacy is maintained by allowing each device to only process separate parts of the data. From this perspective, federated learning [38,39] and distributed deep learning models [40,41] are proved to be suitable models in this architecture. Table 2 presents the commonly used ML models and their relationship with the network layer. Two main drawbacks are identified here. The first is concerned with the use of deep models in fog-enhanced IoT. Each device in these models can handle only a portion of the training data. The deep model performance is largely determined by the amount of training data available. As a result, decentralization can have a negative impact on the accuracy of these models. The second drawback is that noisy data are employed throughout the learning process. For the federated learning strategy, this poses a significant issue. With the help of numerous training modules, who each have access to a smaller amount of training data, the training data are distributed in this manner.

When used in such a dispersed situation, noisy data might have a major influence on the performance of ML models.

Table 2. Common ML models in previous studies and their relationship with IoT network layer.

Authors	Summary
[42]	The authors presented a federated learning technique for edge-enhanced IoT systems. Because this method avoids raw data transfer, only regional and international learning variables between learning blocks should be stated. The authors also created a control algorithm that adjusts the global aggregation frequency to minimize learning loss.
[43]	The authors suggested a privacy-preserving architecture, where deep federated and reinforcement learning are combined and work on IoT platform edge devices.
[44]	The author proposed utilizing low-resource machine learning to cluster data in the behavior category. By means of the unsupervised k-means algorithm, this solution employs smart wearables to deliver analysis of health data with privacy protection in fog nodes.
[45]	The authors proposed a framework for safeguarding data aggregation with federated learning while working with inadequate resources. Messages were routed by the authors using a server employing data from the traffic data category. This server reduces the model's complexity while improving its performance.
[46]	The authors proposed that data be aggregated using edge computing before being sent to cloud storage. To de-identify data, they developed a local differential privacy technique. The authors of this study advised using regression analysis to estimate data distribution. The majority of research studies that employ differential privacy to generate representative data values try to apply the approaches to a specific aggregation function.
Our Study	Our model utilizing federated learning and distributed deep learning models at the network layer. By allowing each device to process separate parts of the data, our model maintains data privacy while reducing the impact of noisy data. Additionally, we take into account the potential trade-off between decentralization and accuracy, ensuring that our model achieves optimal performance, even in fog-enhanced IoT scenarios.

2.3. Application Layer

In this layer, the different applications of ML techniques can ironically cause data leakage [47,48]. The literature shows that, in the case of improperly designed models, adversaries can access or derive sensitive data, which highlights the importance of using well-designed models. In order to address this privacy issue, recent research studies created privacy-preserving ML algorithms. The following ML approaches are investigated in the relevant literature: clustering [49,50], linear regression [51,52], decision tree [53,54], SVM [55,56], logistic regression [57,58], and naïve Bayesian [59,60]. Table 3 depicts studies that use common ML models and their relationship with the application layer. These studies report two main limitations. Firstly, using blockchain-based smart contracts that implement a consensus mechanism along with the reinforcement learning approach reduces the speed of the system. This, in turn, restricts the ability to use extensive data in training the model. The second limitation is related to using user behavior data in deep models. To learn about user behavior using such models, a significant amount of data is needed. The dilemma is that users do not give share their behavior data.

Table 3. Common ML models in previous studies and their relationship with IoT application layer.

Authors	Summary
[61]	The authors introduced a wireless network authentication technique that employs radio channel information to authenticate on the physical layer. The authors employed Dyna-Q and Q-learning reinforcement learning algorithms to determine optimal threshold values about information that might be considered public data on radio channels.
[62]	The authors combined ML and blockchain technologies. A reinforcement learning approach is included in their suggested solution, which enables smart contracts to handle control choices using information from the limited disclosure category, dynamically dependent on environmental input.

Table 3. *Cont.*

Authors	Summary
[63]	The authors employed machine learning role-based provisioning for new applications to be automated. They also proposed solutions to two related issues in this domain: adjusting to changes in job definitions and imposing limits. This research investigated several ML approaches; however, the major focus was on the support vector machine (SVM). This study showed that by keeping an eye on misbehavior data, the access controller may identify questionable objects and limit their access to the system's resources. Machine learning has also been used to examine privacy regulations and legal agreements. These data sources are covered by privacy policies.
[47]	The authors designed and demonstrated a general threat model for categorizing various threats. The essay focused on classification approaches.
[64]	The authors used Google and Amazon ML services to perform a membership inference. They proposed a shadow training approach and evaluated the findings on several datasets, including patient data from a Texas hospital. The findings showed certain serious flaws that allow attackers to infer data records.
[47]	The authors created a protocol for privacy-preserving machine learning. For the neural network model, logistic regression, and linear regression, the authors employed stochastic gradient descent. An offline phase was introduced to a two-server model to encrypt datasets before utilizing them to train ML models.
[65]	The authors provided a method for protecting input information as well as learning parameters. The garbled circuit, a cryptography approach, was used in this study, which relies on analytic and synthetic tools. To increase the framework's speed, the authors proposed an efficient implementation of this approach as well as additional preprocessing processes.
Our Study	Our model incorporates privacy-preserving ML algorithms at the application layer, such as clustering, linear regression, decision trees, SVM, and logistic regression. By employing these techniques, we ensure that user data are protected, and sensitive information is not accessed or derived by adversaries. Furthermore, we mitigate the issue of user behavior data by using anonymized or aggregated data for learning user patterns, thus respecting user privacy and improving model accuracy.

2.4. Common ML Models for IoT Malware Detection

Large strands of the literature have been exploring the different ways of using ML classifiers to detect malware [66,67]. These studies classify the already existing malware samples in order to train ML models and enable them to accurately predict other non-existing potential malware samples. Table 4 presents relevant studies, ML models and their relationship with malware detection, and Table 5 depicts the different ML techniques that are used in privacy-preserving IoT. This study focuses on identifying known malware samples in order to train an ML model capable of making (accurate) predictions on previously unseen malware samples. In this study, we compare many ML techniques, including single decision tree (DT), random forest (RF), and AdaBoost learner. To compare detection accuracy on IoT malware, in addition, we include the k-nearest neighbor classifier (KNN) and support vector machine (SVM) based methods.

Table 4. Common ML models in previous studies and their relationship with IoT malware detection.

Authors	Summary
[68]	The authors described the two-step procedure for detecting malware using machine learning: "feature extraction and classification/clustering". They next went through several feature selection and classification techniques such as SVM, DT, and ANN.
[69]	The authors employed SVM to identify malware in the Android operating system, and their created dataset reached 99% accuracy and precision.
[70]	The authors employed the naive Bayes classifier to detect malware in Android-based IoT devices. Using the naive Bayes classification based on a decision tree, they attained 98% accuracy.
[71]	For the malware detection technique, the authors attained a 97% F-score on decision trees. The researchers also used naive Bayes and logistic regression, which yielded f-scores of 51% and 94%, respectively.

Table 4. *Cont.*

Authors	Summary
[72]	The authors obtained 98.2% accuracy by using a KNN classifier with a fingerprint feature for IoT malware detection on the device layer.
[73]	The authors employed a deep Eigenspace learning strategy to identify IoT malware and obtained 99.68% accuracy.
[74]	The authors discussed three ways for detecting IoT malware: CNN on byte sequences, CNN on color pictures, and CNN on assembly sequences. To build their training dataset, the authors employed 15,000 pieces of IoT malware and 1000 copies of benign ware.Their findings revealed that CNN on pictures and assembly sequences performed better than CNN on byte sequences.
[75]	The authors used ML techniques (supervised, unsupervised, and reinforcement learning) in IoT contexts to identify malware, authenticate users, and control access.
[76]	The authors employed adversarial learning against assaults to detect IoT malware. Off-the-shelf approaches and graph embedding and augmentation (GEA) methods were utilized, with off-the-shelf methods achieving 100% misclassification and GEA methods classifying all malware as benign.
Our Study	By utilizing diverse ML classifiers, our model can accurately detect and predict known and unknown malware samples within the IoT environment. This comprehensive approach enhances the robustness and reliability of our model for malware detection.

Table 5. The used ML models for data privacy in IoT.

Ref	IoT Layers	Machine Learning Technique														
		Gaussian Regression	Self-Organizing Map (SOM)	Principal Component Analysis (PCA)	Regression Analysis	K-Nearest Neighbors (KNN)	Linear Regression	Support Vector Machine (SVM)	Logistic Regression	Decision Tree	Random Forest	K-Means	Reinforcement Learning	Neural Network	Deep Learning	Federated Learning
[77]	Perception Layer	•														
[78]		•														
[34]			•													
[79]				•												
[80]				•												
[81]				•												
[82]				•												
[83]													•			
[36]																•
[84]												•				
[85]												•				
[42]	Network Layer															•
[43]													•		•	•
[86]														•		
[87]														•		
[88]															•	
[44]												•				
[45]																•
[89]																•
[46]					•											
[90]					•											
[61]	Application Layer													•		
[91]									•							
[62]													•			
[92]														•		
[93]															•	
[63]								•								
[94]						•		•		•						
[95]										•						
[96]															•	
[47]							•		•					•		
[65]														•		
[97]														•		

Prior demonstrations concentrated on building classifiers or deep learning models using IoT malware datasets. This research focuses on recognizing existing malware samples in order to train an ML model capable of generating accurate predictions on previously undiscovered malware samples. We examine various ML approaches in this work, including single decision tree (DT), random forest (RF), and AdaBoost learner. This is in addition to the k-nearest neighbor (KNN) and support vector machines (SVMs) based approaches for assessing detection accuracy on IoT malware.

3. Background

This section sheds light on IoT architecture in Section 3.1 and the importance of GDPR and its implications in IoT in Section 3.2 for better understanding of related privacy issues from regulatory perspective.

3.1. IoT Architecture

While the IoT architecture design has not been agreed on universally, a three-layer architecture is the most common and frequently accepted model as shown in Figure 2. It contains the following layers [31]:

- Perception layer: This is the architecture's physical layer. The sensors and associated equipment are used to collect various data quantities, depending on the project requirement. These may be edge systems, sensors, and drives interacting with their surroundings.
- Network layer: The function of the network layer is to transfer and process information that is collected by all of these devices. These devices link with other smart things, servers and network devices. It is also responsible for data transfer.
- Application layer: The user interacts with the application layer. It is responsible for providing the user with specialized application services. For instance, this can be an intelligent home implementation, where users tap on an app to switch on a coffee maker.

3.2. GDPR and Its Implications

The newly enacted general data protection regulation (GDPR) intends to reinforce user rights and establish new criteria for data management, while encouraging greater user participation in privacy protection [98]. Data protection through architecture and GDPR are used to ensure solutions centered on remote and portable services, and ensure that network access cannot be restricted to authorized users with predefined permissions. It may also be used to check data completeness, transfer, and retrieval. In general, cloud computing security issues are classified into three categories.

The IoT sector is currently confronted with the EU's "General Data Protection Regulation (GDPR)" [99], which went into effect in 2018 [100]. The regulation lays out important principles for creating parity between users and third parties. It also establishes standards for the protection of user data in the IoT and fundamentally alters how data are handled across all applications. The GDPR tackles concerns such as what categories of data can be processed and in what situations, as well as the reasons for data collection, the amount of data that can be collected, the required period of data retention, and the rights of users about their data.

It thus places a greater emphasis on the rights of users, whose data are being processed, including the rights of notice, access, rectification, deletion, restrict processing, data portability, objection, and the right to stop automated decision making [101]. Simply put, the GDPR seeks to return ownership of personal data to the user. Because the IoT relies on extensive user data gathering and sharing, the risk to user privacy grows. Traditional privacy measures must shift their attention from service providers to users in this regard.

Privacy by design and privacy by default are required under GDPR for any enterprises that gather and process user information [102]. Designing services with data protection and privacy in mind is known as privacy by design, and it requires enterprises to design all services that process personal data with this in mind. Default privacy states that, with no user input necessary, all public services must employ the most stringent confidentiality settings by default.

The authors in [103] studied the effect of the incorporation of the GPDR act into ML models to prevent model reversal and membership inferences. These investigators concluded that certain ML models may have to be classified in the new regulations as personal data.

GDPR compliance is included in the relatively new privacy-preserving approaches that have been proposed in the literature. Users' consent and the right to access are fundamental aspects in personalized data repositories. In order to exercise the right to be forgotten, people can request that organizations that violate their privacy preferences remove their access to their data.

The users' informed consent is required before companies can perform data analysis on them. With these solutions, consumers may also express their privacy preferences and verify that the service providers are abiding by them. The ability of users to utilize their right to access the information to which they are entitled may be hindered because data-summarizing systems employ changed data. Furthermore, although cryptography-based techniques may provide privacy by design and default, they may also restrict the right to access acquired information.

Another difficulty with ML-based approaches is that they might be violating the right of access to information by revealing sensitive facts about how businesses train their ML models. In our perspective, it is difficult to use the right to forget ML approaches after data have been used to train a model for ML; therefore, they should be classified as personal data as indicated by [103], even if the effects of data points on the trained model may progressively diminish.

4. Common Vulnerabilities, Privacy Threats and Attacks in IoT

This section explores, compares and analyzes common vulnerabilities, privacy threats and attacks in the IoT context in Sections 4.1–4.3, respectively.

4.1. IoT Vulnerabilities

IoT systems or projects that include flaws let unauthorized individuals send commands, gain unauthorized access to data, or perform DDoS attacks [104]. Vulnerabilities can be found across the IoT system. This might include deficiencies with the system's hardware or software, policy decisions made by the system or its user, and more [105].

The Open Web Application Security Initiative (OWASP) project aims to promote and help manufacturers in designing their devices with security in mind, resulting in safe gadgets by design. Its purpose is to assist companies and people in determining acceptable risk and taking suitable risk-mitigation steps.

The OWASP revised its top 10 IoT vulnerabilities in 2018 [106]. The list includes (a) hard-coded or easily crackable passwords; (b) network services that are not secure; (c) ecosystem interfaces that are not secure; (d) a lack of secure update mechanisms; (e) the use of faulty or outdated components; (f) insufficient privacy protection; (g) insecure data transfer and storage; (h) a lack of device management; (i) insecure default settings; and (j) a lack of physical hardening.

The OWASP top 10 IoT vulnerability list adopts a unified approach to tackling IoT vulnerabilities that may affect IoT devices rather than including disparate criteria for various stakeholders. The top OWASP IoT project teams steered clear of specific IoT security vulnerability guidelines. The OWASP project compiled a list of attack zones and related vulnerabilities for the IoT environment. This is shown in Figure 3 below [18,107].

4.2. Common Privacy Threats in IoT

The dynamic nature of IoT in terms of technologies and capabilities, as well as the emergence of new ways of interacting with the IoT, creates unique privacy concerns and challenges. This section describes how we classified and summarized those common threats.

1. Identification Threat:
   If an individual's address and name, or any other type of pseudonym, can be linked together and information is inferred, this is called identification. The threat is to link a particular private identity, to breach the context, and further threats are activated and facilitated. The profiling and tracking of individuals, or the collecting of multiple

data sources are only a few examples of related threats. Current threats to identity are concentrated in the backend services, where massive volumes of data are collected and stored in a single location that is out of the subject's control [108,109]. IoT systems that prioritize local processing over centralized processing, as well as horizontal interactions over vertical interactions provide the greatest obstacle in identifying users, as only a minimal quantity of identifying data are available outside their own private worlds. This threat is rated as the most frequent and has an impact on network layer information processing [110].

2. Localization and Tracking Threat:
   It is possible to determine and document an individual's physical location in time and space through localization and tracking and recording without permission or consent. Identifying a person is necessary for tracking to be a continuous process [111]. A variety of methods are available for tracking, such as mobile phone location or internet traffic. As a result of this danger, the vast majority of tangible privacy violations have been identified: GPS stalking, publication of private information, or the general feeling of being pursued. Localization and tracking are most dangerous during the processing of data, when back-end locations traces are generated without the subject's knowledge. On the other hand, the key hurdles in tracking and localization lie in making people aware that they are being tracked, controlling the sharing of location data in indoor spaces, and developing privacy-preserving protocols for connection with IoT devices that impact all layers of the IoT architecture [112]. Figure 4 shows locating cell phones by IoT sensors in comparison to the cell towers that are fewer in number than IoT sensors. As a result, discovering locations using IoT sensors is more convenient and accurate than using cellular networks.
3. Profiling Threat:
   Profiling refers to the risk of data files being collected or arranged by people to identify interest in connection to other profiles and data. The profiling approaches are generally employed for e-commerce personalization (recommenders systems, newsletters and advertising), but also for customer demographic and interest-based internal optimization. To personalize services, users are profiled; however, this often leads to unwanted advertising, price discrimination, or biased automatic choices. This threat is more prominent because numerous information sources are available in the IoT ecosystem, and it is possible to gather complete information on individuals and to associate user preferences with other profiles. This has an impact on how information is processed at the network layer, especially when it is important to share or exchange data with other parties [111].
4. Interaction and Presentation Threat:
   With this threat, personal information is disseminated via public media to those who should not know it. Many IoT applications, include systems for industry, infrastructure, and the medical and healthcare fields, and so on, necessitate numerous connections between the device and the user [5]; it is feasible that information is delivered to users through the use of smart items in the surrounding environment, for instance, using lighting methods and television or computer screens to display videos. To put it another way, users take control of systems using an instinctual methodology that makes use of smart objects in the surroundings. However, several intercommunications and organizational operations are inherently public. This causes privacy concerns when the user and the system exchange confidential information [108,109].
5. Lifecycle Transitions Threat:
   Privacy is a major concern when smart objects divulge their private information over the course of their life cycle as management domains change. This problem is noted with regard to damaging images and videos commonly viewed on cameras and other modern gadgets. Many customer support products now have life cycles that are designed to be continuously purchased once, even when the results have not improved. Smart items can attribute to a more interesting life cycle, involving

exchange, loan, donation and disposal. We thus acknowledge the necessity for adaptive outcomes which plainly pose certain challenges. Some life cycle adjustments, like sharing a smart item, need to hide information for a while. The owner can unclamp the classified information and continuously monitor this device [108,109].

6. Inventory Information-Gathering Threat:
It is described as the unauthorized gathering of information on the actuality and characteristics of personal equipment. This vulnerability is mostly caused by the sensor devices' communication capabilities, which allow illegal access to or information gathering. The communication pattern and other recognizable elements can also be seen by unauthorized entities, and the existence of devices can also reveal the model and the kind of device from that information.Inventory lists can provide information about user preferences, which law-enforcement authorities or burglars can use to conduct illegal searches or plan targeted break-ins. Moreover, inventories can reveal user preferences [108,109]. When it comes to defeating IoT inventory threats, there are two types of challenges: (i) query validation—an effective defense against agile inventory assaults begins with making smart objects capable of validating queries sent by authorized entities and responding appropriately; and (ii) fingerprinting mitigation—techniques that defend well-being are required to protect fingerprint transmissions of smart devices and prevent passive inventory assaults [111].
7. Linkage Threat:
This threat connects previously independent system devices, such as the collection of information on various data, which were never exposed to previously opaque sources. The users have no idea about the inferior evaluation and lost data that come from combining diverse data and authorizations. The increasing expansion of unknown data is another example of linking violating privacy [108]. There will be an increase in linkage hazards as the IoT develops for two reasons. The first is that by connecting systems from many firms, a parallel interconnection can eventually create a diversified system that offers novel services that no one system has ever provided on its own. Second, the successful connectivity of such things necessitates a fluid exchange of information and ongoing maintenance among various stakeholders [113].

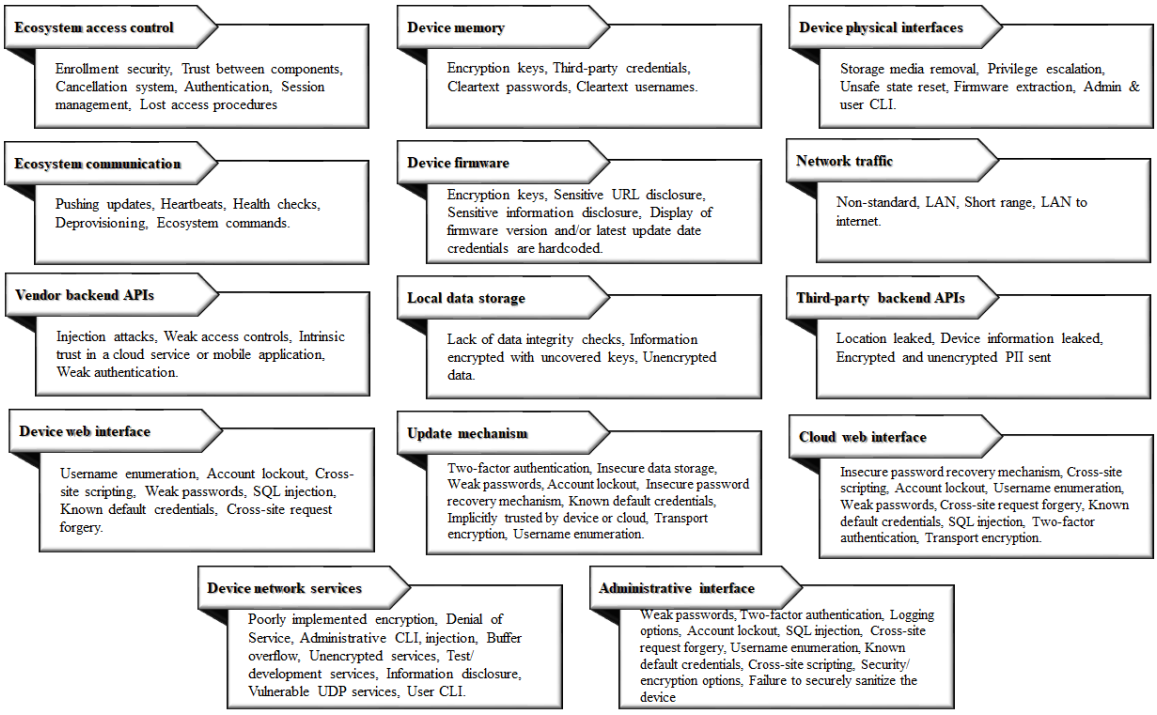

Figure 3. OWASP IoT attack surface and associated vulnerabilities.

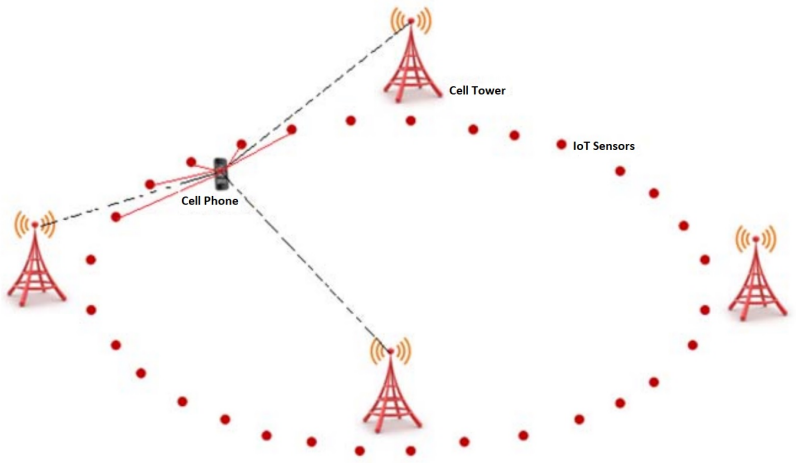

Figure 4. Locating cell phone by IoT sensors [114].

In Table 6, we summarize the most frequent IoT privacy threats and their effects.

Table 6. Summary of identified privacy threats and their effects.

Threat Name	Threat Description	Threat Effect
Identification	It refers to the risk that a person or their data may be linked to a permanent identifier. For example, an individual from a database or collection can be linked by name, pseudonyms, images, voice or an address.	Link a particular private identity to breach the context, and those further threats are activated and facilitated.
Localization and Tracking	Using various methods, such as GPS, internet traffic, or smartphone location; To identify and record an individual's physical location over time	Determine and record a person's precise location in time and space, capturing it without the subject's knowledge or consent.
Profiling	Users are profiled for customization; data files are collected or arranged by people to identify interest.	Leads to undesirable advertising, pricing discrimination or automated biased judgments.
Interaction and presentation	The potential of encroaching on user privacy by sending specific individualized private information through a public medium.	This causes privacy concerns when the user and the system exchange confidential information
Lifecycle transitions	This issue arises when IoT devices are changed ownership. The majority of IoT gadgets are offered with the premise of "purchase once, use forever," and they accumulate a ton of personal data over the course of their lifespan.	When smart objects reveal their private information over the course of their life cycle as management domains change, privacy is a major concern.
Inventory attacks	This is mostly because sensor devices communicate and unlicensed data access or collection is permitted. Those inventories can provide user preferences information that can be used illegally.	Inventory lists can reveal information about user preferences, which criminals might use to perform unlawful searches or organize targeted break-ins.
Linkage	The combining of data from various sources may expose details about people that they did not initially agree to disclose.	Users are unaware of the poor evaluation and lost data that result from mixing several data and authorizations. The rapid proliferation of unknown data is another form of connection that violates privacy.

The literature on IoT threats shows that location monitoring 31.5 percent is the hazard that people are most concerned about, followed by the sharing of unanonymized data 25.9 percent. Concerns about profiling were brought up in 21.3 percent of the research, followed by inventory attacks at 8.3 percent, interaction and presentation at 6.5 percent, life cycle transitions at 3.7 percent, and connecting at 2.7 percent (Figure 5) [108,109,115,116].

4.3. Common Privacy Attacks in IoT

Some of the most frequent IoT privacy attacks are briefly described below. In addition, Figure 6 links identified privacy threats with associated privacy attacks. Additional assaults on IoT eco-system elements (such as databases and ML models) are detailed in [117,118], which, in turn, jeopardize users' privacy. In the following, we present some of the IoT privacy attacks.

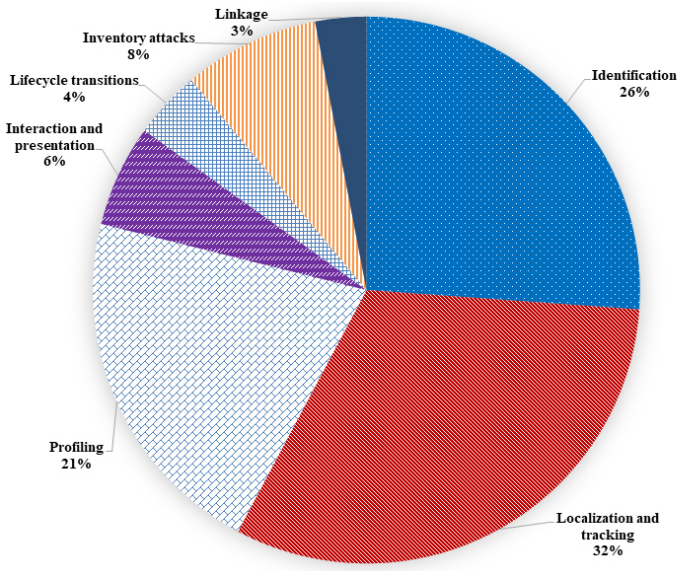

Figure 5. Highlighted IoT privacy threats [109,116].

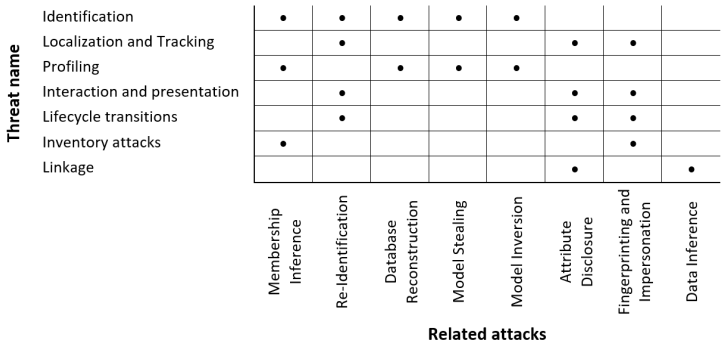

Figure 6. Identified IoT privacy threats with associated privacy attacks.

1. Membership Inference Attack: The attackers may use this attack to find out whether a certain data record was utilized to form the ML model or not, as the opponent knows the ML model and each data record [64,119]. Data privacy is compromised in this attack if an individual is sensitive to inclusion in a training set. A health-related ML model that contains that person's health information as a data record, for instance, leaks information. This attack concerns a person's identification, can help with profiling, and can take advantage of connecting and inventory attacks in terms of personal data security issues.
2. Data Inference Attack: The attack is often linked to the encryption-based privacy preservation methods, as observed by [120]. It attempts to retrieve some information about a particular data record via the linkage with public information, also making tailor-made system queries to check answers to discover whether information about underlying data is being leaked. The frequency analysis used for deciphering is a famous case of this attack.
3. Attribute Disclosure Attack: The disclosure of an attribute happens when data records allow a person to obtain more accurate characteristics [121]. In other words, the data release reveals fresh information about certain people. This attack usually leverages links to generate information from various data sources.
4. Fingerprinting and Impersonation Attack: An attacker might monitor a device's communication behavior and attempt to imitate it with an inventory of attacks [122,123]. If privacy is breached, the attacker might obtain device credentials to modify user privacy choices and introduce fake data into the system.

5. Re-Identification Attack: An attacker can use this technique to re-identify a record from outsourced data records, or public or open data records by combining data from other collections [124]. Re-identification is a relatively common attack, with a traditional example being the voters list being exploited in 1997 to re-identify the state health records of government officials [125].
6. Database Reconstruction Attack: According to [126], when statistics information is published in research organizations, sensitive data may be exposed to data base reconstruction attacks. This enables the rebuilding of the original databases in part or whole, which can allow some users, depending on their relationship with specific features, to be identified or unintentionally profiled in the target database.
7. Model Stealing Attack:Internal training settings and other sensitive ML details may also be re-established or disclosed by using model-stealing approaches [127–130]. This discloses sensitive information on the training data used for these algorithms and can lead to people being profiled unintentionally.
8. Model Inversion Attack:By following the forecasts of the ML model, model reverse-invasive attacks allow attackers to obtain data from underlying training as shown in [47,95,131]. As a result of this attack, a particular record of training cannot always be retrieved. Instead, the attacker extracts an average representation of equally categorized inputs. However, if exposed classes are sparsely filled, i.e., a class might correspond to a single person in a record [95], this might be a great risk to privacy.

IoT systems are also vulnerable to other different security attacks that could affect data privacy, such as the following:

1. DoS attack:To prevent IoT devices from accessing services, attackers send too many queries to the target server [132]. When DDoS attackers send requests for IoT services from hundreds of IP addresses, the server has a tough time distinguishing between real IoT devices and attackers. DDoS attacks are especially dangerous to distributed IoT devices that use lightweight security mechanisms [133].
2. Jamming:Attackers broadcast fake signals during failed communication attempts to interfere with IoT devices' continuing radio communications and further deplete their bandwidth, energy, CPU, and memory resources [134].
3. Spoofing:With the use of its identification, such as the MAC address and RFID tag, a spoofing device impersonates a valid IoT device in order to gain unauthorized access to the IoT system. Assaults like DOS and MiTM attacks might potentially be carried out by it [61].
4. Sniffing/Man in The Middle attack (MiTM):Jamming and spoofing signals are sent by a man-in-the-middle attacker to covertly track, spy on, and alter the private communication between IoT devices [132]. The sniffer intercepts traffic coming in and out IoT devices, for example, passwords, emails, credit card data. A Wi-Fi router is the chosen target because it stores all network traffic data and can be used to control any device linked to it, including PCs and cellphones.
5. Software attacks:Mobile malware, such as Trojans, worms, and viruses, can cause privacy breaches, economic losses, power depletion, and network performance deterioration in IoT systems [135]. Users may unwittingly click on malicious links or download infected software; smart TVs and other similar gadgets are particularly susceptible to this sort of danger.
6. Password Attacks: Password assaults, such as dictionary or brute force, target a device's login information by blasting it with numerous password and username permutations until it discovers the correct one. Because most individuals choose a basic password, these attacks are fairly effective. Furthermore, approximately 60% of users repeat the same password [136]; individuals generate passwords and may choose to use the same password across several websites, accounts, and gadgets. These actions rely on the characteristics of users, including their demographics, situational awareness, psychology, and cognitive abilities [137]. Because many users share the same password across several websites or because systems have prioritized third-

party access in their system architecture [138], authors of [139] discovered that more secure sites are vulnerable to less-secure sites. Despite extra security measures, human nature encourages users to use the same or slightly modified passwords for several accounts or multiple users using the same password. As a result, if an attacker gains access to one device, he or she gains access to all devices.

5. IoT Privacy Requirements and Preserving Solutions

In this section, we cover the IoT privacy requirements in Section 5.1, and current related privacy-preserving solutions in the context of IoT in Section 5.2.

5.1. IoT Privacy Requirements

IoT privacy requirements can be divided into two categories, "institutional" and "technical" requirements [140–143] as shown in Figure 7.

1. Institutional requirements:
   (a) Regulation to keep pace with technology: In the internet-of-things age, information is gathered when a person connects to public and work spaces. This information can recognize user actions within the private domain. Behavioral evaluation can even extend to the realm of mobility via data obtained via linked gadgets.
   (b) Standards: One crucial component of this convergence is the development of privacy standards. They can function on a global scale across countries, allowing governments to develop these initiatives by incorporating industry guidelines into legislation and establishing formal standards for third-party verification/auditing of standard provisions. It is also important to assist enterprises in tracking data streams in IoT applications in order to address any special privacy rules that may be in place for different nations or locations.
   (c) Planning privacy by design: The adoption of privacy by design methodologies from the beginning of solution development is required to provide end-to-end privacy protection. Endpoint hardware security, communications security, protocol dependencies and other requirements, such as firewalls, network segmentation, supporting computing and storage system security, should all be considered upfront to ensure data life cycle protection as it traverses IoT systems.
   (d) Cultural expectations: Privacy by design, concentrated senior leadership on possible risk and remediation efforts, and improved collaboration across operational departments are all essential cultural adjustments inside firms that can drive greater privacy protection in IoT contexts.
2. Technological requirements:
   (a) Sensor to edge versus multiple endpoints: While business requirements eventually decide the right architectures, there are two main models for analyzing privacy requirements: an end-to-end solution and an architecture in which network termination happens at the local level (the edge).
   (b) Device constraints: In terms of functionality, underlying vendor technology, and even models, IoT sensors represent the apex of device variety. However, whether the device is of a low form factor (such as a chip in a car) or high form factor (typically a new device, such as a PC or laptop, that is very similar across verticals, has computing power, and can run an operating system and a controller), security needs must be addressed across the IoT solution.
   (c) IoT scale: In deployments involving hundreds of sensors, implementation is often staggered. Privacy and other management needs are fine-tuned on the second or subsequent phases, a method that might introduce new risks. Architecture, tools, management, and procedures must all be synchronized in

large-scale deployments, a project management reality that requires time, even though a staggered approach may introduce additional dangers.

(d) Solution maturity: The risk of "fail fast" solution creation is heightened because acceptable levels of security, privacy, or regulatory concern may not be given proper priority. A small vendor or startup developing a connected device may be more concerned with bringing a minimum viable product to market quickly than with ensuring proper privacy protection on the device. Smaller companies are less likely to have devoted adequate attention to system development and life cycle approaches that incorporate threat modeling across IoT components.

(e) Data governance: Because various countries have varying data management standards, the location of data acquisition and storage must be addressed. Aside from privacy concerns, big sensor installations may collect information that poses a new threat to national security. For example, environmental or seismic data may be altered or used for terrorist objectives.

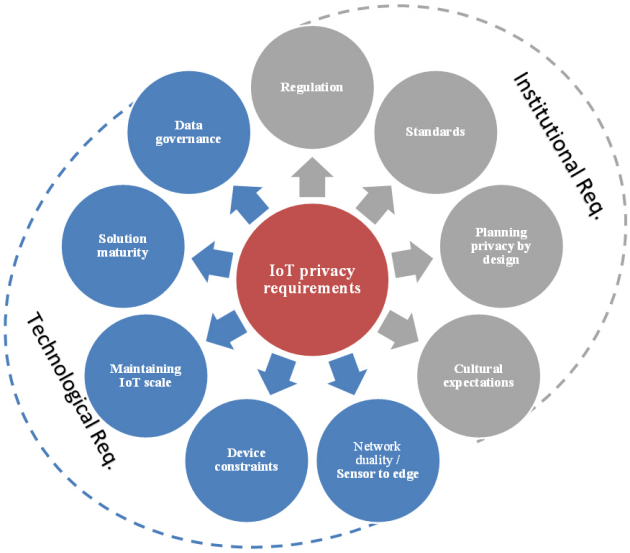

Figure 7. IoT privacy requirements.

5.2. Privacy Preserving Solutions

There are numerous privacy-preserving solutions that could mitigate the risk of IoT privacy threats and prevent attacks on user privacy. We summarize them as follows:

1. Data Perturbation mechanisms:
   These methods use a sequence of processes to alter or conceal private information in the original data [144]. As a result, techniques such as noise addition and anonymization are used.

   (a) Noise Addition mechanisms:
   These approaches aim to modify confidential characteristics by introducing noise to the original information in order to avoid a person's identity [145]. These may be classified into four categories as follows:

   i. Data-sampling mechanisms, which seek to provide a new table, including solely sample data for the entire population.
   ii. Random-noise mechanisms, which involve adding or multiplying a randomized number to the value of the sensitive characteristic.
   iii. Data-swapping mechanisms, which change a data subset by creating ambiguity over the real value of the data [146].
   iv. Differential privacy mechanisms, by adding Laplace noise to the outcome of the database query [145].

(b) Anonymization Protection Mechanisms:
By deleting any explicit identifiers, these methods obscure the identity of the data owner and reduce the accuracy of the information. The k-anonymity [147], l-diversity [148] and t-closeness [121] techniques are considered well-known privacy-preserving approaches. The k-anonymity formal technique is developed to fight the re-identification problem induced by the quasi-identifier features; k-anonymity, on the other hand, is vulnerable to assaults using preexisting knowledge. The researchers have therefore developed other versions, such as l-diversity [148], the main idea being that in each quasi-identifier group, the different values for the sensitive attribute should be present for at least l distinctions, and the method for t-closeness [121] requires distribution in each quasi-identifier group of a sensitive attribute that is close to distributing the attribute in the table as a whole.

2. Data Restriction mechanisms:
Data use can be restricted using these methods, which either prevent access or encrypt inputs. Access control and cryptography-based solutions are two options for limiting access to sensitive information.

(a) Access Control:
These methods are efficient in ensuring data exchange [144]. Owners of data can specify their personal preferences for who has access to what data and how those data can be manipulated by others. Role-basic access control (RBAC) and attribute-based access control (ABAC) are examples of controls. When it comes to assigning access permissions, RBAC uses roles, but ABAC uses attributes, such as the resource and environment attributes of a user's role [149].

(b) Cryptographic protection:
When it comes to privacy preservation, these methods are heavily used. There are three major categories under which they fall:

i. Secure multiparty computation combines inputs of scattered entities for the production of outputs while safeguarding the input privacy of the individual [146].
ii. Symmetric/asymmetric encryption employs data-protection keys.
iii. Public key infrastructure provides the entity with a certificate to ensure that the specified entity holds a public key.

Despite the fact that many sensors are unable to provide acceptable security procedures due to the limited number of storage and processing resources, encryption remains the most dominant technology in nearly all currently suggested solutions [150].
The cryptographic method known as homomorphic encryption [151], on the other hand, enables processing on encrypted data directly. It enables the execution of quadratic, addition, and multiplication operations. Furthermore, homomorphic encryption offers privacy-preserving capabilities in both the training and classification stages of ML models, in contrast to the majority of earlier studies, which only concentrated on the training step. The following are the divisions into partial and full homomorphic schemes:

i. Partial homomorphic schemes:
In contrast to arbitrary computation on ciphertexts, they offer restricted operations on ciphertexts, such as addition and multiplication, in addition to other operations. Due to their lower computational cost, these approaches perform significantly well and outperform fully homomorphic systems. However, fewer algorithms may be used because of the constrained amount of operations [152].
ii. Fully homomorphic encryption (FHE):

This method allows for unrestricted computation in addition to quadratic functions and multiplication and addition on ciphertexts. Since they provide unconstrained processing, classifiers created using this schema are naturally privacy-preserving and more suitable for real-world applications in terms of privacy guarantees. There are not many completely homomorphic encryption systems though, and they are typically expensive to compute, requiring two to five seconds for each operation [152]. Although certain efficient FHE algorithms have been devised [153], it has been demonstrated that they are susceptible to data inference attacks, such as those that recover encryption keys and decode data in situations when the message is known (broadcast) and the message is unknown (secrets) [154].

iii. Data Minimization Principle
Because of this, IoT service providers must restrict or concentrate personal data only when it is truly necessary. The data should also only be kept for as long as it is necessary for the technical services. Other options, such as hitchhiking, have been proposed as alternatives to the above four options. This is a fresh method for protecting the privacy of people who share their physical locations online. Hitchhiking apps treat locales as though they are objects of study. There is no longer a trade-off between fidelity and knowing who is at a specific spot [111].

3. Decentralized Machine Learning:
Technologies for decentralized ML offer a new paradigm for computing that improves privacy, instead of transferring potentially sensitive user data to a computer. End-user devices are used to offload some calculations, and each one updates the system model in part. By doing this, the risk of exposing the service provider and other trustworthy but motivated environmental foes to sensitive and confidential raw data is reduced. Federated machine learning [38,45] in recent years has been more common in ML and recommendation systems, and is being extensively explored and utilized [39,155]. It offers the construction of a global model through the learning of user-pushed updates. Since the technology is still in its infancy, the IoT ecosystem's distributed computing systems are taken into consideration. It is very adaptable and efficient. However, it is important to consider how this approach may be implemented in a variety of applications and usage scenarios. Inference attacks may be possible with federated machine learning [156]. Ref. [157] presents preliminary findings that might reveal very important user data.
4. Multi-tier Machine Learning:
On sensitive data, open ML models enable data memorization during training. This approach uses many layers of training, which can reduce the impact of unique, sensitive training data on the resulting models. A multi-level ML technique called semi-supervised aggregation and transfer of knowledge [158] proposes a hierarchy of "teacher" and "student" models. In order to preserve a student model's privacy, teacher models aggregate sensitive data divisions directly while using student models to create non-sensitive data. Different privacy is used in this method to specify the privacy-protecting features throughout the training stage of student models. Because the models of students alone are released, model inversion attacks cannot affect the original training examples because they are not the absolute majority of teacher models' categorization decisions that are utilized for the training method. This approach with high confidentiality, suitable for a large variety of ML models, is relatively recently distributed. The usefulness of the recommendations in terms of quality nevertheless must be examined.
5. Output Obfuscation Techniques:
User reconfiguration through model inversion attacks may be stopped by concealing the output of ML models from the specified range. A technique called differential

privacy is intended to increase the precision of statistical database searches while reducing the likelihood that their records may be recognized [159]. ϵ-differential privacy provides the differential protection of privacy by adding a selected random noise to the real answer of an ML model, using a Laplacian distribution. This indicates a constant incertitude in all measures, which means that a given record is less likely to be exposed. However, due to specific functional constraints, differential privacy alone cannot offer safeguards for all scenarios. The most frequently sought and implemented approach for protecting privacy in the present age may be characterized as differential privacy. Along with other techniques, it is used to construct privacy preservation apps and services because it is very successful against model reversal and inference assaults [119,158,160].

6. Ensuring Privacy With Dataflow Models:
   The technique enables the development of data flow models at each level with corresponding authorizations for ensuring user privacy and transparent responsibility.
   (a) Ensure privacy and verification by using Blockchain:
   Researchers suggested that a blockchain be used for verified data collection, storage and access accountability in IoT environments [161,162]. For example, data from blockchain can offer flawless records and allow cloud accountability [163]. In addition, as assessed in [164], blockchains are expanded for IoT applications in medical care. However, research on scalability in blockchains may be made available such that they are ideally adapted to IoT settings.
   (b) Languages and platforms of privacy programming:
   These solutions need prior information flows and privileges so that all data components are connected to the relevant policies. This means that the data items must be stated previously [122,165]. As an example, Jeeves is an additional library with Java, a privacy-oriented programming language [166]. Homepad [167] apps are used as guided element graphs (instances of functions that process data in isolation). It allows the program to check its flow chart automatically against user-defined privacy standards with minimal overall computing costs by modeling these aspects and the data flow chart. Furthermore, [168] provides certain privacy standards for the development of IoT applications.

In Figure 8, we summarize the privacy-preservation solutions discussed earlier, while in Table 7, we analyze the privacy-preservation solutions and discuss their limitations.

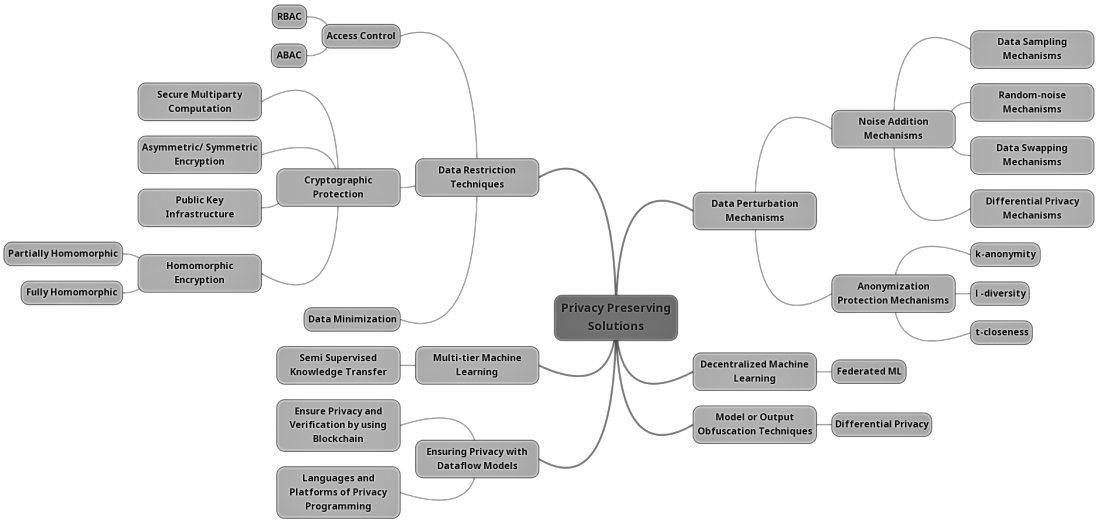

Figure 8. Possible privacy-preserving solutions mindmap.

Table 7. IoT privacy-preservation solutions analysis.

Preserving Techniques		Advantages	Relevant Privacy Threat(s)	Limitations	Relevant Attack(s)
Anonymization	k-anonymity [121,169]	Easy to implement	Identification, localization and tracking, profiling, linkage	Varied data are needed	Re-identification, Database reconstruction, Data inference, Attribute disclosure
	l-diversity [121,170]	Minimally complex		Varied data are needed	Attribute disclosure
	t-closeness [171]	Preserving delicate characteristics		Requires strong dataset diversification	
Model or output Obfuscation	Differential privacy [119,160,172]	Easy to integrate with solutions	Identification, profiling, linkage	Works for low sensitivity data queries	Model Inversion, Inference attacks
Multi-tier ML	Semi-supervised knowledge transfer [158]	Distributed, applicable to any ML model	Profiling, linkage	Effect on accuracy of ML models is unknown	Model stealing and inversion, Inference attacks
Decentralized ML	Federated ML [157]	Highly scalable and efficient	Inventory attacks, linkage, profiling	Potential information leakage	Inference, Fingerprinting and impersonation attacks
Cryptography	Fully Homomorphic encryption [154]	Private ML models training and classification	Inventory attacks	Large computational overhead	Data inference (data/key recovery)
	Partially Homomorphic encryption [154]	Relatively lower computational overhead		Not applicable to all ML models	Inference attacks
Data summarization	Public-private data summarization	Low accuracy loss and very effective solution	Identification	Unquantified Privacy guarantees	Inference attacks
Data flow models	Blockchain for privacy [163]	Verifiable privacy	Inventory attacks	Cost of computation, limited scalability	Fingerprinting and impersonation attacks
	Technologies and programming languages that safeguards privacy [122]	Low overhead with verifiable privacy		Information flows must be announced in advance	

6. Experiments and Evaluation

The core idea behind this work is to tackle the privacy issues in the IoT environment. As discussed in the previous section, several solutions have been proposed to solve this problem. However, such solutions have several limitations and are not able to maintain long-term network protection. In contrast, ML techniques give defense systems the capacity to learn for themselves from a sizable dataset in order to spot hidden patterns and make choices without explicit guidance [173]. Motivated by the potential of ML in many real-world applications, we also utilize different ML algorithms to solve the problem of privacy in IoT. For this purpose, we use "MalMemAnalysis" dataset as a case study; this dataset focuses on simulating real-world privacy related obfuscated malware as closely as possible, such as spyware, ransomware, and Trojan horse.

In this section, we provide a succinct explanation of the experimental setup and classification outcomes of several ML algorithms. The MalMemAnalysis dataset is subjected to a number of experiments and analyzes utilizing a variety of methodologies. The performance of ML in known and unknown attacks is evaluated using some attack classes in the testing phase. These attack classes have different distribution from those used during the training. Furthermore, various metrics, such as the precision, recall, and F-score, are used for more elaboration.

6.1. Experiment Dataset

The quality of the training datasets has a substantial impact on how well ML approaches function [174]. One of the key problems impeding the development of detection systems is the lack of a benchmark dataset for the detection of privacy attacks. We can find many datasets to investigate various ML algorithms in a variety of fields, such as language translation or the biomedical business. However, the paucity of attack detection datasets is mostly due to privacy and security concerns. Additionally, the majority of publicly available datasets are dated, painstakingly anonymized, and do not reflect current network risks. We utilize the MalMemAnalysis dataset [175] to address all of these issues and verify the effectiveness of the suggested ML models by emulating real-world obfuscated malware as closely as is feasible. In order to prevent the memory dump process from being visible in the memory dumps, this dataset uses the memory dump method in debug mode [175].

Malware that has been obfuscated hides itself to avoid being found and removed. The goal of the dataset for obfuscated malware is to evaluate in-memory obfuscated malware detection techniques. The dataset was created to be as realistic as possible. By using malware that is ubiquitous in the real world, this was made possible. An obfuscated malware detection system may be tested using this balanced dataset comprising Trojan

horse, ransomware, and spyware malware. The dataset mimics actual world conditions. Given that it is split equally between malicious and benign memory dumps, the dataset is balanced as shown in Figure 9. It has 58,596 records in total, 29,298 of which are benign and 29,298 of which are malicious as indicated in Table 8.

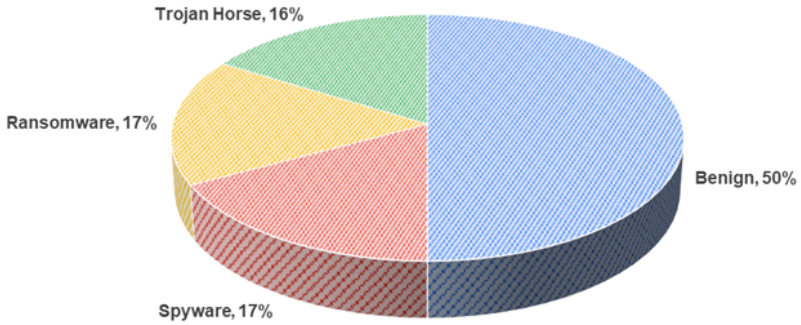

Figure 9. Percentage of different malware families in used dataset.

Table 8. Dataset label distribution.

Categories	Records
Benign	29,298
Spyware	10,020
Ransomware	9791
Trojan Horse	9487
Total	58,596

The malicious traffic was captured by using malicious memory dumps, where 2916 malware samples were collected from VirusTotal. The collected samples have different malware categories, including ransomware, spyware, and Trojan horse as shown in Figure 10. Similarly, for the creation of benign memory dumps, normal user behavior was captured by using various applications in the virtual machine.

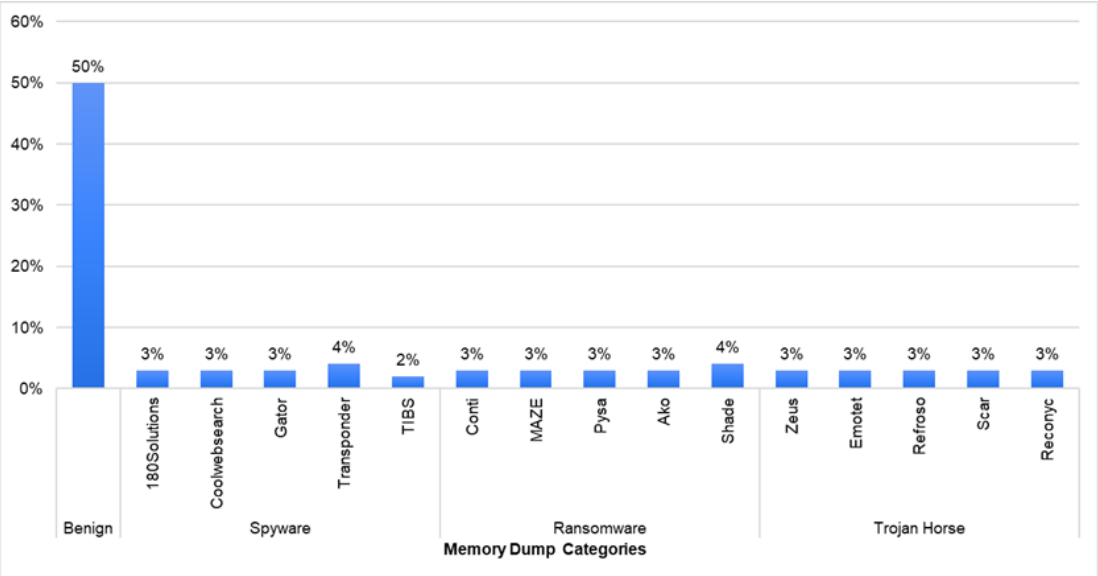

Figure 10. Overall malware families found in the dataset.

6.2. Machine Learning Analysis Techniques

ML derives relevant information from raw data while disguising the information to protect privacy [176]. By learning from their previous performance and adjusting them to provide better outcomes, it makes machines smarter [177,178]. Several ML techniques have shown to be incredibly useful in minimizing privacy threats. These approaches are used to generate meaningful outputs from large and mixed datasets, and the outputs may be used to foresee and detect vulnerabilities in IoT-based models. In the next section, we will perform a practical simulation using several ML algorithms to prove their capabilities to detect related malicious and anomalous attacks against privacy. Eight common supervised learning algorithms were used to train and evaluate the obfuscated malware dataset. Specifically, we employed three tree-based algorithms: logistic regression (LR) [179], gradient boosting (GB) [180], a single decision tree (DT) [181], random forest (RF) [182], Gaussian naive_bayes (GNB) [183], and AdaBoost [184] learner. Additionally, we used the k-nearest neighbor classifier (KNN) [185] and support vector machines (SVM) [186] based methods. The default parameters were used in all the implemented algorithms.

Python was used to plan and carry out the experiment, while the backend libraries Sklearn and Tensorflow were utilized for all suggested methods. The experimental setup used to assess the model parameters is described in Figure 11. In addition, the dataset includes 26 additional characteristics that were retrieved using VolMemLyzer-V2 as part of the suggested model to find hidden and obfuscated malware.

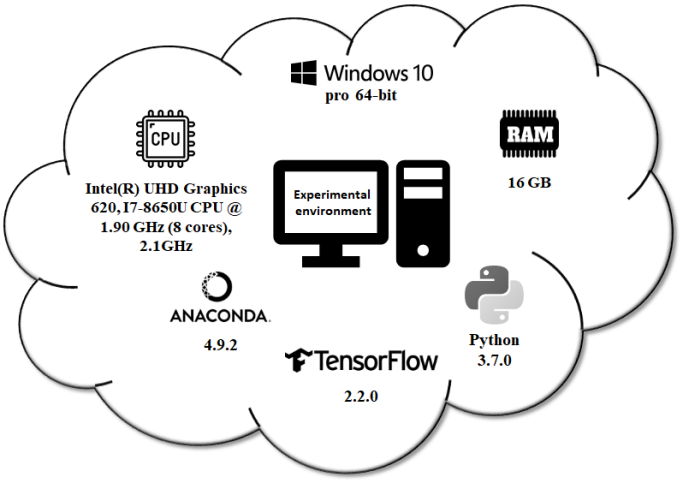

Figure 11. Experimental environment.

6.3. The Evaluation Metrics

To assess the performance of each model, we employed the most used performance metrics, such as accuracy, precision, recall, and F-score metrics as shown in the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN1} \tag{3}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

where true positive *(TP)* and true negative *(TN)* represent the correctly predicted values, and false positive *(FP)* and false negative *(FN)* indicate misclassified events.

7. Experimental Results and Analysis

First, each base learner is evaluated using the dataset, and the results are analyzed using different evaluation metrics, including precision, recall, F1-score and accuracy. In this work, we tested several experiments to tackle the privacy issue using ML algorithms as represented in the following sections.

7.1. Scenario 1

In this experiment, we divided the dataset into 70% (including normal and malware data) for training and 30% testing purpose. The train_test_split method was used from Scikit-Learn library with test_size = 0.3 for this purpose as shown in Table 9. We only considered the binary classification for the experiments, as all Malware classes are considered malicious traffic. The obtained results are represented in Table 10 and Figure 12.

The results show that all ML algorithms provide high evaluation metrics for both normal and malicious malware attacks Figure 13. We can observe that the average accuracy is 99.88%, while the average precision, recall and F1-score for benign are 99.90%, 99.85%, and 99.88% and those for attack are 99.86%, 99.91%, and 99.88% respectively.

Table 9. Used data for Binary Experiment.

	Train		**Test**	
Df	**Benign**	**Malware**	**Benign**	**Malware**
Value_Count	20,548	20,469	8788	8790
Total	41,017		17,578	

Table 10. Individual classifiers' result (binary class).

		Evaluation Results %						
		Precision		**Recall**		**F1-Score**		**Accuracy Score**
Binary Class		**Benign**	**Attack**	**Benign**	**Attack**	**Benign**	**Attack**	
Techniques	LR	0.9985	0.9992	0.9992	0.998528	0.9988	0.99886	0.9988
	AB	1	1	1	1	1	1	1
	GB	0.9995	1	1	0.9995	0.9997	0.9997	0.9997
	GNB	0.9952	0.9897	0.9896	0.9953	0.9924	0.9925	0.9924
	KNN	0.9995	0.9997	0.9997	0.9995	0.9996	0.9996	0.9996
	DT	0.9998	1	1	0.9998	0.9999	0.9999	0.9999
	RF	0.9998	1	1	0.9998	0.9999	0.9999	0.9998
	SVM	1	0.9998	0.9998	1	0.9999	0.9999	0.9999

7.2. Scenario 2

In this experiment, we used Ransomware, and Trojan Horse classes for training, while the Spyware is used for testing. The main objective is to show how ML can work sufficiently in unknown attacks. The Distribution of the used samples for training and testing is depicted in Table 11, while the output results is represented in Table 12, Figures 14 and 15.

The experimental results show that the ML techniques have the capability to detect obfuscated and hidden malware (spyware) with high accuracy, reaching 99.61%, and average precision, recall and F1-score for benign being 99.34%, 99.85%, and 99.59% and those for attack being 99.84%, 99.34%, and 99.58% respectively.

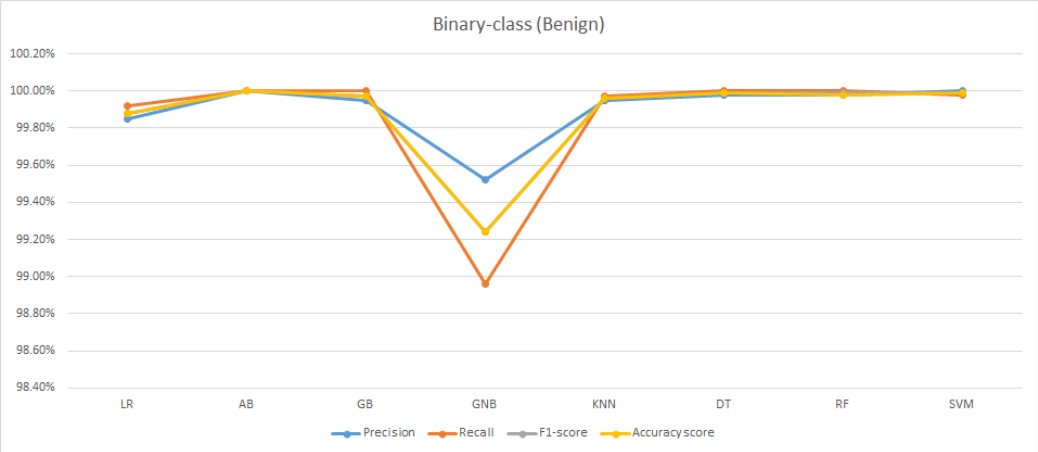

Figure 12. The average results of binary classification (benign).

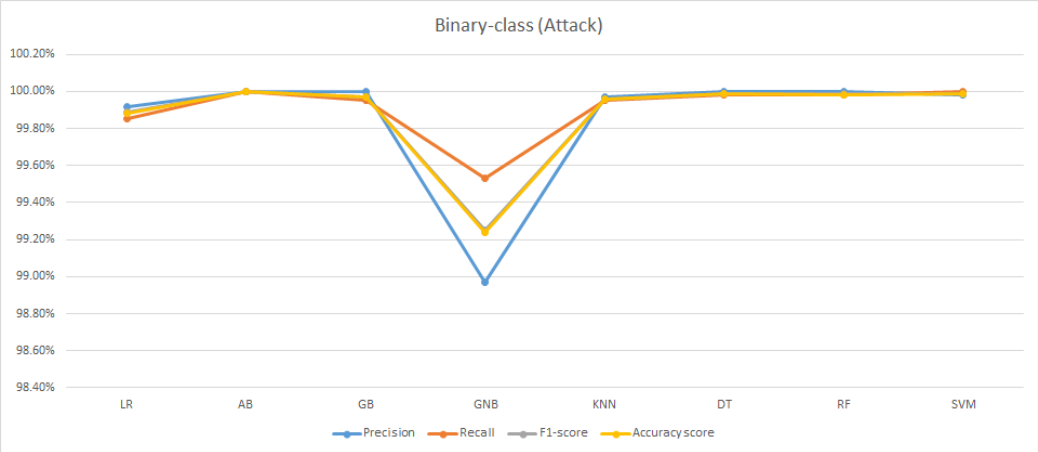

Figure 13. The average results of binary classification (attack).

Table 11. Used data for unknown attack experiment.

	Train		Test
Df	**Ransomware**	**Trojan Horse**	**Spyware**
Value_Count	9791	9487	10,020

Table 12. Individual classifiers result (unknown class).

		Evaluation Results %						
		Precision		Recall		F1-Score		Accuracy Score
Binary-Class		**Benign**	**Attack**	**Benign**	**Attack**	**Benign**	**Attack**	
Techniques	LR	0.9942	0.9992	0.9993	0.9942	0.9967	0.9967	0.9988
	AB	0.994	1	1	0.994	0.997	0.9969	0.9967
	GB	0.994	0.9997	0.9998	0.994	0.9969	0.9968	0.997
	GNB	0.9877	0.9897	0.9899	0.9875	0.9888	0.9886	0.9969
	KNN	0.9943	0.9996	0.9997	0.9942	0.997	0.9969	0.9887
	DT	0.994	0.9996	0.9997	0.9939	0.9968	0.9967	0.9969
	RF	0.9941	0.9996	0.9997	0.994	0.9969	0.9968	0.9968
	SVM	0.9951	0.9995	0.9996	0.995	0.9973	0.9972	0.9968

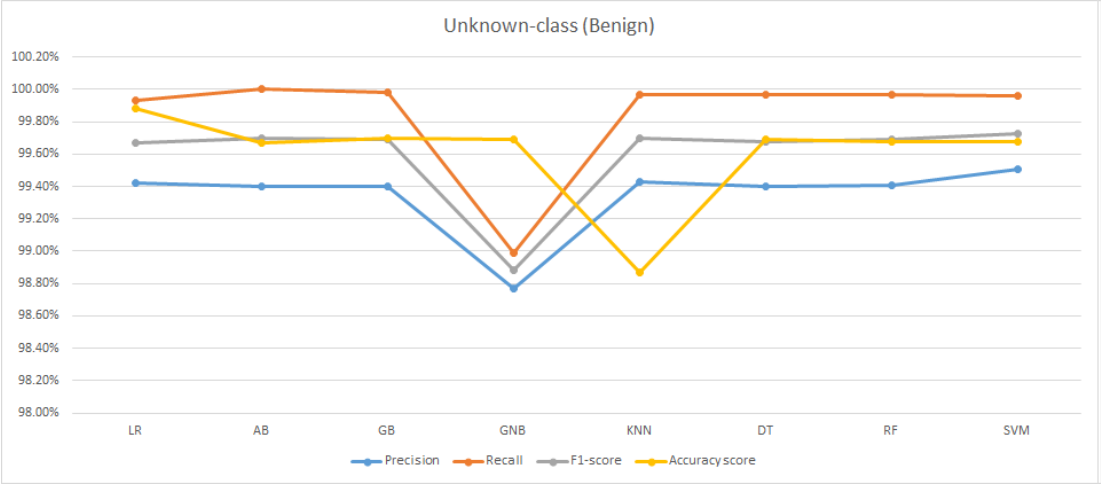

Figure 14. The average results of unknown classification (benign).

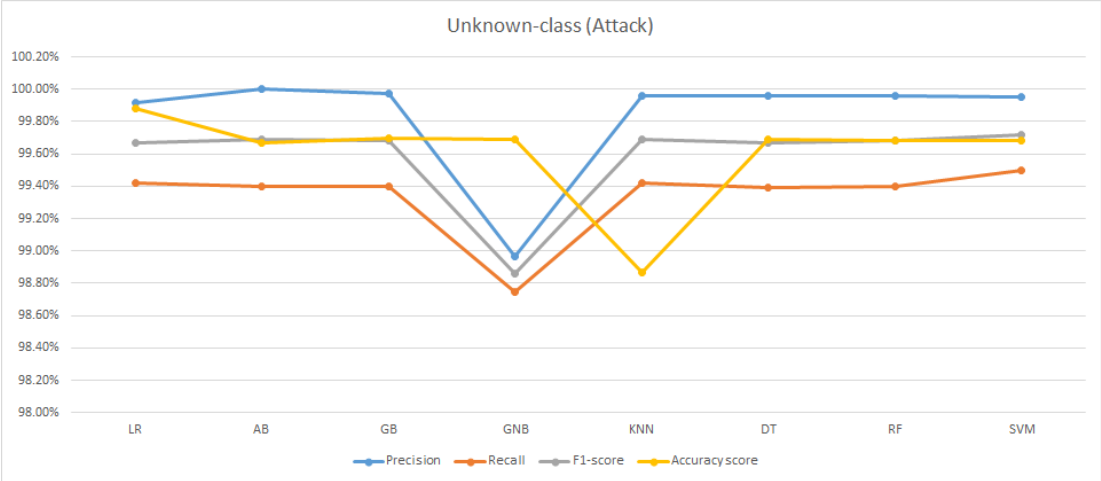

Figure 15. The average results of unknown classification (attack).

7.3. Comparative Analysis with State-of-the-Art

A comparative analysis of possible conventional ML algorithms is performed in this section. It is conducted for comparing and analyzing the accuracy of all the conventional algorithms.

The proposed methodology is compared with the state-of-the-art methods. It is illustrated in Table 13. The results show that the proposed approach achieves better results by 99.50% as compared with the state-of-the-art methods.

Table 13. Comparative analysis using existing ML classification methods-based IoT privacy solutions.

No.	Authors	Classification Methods	Accuracy
1	[70]	Naive Bayes	98%
2	[72]	KNN	98.20%
3	[73]	deep Eigenspace learning	99.68%
4	[187]	SVM	94%
5	[186]	SVM	98.50%

Table 13. *Cont.*

No.	Authors	Classification Methods	Accuracy
6	[188]	CNN	91.34%
7	[69]	SVM	99%
8	[71]	f-score Naïve Bayes Logistic Regression	97% 51% 94%
9	This Study	Logistic Regression AdaBoost	99.88% 100%

7.4. Discussion

With the continuous increase in the number of IoT devices linked to the internet and the quick rate at which these gadgets are becoming an intimate part of human everyday life, there is also an increase in privacy threats. As a result, privacy stands out as a significant risk impeding the mainstream adoption of IoT. IoT device vulnerabilities can result in serious security breaches and drastically harm user privacy by revealing personal data. Platforms, apps, and infrastructures must take privacy seriously in order to boost IoT adoption and alleviate user concerns. For resource-constrained IoT devices, training a model locally may be problematic. Second, several iterations are necessary for the learning process to converge, resulting in high communication overhead since certain local classes have patterns that differ greatly from the public data.

On the other hand, the last decade has witnessed a significant increase in the use of machine learning (ML) techniques to satisfy the needs for creating effective IDSs. It is considered one of the most significant solutions to solve the weaknesses of traditional solutions, as it offers the construction of a global model through the learning of user-pushed updates. As a result, numerous businesses, like Google, Microsoft, and Facebook, employ ML extensively across a range of applications, including speech recognition and image processing. The crucial component of ML approaches is the automated extraction of intense characteristics from raw datasets. As a result, they may be used for a variety of cybersecurity activities, including intrusion detection and traffic analysis [177,189]. However, few works have applied ML algorithms for the privacy attack issue in the IoT context. Since the technology is relatively new, the IoT ecosystem takes into consideration the nature of distributed computing systems as follows: it is very versatile and efficient.

8. Conclusions and Future Work

This paper provided a comprehensive survey on the main privacy issues in IoT and examined how the growth of IoT affects each threat. Concerns about privacy can lead to far larger threats. The associated attacks to the privacy threats were presented. In addition, we conducted a study of possible solutions for addressing different privacy problems and threats in IoT. While some of the privacy issues in IoT contexts are lessened by the ways that are provided, performance evaluation and assessment in real-world contexts are clearly lacking. Furthermore, there is a conflict between the requirement to safeguard user privacy and the degree of data access required to deliver improved services. This brings up the question of how to meet the demands of client privacy while retaining the same quality of service.

In order to produce meaningful outputs from enormous and varied datasets, ML algorithms are required. The outcomes may be used to forecast and identify weaknesses in IoT-based models. We conducted an in-depth review of the current literature related to privacy-preserving ML techniques within the IoT ecosystem. By thoroughly examining and discussing various approaches, we shed light on the limitations and potential areas for improvement in IoT privacy. Our work can be regarded as a comprehensive survey that aims to provide valuable insights for researchers, practitioners, and policymakers in

the domain of IoT security and privacy. We also implemented practical experiments to demonstrate the capability of ML to detect malicious and anomalous attacks and preserve IoT privacy. This approach introduces several layers of training that can decrease the impact of separate, delicate training data on output models. The experiment primarily targeted malware that was obfuscated or masked and belonged to one of the three malware types: ransomware, spyware, and Trojan horse malware. The results of the experiments show that the additional features and ML algorithms enhanced the overall accuracy for detecting obfuscated and concealed malware. The experimental results also show that the suggested technique outperforms the state-of-the-art approaches for detecting malicious and anomalous attacks by 99.50%, which effectively helps with the necessary safeguarding to maintain IoT privacy.

In our future work, we intend to work with other types of IoT privacy issues, as well as testing the suggested model in various use cases and investigating how it may interact with the assaults in real time.

Author Contributions: Methodology, S.E.-G., M.S.E. and M.A.A.; Software, S.E.-G. and A.J.; Validation, S.E.-G., M.S.E. and A.J.; Formal analysis, M.S.E.; Resources, S.E.-G., M.S.E., A.J. and M.A.A.; Writing—original draft, S.E.-G.; Writing—review & editing, M.A.A.; Supervision, M.S.E. and A.J.; Project administration, M.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no close personal links or financial conflicts of interest that may seem to have influenced the research described in this paper.

References

1. Zhu, C.; Leung, V.C.; Shu, L.; Ngai, E.C.H. Green internet of things for smart world. *IEEE Access* **2015**, *3*, 2151–2162. [CrossRef]
2. Shen, M.; Liu, Y.; Zhu, L.; Du, X.; Hu, J. Fine-grained webpage fingerprinting using only packet length information of encrypted traffic. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 2046–2059. [CrossRef]
3. Shen, M.; Zhang, J.; Zhu, L.; Xu, K.; Du, X. Accurate decentralized application identification via encrypted traffic analysis using graph neural networks. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 2367–2380. [CrossRef]
4. Shen, M.; Zhang, J.; Zhu, L.; Xu, K.; Tang, X. Secure SVM training over vertically-partitioned datasets using consortium blockchain for vehicular social networks. *IEEE Trans. Veh. Technol.* **2019**, *69*, 5773–5783. [CrossRef]
5. Kaissis, G.A.; Makowski, M.R.; Rückert, D.; Braren, R.F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2020**, *2*, 305–311. [CrossRef]
6. Singh, U. Role of Data Analytics in Bio Cyber Physical Systems. In *Trends of Data Science and Applications*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 954, pp. 129–146.
7. Kanellos, M. 152,000 Smart Devices Every Minute in 2025: IDC Outlines the Future of Smart Things. Forbes. 2016. Available online: https://www.forbes.com/sites/michaelkanellos/2016/03/03/152000-smart-devices-every-minute-in-2025-idc-outlines-the-future-of-smart-things/?sh=3cc5cdc54b63 (accessed on 8 August 2023).
8. Mahalle, P.; Babar, S.; Prasad, N.R.; Prasad, R. Identity management framework towards internet of things (IoT): Roadmap and key challenges. In Proceedings of the International Conference on Network Security and Applications, Chennai, India, 23–25 July 2010; pp. 430–439.
9. Agarwal, R.; Fernandez, D.G.; Elsaleh, T.; Gyrard, A.; Lanza, J.; Sanchez, L.; Georgantas, N.; Issarny, V. Unified IoT ontology to enable interoperability and federation of testbeds. In Proceedings of the 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA, USA, 12–14 December 2016; pp. 70–75.
10. Ganzha, M.; Paprzycki, M.; Pawłowski, W.; Szmeja, P.; Wasielewska, K. Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective. *J. Netw. Comput. Appl.* **2017**, *81*, 111–124. [CrossRef]
11. Al-Qaseemi, S.A.; Almulhim, H.A.; Almulhim, M.F.; Chaudhry, S.R. IoT architecture challenges and issues: Lack of standardization. In Proceedings of the 2016 Future Technologies Conference (FTC), San Francisco, CA, USA, 6–7 December 2016; pp. 731–738.
12. Ngu, A.H.; Gutierrez, M.; Metsis, V.; Nepal, S.; Sheng, Q.Z. IoT middleware: A survey on issues and enabling technologies. *IEEE Internet Things J.* **2016**, *4*, 1–20. [CrossRef]

13. Chabridon, S.; Laborde, R.; Desprats, T.; Oglaza, A.; Marie, P.; Marquez, S.M. A survey on addressing privacy together with quality of context for context management in the Internet of Things. *Ann. Telecommun.-Ann. Télécommun.* **2014**, *69*, 47–62. [CrossRef]
14. Dwivedi, A.D.; Singh, R.; Ghosh, U.; Mukkamala, R.R.; Tolba, A.; Said, O. Privacy preserving authentication system based on non-interactive zero-knowledge proof suitable for Internet of Things. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *13*, 4639–4649. [CrossRef]
15. Fu, X.; Wang, Y.; Yang, Y.; Postolache, O. Analysis on cascading reliability of edge-assisted Internet of Things. *Reliab. Eng. Syst. Saf.* **2022**, *223*, 108463. [CrossRef]
16. Cucu, P. IoT Security Basics Every Device Owner Needs Now. Available online: https://www.team911.com/news/349442/IoT-Security-Basics-Every-Device-Owner-Needs-Now.htm (accessed on 25 July 2023).
17. Jonsdottir, G.; Wood, D.; Doshi, R. IoT network monitor. In Proceedings of the 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, UK, 3–5 November 2017; pp. 1–5.
18. Lally, G.; Sgandurra, D. Towards a framework for testing the security of IoT devices consistently. In Proceedings of the International Workshop on Emerging Technologies for Authorization and Authentication, Barcelona, Spain, 7 September 2018; pp. 88–102.
19. Cyrus, C. IoT Cyberattacks Escalate in 2021, According to Kaspersky. Available online: https://www.iotworldtoday.com/2021/09/17/iot-cyberattacks-escalate-in-2021-according-to-kaspersky/ (accessed on 25 September 2022).
20. Dovom, E.M.; Azmoodeh, A.; Dehghantanha, A.; Newton, D.E.; Parizi, R.M.; Karimipour, H. Fuzzy pattern tree for edge malware detection and categorization in IoT. *J. Syst. Archit.* **2019**, *97*, 1–7. [CrossRef]
21. Pan, Z.; Sheldon, J.; Mishra, P. Hardware-assisted malware detection using explainable machine learning. In Proceedings of the 2020 IEEE 38th International Conference on Computer Design (ICCD), Hartford, CT, USA, 18–21 October 2020; pp. 663–666.
22. Gibert, D.; Mateu, C.; Planes, J. The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *J. Netw. Comput. Appl.* **2020**, *153*, 102526. [CrossRef]
23. Mahdavinejad, M.S.; Rezvan, M.; Barekatain, M.; Adibi, P.; Barnaghi, P.; Sheth, A.P. Machine learning for Internet of Things data analysis: A survey. *Digit. Commun. Netw.* **2018**, *4*, 161–175. [CrossRef]
24. Chen, M.; Gündüz, D.; Huang, K.; Saad, W.; Bennis, M.; Feljan, A.V.; Poor, H.V. Distributed learning in wireless networks: Recent progress and future challenges. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 3579–3605. [CrossRef]
25. Kumar, J.S.; Patel, D.R. A survey on internet of things: Security and privacy issues. *Int. J. Comput. Appl.* **2014**, *90*, 11.
26. Lin, H.; Bergmann, N.W. IoT privacy and security challenges for smart home environments. *Information* **2016**, *7*, 44. [CrossRef]
27. Borgohain, T.; Kumar, U.; Sanyal, S. Survey of security and privacy issues of internet of things. *arXiv* **2015**, arXiv:1501.02211.
28. Lin, J.; Yu, W.; Zhang, N.; Yang, X.; Zhang, H.; Zhao, W. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet Things J.* **2017**, *4*, 1125–1142. [CrossRef]
29. Yang, Y.; Wu, L.; Yin, G.; Li, L.; Zhao, H. A survey on security and privacy issues in Internet-of-Things. *IEEE Internet Things J.* **2017**, *4*, 1250–1258. [CrossRef]
30. Salman, T.; Jain, R. Networking protocols and standards for internet of things. In *Internet of Things and Data Analytics Handbook*; Wiley: Hoboken, NJ, USA, 2017; pp. 215–238. [CrossRef]
31. El-Gendy, S.; Azer, M.A. Security Framework for Internet of Things (IoT). In Proceedings of the 2020 15th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 15–16 December 2020; pp. 1–6.
32. Guan, Z.; Zhang, Y.; Wu, L.; Wu, J.; Li, J.; Ma, Y.; Hu, J. APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT. *J. Netw. Comput. Appl.* **2019**, *125*, 82–92. [CrossRef]
33. Tonyali, S.; Akkaya, K.; Saputro, N.; Uluagac, A.S.; Nojoumian, M. Privacy-preserving protocols for secure and reliable data aggregation in IoT-enabled smart metering systems. *Future Gener. Comput. Syst.* **2018**, *78*, 547–557. [CrossRef]
34. Lee, S.; Chung, T. Data aggregation for wireless sensor networks using self-organizing map. In Proceedings of the International Conference on AI, Simulation, and Planning in High Autonomy Systems, Jeju Island, Republic of Korea, 4–6 October 2004; pp. 508–517.
35. Rooshenas, A.; Rabiee, H.R.; Movaghar, A.; Naderi, M.Y. Reducing the data transmission in wireless sensor networks using the principal component analysis. In Proceedings of the 2010 Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Brisbane, QLD, Australia, 7–10 December 2010; pp. 133–138.
36. Su, D.; Cao, J.; Li, N.; Bertino, E.; Jin, H. Differentially private k-means clustering. In Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy, New Orleans, LA, USA, 9–11 March 2016; pp. 26–37.
37. Dwork, C. Differential Privacy. In *Encyclopedia of Cryptography and Security*; Springer: Berlin/Heidelberg, Germany, 2011.
38. Konečnỳ, J.; McMahan, H.B.; Yu, F.X.; Richtárik, P.; Suresh, A.T.; Bacon, D. Federated learning: Strategies for improving communication efficiency. *arXiv* **2016**, arXiv:1610.05492.
39. Smith, V.; Chiang, C.K.; Sanjabi, M.; Talwalkar, A.S. Federated multi-task learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
40. Dean, J.; Corrado, G.; Monga, R.; Chen, K.; Devin, M.; Mao, M.; Ranzato, M.; Senior, A.; Tucker, P.; Yang, K.; et al. Large scale distributed deep networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1–11.
41. Mnih, V.; Badia, A.P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; Kavukcuoglu, K. Asynchronous methods for deep reinforcement learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1928–1937.

42. Wang, S.; Tuor, T.; Salonidis, T.; Leung, K.K.; Makaya, C.; He, T.; Chan, K. Adaptive federated learning in resource constrained edge computing systems. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1205–1221. [CrossRef]
43. Wang, X.; Han, Y.; Wang, C.; Zhao, Q.; Chen, X.; Chen, M. In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning. *IEEE Netw.* **2019**, *33*, 156–165. [CrossRef]
44. Borthakur, D.; Dubey, H.; Constant, N.; Mahler, L.; Mankodiya, K. Smart fog: Fog computing framework for unsupervised clustering analytics in wearable internet of things. In Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, Canada, 14–16 November 2017; pp. 472–476.
45. Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 1175–1191.
46. Xu, C.; Ren, J.; Zhang, D.; Zhang, Y. Distilling at the edge: A local differential privacy obfuscation framework for IoT data analytics. *IEEE Commun. Mag.* **2018**, *56*, 20–25. [CrossRef]
47. Mohassel, P.; Zhang, Y. Secureml: A system for scalable privacy-preserving machine learning. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 26 2017; pp. 19–38.
48. Tanuwidjaja, H.C.; Choi, R.; Baek, S.; Kim, K. Privacy-preserving deep learning on machine learning as a service—A comprehensive survey. *IEEE Access* **2020**, *8*, 167425–167447.
49. Beye, M.; Erkin, Z.; Lagendijk, R.L. Efficient privacy preserving k-means clustering in a three-party setting. In Proceedings of the 2011 IEEE International Workshop on Information Forensics and Security, Iguacu Falls, Brazil, 29 November–2 December 2011; pp. 1–6.
50. Rösner, C.; Schmidt, M. Privacy preserving clustering with constraints. *arXiv* **2018**, arXiv:1802.02497.
51. Gascón, A.; Schoppmann, P.; Balle, B.; Raykova, M.; Doerner, J.; Zahur, S.; Evans, D. Privacy-Preserving Distributed Linear Regression on High-Dimensional Data. *Proc. Priv. Enhancing Technol.* **2017**, *2017*, 345–364. [CrossRef]
52. Cock, M.d.; Dowsley, R.; Nascimento, A.C.; Newman, S.C. Fast, privacy preserving linear regression over distributed datasets based on pre-distributed data. In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, Denver, CO, USA, 16 October 2015; pp. 3–14.
53. Ravi, A.; Chitra, S. privacy preserving data mining using differential evolution—Artificial bee colony algorithm. *Int. J. Appl. Eng. Res.* **2014**, *9*, 21575–21584.
54. Fong, P.K.; Weber-Jahnke, J.H. Privacy preserving decision tree learning using unrealized data sets. *IEEE Trans. Knowl. Data Eng.* **2010**, *24*, 353–364. [CrossRef]
55. Yu, H.; Vaidya, J.; Jiang, X. Privacy-preserving svm classification on vertically partitioned data. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, 9–12 April 2006; pp. 647–656.
56. Vaidya, J.; Yu, H.; Jiang, X. Privacy-preserving SVM classification. *Knowl. Inf. Syst.* **2008**, *14*, 161–178. [CrossRef]
57. Aono, Y.; Hayashi, T.; Phong, L.T.; Wang, L. Privacy-preserving logistic regression with distributed data sources via homomorphic encryption. *IEICE Trans. Inf. Syst.* **2016**, *99*, 2079–2089. [CrossRef]
58. Xie, W.; Wang, Y.; Boker, S.M.; Brown, D.E. Privlogit: Efficient privacy-preserving logistic regression by tailoring numerical optimizers. *arXiv* **2016**, arXiv:1611.01170.
59. Huai, M.; Huang, L.; Yang, W.; Li, L.; Qi, M. Privacy-preserving naive bayes classification. In Proceedings of the International Conference on Knowledge Science, Engineering and Management, Chongqing, China, 28–30 October 2015; pp. 627–638.
60. Li, P.; Li, J.; Huang, Z.; Gao, C.Z.; Chen, W.B.; Chen, K. Privacy-preserving outsourced classification in cloud computing. *Clust. Comput.* **2018**, *21*, 277–286. [CrossRef]
61. Xiao, L.; Li, Y.; Han, G.; Liu, G.; Zhuang, W. PHY-layer spoofing detection with reinforcement learning in wireless networks. *IEEE Trans. Veh. Technol.* **2016**, *65*, 10037–10047. [CrossRef]
62. Outchakoucht, A.; Hamza, E.S.; Leroy, J.P. Dynamic access control policy based on blockchain and machine learning for the internet of things. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 417–424. [CrossRef]
63. Ni, Q.; Lobo, J.; Calo, S.; Rohatgi, P.; Bertino, E. Automating role-based provisioning by learning from examples. In Proceedings of the 14th ACM Symposium on Access Control Models and Technologies, Stresa, Italy, 3–5 June 2009; pp. 75–84.
64. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; pp. 3–18.
65. Rouhani, B.D.; Riazi, M.S.; Koushanfar, F. Deepsecure: Scalable provably-secure deep learning. In Proceedings of the 55th Annual Design Automation Conference, San Francisco, CA, USA, 24–29 June 2018; pp. 1–6.
66. HaddadPajouh, H.; Dehghantanha, A.; Khayami, R.; Choo, K.K.R. A deep recurrent neural network based approach for internet of things malware threat hunting. *Future Gener. Comput. Syst.* **2018**, *85*, 88–96. [CrossRef]
67. Kumar, A.; Lim, T. EDIMA: Early detection of IoT malware network activity using machine learning techniques. In Proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), Limerick, Ireland, 15–18 April 2019.
68. Ye, Y.; Li, T.; Adjeroh, D.; Iyengar, S.S. A survey on malware detection using data mining techniques. *ACM Comput. Surv.* **2017**, *50*, 1–40. [CrossRef]
69. Ham, H.S.; Kim, H.H.; Kim, M.S.; Choi, M.J. Linear SVM-based android malware detection for reliable IoT services. *J. Appl. Math.* **2014**, *2014*, 594501. [CrossRef]

70. Kumar, R.; Zhang, X.; Wang, W.; Khan, R.U.; Kumar, J.; Sharif, A. A multimodal malware detection technique for Android IoT devices using various features. *IEEE Access* **2019**, *7*, 64411–64430. [CrossRef]
71. Markel, Z.; Bilzor, M. Building a machine learning classifier for malware detection. In Proceedings of the 2014 Second Workshop on Anti-Malware Testing Research (WATeR), Canterbury, UK, 23 October 2014; pp. 1–4.
72. Nguyen, T.D.; Marchal, S.; Miettinen, M.; Asokan, N.; Sadeghi, A. DÏoT: A self-learning system for detecting compromised IoT devices. *arXiv* **2018**, arXiv:1804.07474.
73. Azmoodeh, A.; Dehghantanha, A.; Choo, K.K.R. Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning. *IEEE Trans. Sustain. Comput.* **2018**, *4*, 88–95. [CrossRef]
74. Nguyen, K.D.T.; Tuan, T.M.; Le, S.H.; Viet, A.P.; Ogawa, M.; Le Minh, N. Comparison of three deep learning-based approaches for IoT malware detection. In Proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh, Vietnam, 1–3 November2018; pp. 382–388.
75. Xiao, L.; Wan, X.; Lu, X.; Zhang, Y.; Wu, D. IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? *IEEE Signal Process. Mag.* **2018**, *35*, 41–49. [CrossRef]
76. Abusnaina, A.; Khormali, A.; Alasmary, H.; Park, J.; Anwar, A.; Mohaisen, A. Adversarial learning attacks on graph-based IoT malware detection systems. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–10 July 2019; pp. 1296–1305.
77. Ertin, E. Gaussian process models for censored sensor readings. In Proceedings of the 2007 IEEE/SP 14th Workshop on Statistical Signal Processing, Madison, WI, USA, 26–29 August 2007; pp. 665–669.
78. Kho, J.; Rogers, A.; Jennings, N.R. Decentralized control of adaptive sampling in wireless sensor networks. *ACM Trans. Sens. Networks (TOSN)* **2009**, *5*, 1–35.
79. Kohonen, T. Essentials of the self-organizing map. *Neural Netw.* **2013**, *37*, 52–65. [CrossRef]
80. Masiero, R.; Quer, G.; Munaretto, D.; Rossi, M.; Widmer, J.; Zorzi, M. Data acquisition through joint compressive sensing and principal component analysis. In Proceedings of the GLOBECOM 2009-2009 IEEE Global Telecommunications Conference, Honolulu, HI, USA, 30 November–4 December 2009; pp. 1–6.
81. Masiero, R.; Quer, G.; Rossi, M.; Zorzi, M. A Bayesian analysis of compressive sensing data recovery in wireless sensor networks. In Proceedings of the 2009 International Conference on Ultra Modern Telecommunications & Workshops, St. Petersburg, Russia, 12–14 October 2009; pp. 1–6.
82. Macua, S.V.; Belanovic, P.; Zazo, S. Consensus-based distributed principal component analysis in wireless sensor networks. In Proceedings of the 2010 IEEE 11th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Marrakech, Morocco, 20–23 June 2010; pp. 1–5.
83. Mihaylov, M.; Tuyls, K.; Nowé, A. Decentralized learning in wireless sensor networks. In Proceedings of the International Workshop on Adaptive and Learning Agents, Budapest, Hungary, 12 May 2009; pp. 60–73.
84. Xiong, J.; Ren, J.; Chen, L.; Yao, Z.; Lin, M.; Wu, D.; Niu, B. Enhancing privacy and availability for data clustering in intelligent electrical service of IoT. *IEEE Internet Things J.* **2018**, *6*, 1530–1540. [CrossRef]
85. Guan, Z.; Lv, Z.; Du, X.; Wu, L.; Guizani, M. Achieving data utility-privacy tradeoff in Internet of medical things: A machine learning approach. *Future Gener. Comput. Syst.* **2019**, *98*, 60–68. [CrossRef]
86. Canedo, J.; Skjellum, A. Using machine learning to secure IoT systems. In Proceedings of the 2016 14th Annual Conference on Privacy, Security and Trust (PST), Auckland, New Zealand, 12–14 December 2016; pp. 219–222.
87. Kulkarni, R.V.; Venayagamoorthy, G.K. Neural network based secure media access control protocol for wireless sensor networks. In Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; pp. 1680–1687.
88. Lane, N.D.; Bhattacharya, S.; Georgiev, P.; Forlivesi, C.; Jiao, L.; Qendro, L.; Kawsar, F. Deepx: A software accelerator for low-power deep learning inference on mobile devices. In Proceedings of the 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Vienna, Austria, 11–14 April 2016; pp. 1–12.
89. Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra, V. Federated learning with non-iid data. *arXiv* **2018**, arXiv:1806.00582.
90. Yang, M.; Zhu, T.; Liu, B.; Xiang, Y.; Zhou, W. Machine learning differential privacy with multifunctional aggregation in a fog computing architecture. *IEEE Access* **2018**, *6*, 17119–17129. [CrossRef]
91. Xiao, L.; Wan, X.; Han, Z. PHY-layer authentication with multiple landmarks with reduced overhead. *IEEE Trans. Wirel. Commun.* **2017**, *17*, 1676–1687. [CrossRef]
92. Das, R.; Gadre, A.; Zhang, S.; Kumar, S.; Moura, J.M. A deep learning approach to IoT authentication. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas, MO, USA, 20–24 May 2018; pp. 1–6.
93. Shi, C.; Liu, J.; Liu, H.; Chen, Y. Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Chennai, India, 10–14 July 2017; pp. 1–10.
94. Guntamukkala, N.; Dara, R.; Grewal, G. A machine-learning based approach for measuring the completeness of online privacy policies. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 289–294.
95. Fredrikson, M.; Jha, S.; Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1322–1333.

96. 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.
97. Hitaj, B.; Ateniese, G.; Perez-Cruz, F. Deep models under the GAN: Information leakage from collaborative deep learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 603–618.
98. Kounoudes, A.D.; Kapitsaki, G.M. A mapping of IoT user-centric privacy preserving approaches to the GDPR. *Internet Things* **2020**, *11*, 100179. [CrossRef]
99. Monteiro, R.L. The New Brazilian General Data Protection Law—A Detailed Analysis. Available online: https://iapp.org/news/a/the-new-brazilian-general-data-protection-law-a-detailed-analysis/ (accessed on 25 September 2022).
100. Wolford, B. What Is GDPR, the EU's New Data Protection Law? Available online: https://gdpr.eu/what-is-gdpr/#:~:text=The%20General%20Data%20Protection%20Regulation,to%20people%20in%20the%20EU (accessed on 25 September 2022).
101. Privacy Flag Project Presents New Tools and a Privacy Certification Scheme at IoT Week 2017. Available online: https://digital-strategy.ec.europa.eu/en/news/privacy-flag-project-presents-new-tools-and-privacy-certification-scheme-iot-week-2017 (accessed on 25 September 2022).
102. Drev, M.; Delak, B. Conceptual Model of Privacy by Design. *J. Comput. Inf. Syst.* **2021**, *62*, 888–895. [CrossRef]
103. Veale, M.; Binns, R.; Edwards, L. Algorithms that remember: Model inversion attacks and data protection law. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 20180083. [CrossRef]
104. Kizza, J.M.; Kizza, W. Wheeler. In *Guide to Computer Network Security*; Springer: Berlin/Heidelberg, Germany, 2013.
105. Bertino, E.; Martino, L.D.; Paci, F.; Squicciarini, A.C. Web services threats, vulnerabilities, and countermeasures. In *Security for Web Services and Service-Oriented Architectures*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 25–44.
106. OWASP. OWASP Top Ten Vulnerabilities 2018 Project. Available online: https://owasp.org/www-pdf-archive/OWASP-IoT-Top-10-2018-final.pdf (accessed on 25 September 2022).
107. Miessler, D. Securing the internet of things: Mapping attack surface areas using the OWASP IoT top 10. In Proceedings of the RSA Conference, San Francisco, CA, USA, 20–24 April 2015; Volume 2015.
108. Ziegeldorf, J.H.; Morchon, O.G.; Wehrle, K. Privacy in the Internet of Things: Threats and challenges. *Secur. Commun. Netw.* **2014**, *7*, 2728–2742. [CrossRef]
109. Strous, L.; von Solms, S.; Zúquete, A. Security and privacy of the Internet of Things. *Comput. Secur.* **2021**, *102*, 102148. [CrossRef]
110. Smith, H.J.; Milberg, S.J.; Burke, S.J. Information privacy: Measuring individuals' concerns about organizational practices. *MIS Q.* **1996**, *20*, 167–196. [CrossRef]
111. Aleisa, N.; Renaud, K. Privacy of the Internet of Things: A Systematic Literature Review (Extended Discussion). *arXiv*, **2017**, arXiv: 1611.03340.
112. Voelcker, J. Stalked by satellite-an alarming rise in GPS-enabled harassment. *IEEE Spectr.* **2006**, *43*, 15–16. [CrossRef]
113. Madaan, N.; Ahad, M.A.; Sastry, S.M. Data integration in IoT ecosystem: Information linkage as a privacy threat. *Comput. Law Secur. Rev.* **2018**, *34*, 125–133. [CrossRef]
114. Ramnath, S.; Javali, A.; Narang, B.; Mishra, P.; Routray, S.K. IoT based localization and tracking. In Proceedings of the 2017 International Conference on IoT and Application (ICIOT), Nagapattinam, India, 19–20 May 2017; pp. 1–4.
115. Caron, X.; Bosua, R.; Maynard, S.B.; Ahmad, A. The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective. *Comput. Law Secur. Rev.* **2016**, *32*, 4–15. [CrossRef]
116. Ogonji, M.M.; Okeyo, G.; Wafula, J.M. A survey on privacy and security of Internet of Things. *Comput. Sci. Rev.* **2020**, *38*, 100312. [CrossRef]
117. Papernot, N.; McDaniel, P.; Goodfellow, I.; Jha, S.; Celik, Z.B.; Swami, A. Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, Abu Dhabi, United Arab Emirates, 2–6 April 2017; pp. 506–519.
118. Kellaris, G.; Kollios, G.; Nissim, K.; O'neill, A. Generic attacks on secure outsourced databases. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 1329–1340.
119. Hayes, J.; Melis, L.; Danezis, G.; De Cristofaro, E. Logan: Membership inference attacks against generative models. In Proceedings of the Privacy Enhancing Technologies (PoPETs), Stockholm, Sweden, 16–20 July 2019; Volume 2019, pp. 133–152.
120. Naveed, M.; Kamara, S.; Wright, C.V. Inference attacks on property-preserving encrypted databases. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, Colorado, 12–16 October 2015; pp. 644–655.
121. Li, N.; Li, T.; Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering, Istanbul, Turkey, 17– 20 April 2007; pp. 106–115.
122. Sagirlar, G.; Carminati, B.; Ferrari, E. Decentralizing privacy enforcement for Internet of Things smart objects. *Comput. Netw.* **2018**, *143*, 112–125. [CrossRef]
123. Datta, T.; Apthorpe, N.; Feamster, N. A developer-friendly library for smart home iot privacy-preserving traffic obfuscation. In Proceedings of the 2018 Workshop on Iot Security and Privacy, Budapest, Hungary, 20 August 2018; pp. 43–48.
124. Narayanan, A.; Huey, J.; Felten, E.W. A precautionary approach to big data privacy. In *Data Protection on the Move*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 357–385.
125. Ohm, P. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Rev.* **2009**, *57*, 1701.

126. Abowd, J.; Alvisi, L.; Dwork, C.; Kannan, S.; Machanavajjhala, A.; Reiter, J. Privacy-Preserving Data Analysis for the Federal Statistical Agencies. *arXiv* **2017**, arXiv:1701.00752.
127. Tramèr, F.; Zhang, F.; Juels, A.; Reiter, M.K.; Ristenpart, T. Stealing Machine Learning Models via Prediction APIs. In Proceedings of the 25th USENIX Security Symposium (USENIX Security 16), Austin, TX, USA, 10–12 August 2016; pp. 601–618.
128. Wang, B.; Gong, N.Z. Stealing hyperparameters in machine learning. In Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 21–23 May 2018; pp. 36–52.
129. Juuti, M.; Szyller, S.; Marchal, S.; Asokan, N. PRADA: Protecting against DNN model stealing attacks. In Proceedings of the 2019 IEEE European Symposium on Security and Privacy (EuroS&P), Stockholm, Sweden, 20–22 May 2019; pp. 512–527.
130. Milli, S.; Schmidt, L.; Dragan, A.D.; Hardt, M. Model reconstruction from model explanations. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; pp. 1–9.
131. Carlini, N.; Liu, C.; Kos, J.; Erlingsson, Ú.; Song, D. The secret sharer: Measuring unintended neural network memorization & extracting secrets. *arXiv* **2018**, arXiv:1802.08232.
132. Andrea, I.; Chrysostomou, C.; Hadjichristofi, G. Internet of Things: Security vulnerabilities and challenges. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Washington, DC, USA, 6–9 July 2015; pp. 180–187.
133. Roman, R.; Zhou, J.; Lopez, J. On the features and challenges of security and privacy in distributed internet of things. *Comput. Netw.* **2013**, *57*, 2266–2279. [CrossRef]
134. Han, G.; Xiao, L.; Poor, H.V. Two-dimensional anti-jamming communication based on deep reinforcement learning. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 2087–2091.
135. Xiao, L.; Li, Y.; Huang, X.; Du, X. Cloud-based malware detection game for mobile devices with offloading. *IEEE Trans. Mob. Comput.* **2017**, *16*, 2742–2750. [CrossRef]
136. Halderman, J.A.; Waters, B.; Felten, E.W. A convenient method for securely managing passwords. In Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, 10–14 May 2005; pp. 471–479.
137. Grobler, M.; Gaire, R.; Nepal, S. User, usage and usability: Redefining human centric cyber security. *Front. Big Data* **2021**, *4*, 583723. [CrossRef]
138. Bonneau, J.; Preibusch, S. The Password Thicket: Technical and Market Failures in Human Authentication on the Web. In Proceedings of the WEIS, Cambridge, MA, USA, 14–15 June 2010.
139. Stobert, E.; Biddle, R. The password life cycle: User behaviour in managing passwords. In Proceedings of the 10th Symposium on Usable Privacy and Security (SOUPS 2014), Santa Clara Valley, CA, USA, 9–11 July 2014; pp. 243–255.
140. Allen, M. Privacy and Security in the Internet of Things Era: IoTCC Best Practices Guidance. Available online: https://insightaas.com/new-research-privacy-and-security-in-the-internet-of-things-era-iotcc-best-practices-guidance/ (accessed on 23 May 2022).
141. Alhirabi, N.; Rana, O.; Perera, C. Security and privacy requirements for the internet of things: A survey. *ACM Trans. Internet Things* **2021**, *2*, 1–37. [CrossRef]
142. Yao, X.; Farha, F.; Li, R.; Psychoula, I.; Chen, L.; Ning, H. Security and privacy issues of physical objects in the IoT: Challenges and opportunities. *Digit. Commun. Netw.* **2021**, *7*, 373–384. [CrossRef]
143. Gao, H.; Zhang, Y.; Miao, H.; Barroso, R.J.D.; Yang, X. SDTIOA: Modeling the timed privacy requirements of IoT service composition: A user interaction perspective for automatic transformation from bpel to timed automata. *Mob. Networks Appl.* **2021**, *26*, 2272–2297. [CrossRef]
144. Fang, W.; Wen, X.Z.; Zheng, Y.; Zhou, M. A survey of big data security and privacy preserving. *IETE Tech. Rev.* **2017**, *34*, 544–560. [CrossRef]
145. Mivule, K. Utilizing Noise Addition for Data Privacy, an Overview. In Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2012), Bangkok, Thailand, 16–19 July 2012; pp. 65–71.
146. Sharma, M.; Chaudhary, A.; Mathuria, M.; Chaudhary, S. A review study on the privacy preserving data mining techniques and approaches. *Int. J. Comput. Sci. Telecommun.* **2013**, *4*, 42–46.
147. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertainty Fuzziness Knowl.-Based Syst.* **2002**, *10*, 557–570. [CrossRef]
148. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data* **2007**, *1*, 3-es. [CrossRef]
149. Skarmeta, A.F.; Hernandez-Ramos, J.L.; Moreno, M.V. A decentralized approach for security and privacy challenges in the internet of things. In Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, Republic of Korea, 6–8 March 2014; pp. 67–72.
150. Feng, H.; Fu, W. Study of recent development about privacy and security of the internet of things. In Proceedings of the 2010 International Conference on Web Information Systems and Mining, Sanya, China, 23–24 October 2010; Volume 2, pp. 91–95.
151. Bost, R.; Popa, R.A.; Tu, S.; Goldwasser, S. Machine learning classification over encrypted data. *Cryptol. Eprint Arch.* **2014**, https://eprint.iacr.org/2014/331.pdf. [CrossRef]
152. Padron, A.; Vargas, G. Multiparty Homomorphic Encryption. 2021. Available online: https://courses.csail.mit.edu/6.857/2016/files/17.pdf (accessed on 25 September 2022).
153. Zhou, H.; Wornell, G. Efficient homomorphic encryption on integer vectors and its applications. In Proceedings of the 2014 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 9–14 February 2014; pp. 1–9.

154. Bogos, S.; Gaspoz, J.; Vaudenay, S. Cryptanalysis of a homomorphic encryption scheme. *Cryptogr. Commun.* **2018**, *10*, 27–39. [CrossRef]
155. Wahab, O.A.; Rjoub, G.; Bentahar, J.; Cohen, R. Federated against the cold: A trust-based federated learning approach to counter the cold start problem in recommendation systems. *Inf. Sci.* **2022**, *601*, 189–206. [CrossRef]
156. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [CrossRef]
157. Nasr, M.; Shokri, R.; Houmansadr, A. Comprehensive privacy analysis of deep learning. In Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 20–22 May 2019; pp. 1–15.
158. Papernot, N.; Abadi, M.; Erlingsson, U.; Goodfellow, I.; Talwar, K. Semi-supervised knowledge transfer for deep learning from private training data. *arXiv* **2016**, arXiv:1610.05755.
159. Dwork, C.; Roth, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
160. Lecuyer, M.; Atlidakis, V.; Geambasu, R.; Hsu, D.; Jana, S. On the connection between differential privacy and adversarial robustness in machine learning. *Stat* **2018**, *1050*, 9.
161. Ayoade, G.; Karande, V.; Khan, L.; Hamlen, K. Decentralized IoT data management using blockchain and trusted execution environment. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake, UT, USA, 6–9 July 2018; pp. 15–22.
162. Liang, X.; Zhao, J.; Shetty, S.; Li, D. Towards data assurance and resilience in IoT using blockchain. In Proceedings of the MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM), Baltimore, MD, USA, 23–25 October 2017; pp. 261–266.
163. Liang, X.; Shetty, S.; Tosh, D.; Kamhoua, C.; Kwiat, K.; Njilla, L. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In Proceedings of the 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), Madrid, Spain, 14–17 May 2017; pp. 468–477.
164. McGhin, T.; Choo, K.K.R.; Liu, C.Z.; He, D. Blockchain in healthcare applications: Research challenges and opportunities. *J. Netw. Comput. Appl.* **2019**, *135*, 62–75. [CrossRef]
165. Zavalyshyn, I.; Duarte, N.O.; Santos, N. HomePad: A privacy-aware smart hub for home environments. In Proceedings of the 2018 IEEE/ACM Symposium on Edge Computing (SEC), Seattle, WA, USA, 25–27 October 2018; pp. 58–73.
166. Yang, J.; Yessenov, K.; Solar-Lezama, A. A language for automatically enforcing privacy policies. *ACM Sigplan Not.* **2012**, *47*, 85–96. [CrossRef]
167. Fernandes, E.; Paupore, J.; Rahmati, A.; Simionato, D.; Conti, M.; Prakash, A. FlowFence: Practical Data Protection for Emerging IoT Application Frameworks. In Proceedings of the 25th USENIX Security Symposium (USENIX Security 16), Vancouver, BA, Canada, 16–18 August 2016; pp. 531–548.
168. Celik, Z.B.; Fernandes, E.; Pauley, E.; Tan, G.; McDaniel, P. Program analysis of commodity IoT applications for security and privacy: Challenges and opportunities. *ACM Comput. Surv.* **2019**, *52*, 1–30. [CrossRef]
169. Zhao, P.; Jiang, H.; Wang, C.; Huang, H.; Liu, G.; Yang, Y. On the performance of *k*-anonymity against inference attacks with background information. *IEEE Internet Things J.* **2018**, *6*, 808–819. [CrossRef]
170. Gkoulalas-Divanis, A.; Loukides, G.; Sun, J. Publishing data from electronic health records while preserving privacy: A survey of algorithms. *J. Biomed. Inform.* **2014**, *50*, 4–19. [CrossRef] [PubMed]
171. Wang, R.; Zhu, Y.; Chen, T.S.; Chang, C.C. Privacy-preserving algorithms for multiple sensitive attributes satisfying t-closeness. *J. Comput. Sci. Technol.* **2018**, *33*, 1231–1242. [CrossRef]
172. Dwork, C. Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25–29 April 2008; pp. 1–19.
173. Tsai, C.F.; Hsu, Y.F.; Lin, C.Y.; Lin, W.Y. Intrusion detection by machine learning: A review. *Expert Syst. Appl.* **2009**, *36*, 11994–12000. [CrossRef]
174. Elsayed, M.S.; Le-Khac, N.A.; Jurcut, A.D. InSDN: A novel SDN intrusion dataset. *IEEE Access* **2020**, *8*, 165263–165284. [CrossRef]
175. Carrier, T.; Victor, P.; Tekeoglu, A.; Lashkari, A.H. Detecting Obfuscated Malware using Memory Feature Engineering. In Proceedings of the ICISSP, Copenhagen, Denmark, 9–11 February 2022.
176. Gong, M.; Xie, Y.; Pan, K.; Feng, K.; Qin, A.K. A survey on differentially private machine learning. *IEEE Comput. Intell. Mag.* **2020**, *15*, 49–64. [CrossRef]
177. ElSayed, M.S.; Le-Khac, N.A.; Albahar, M.A.; Jurcut, A. A novel hybrid model for intrusion detection systems in SDNs based on CNN and a new regularization technique. *J. Netw. Comput. Appl.* **2021**, *191*, 103160. [CrossRef]
178. Elsayed, M.S.; Le-Khac, N.A.; Dev, S.; Jurcut, A.D. Machine-learning techniques for detecting attacks in SDN. In Proceedings of the 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 19–20 October 2019; pp. 277–281.
179. Sperandei, S. Understanding logistic regression analysis. *Biochem. Med.* **2014**, *24*, 12–18. [CrossRef]
180. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
181. Hua, Y.; Ge, S.; Li, C.; Luo, Z.; Jin, X. Distilling deep neural networks for robust classification with soft decision trees. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018; pp. 1128–1132.

182. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
183. Moraes, R.M.; Machado, L.S. Gaussian naive bayes for online training assessment in virtual reality-based simulators. *Mathw. Soft Comput.* **2009**, *16*, 123–132.
184. Wyner, A.J.; Olson, M.; Bleich, J.; Mease, D. Explaining the success of adaboost and random forests as interpolating classifiers. *J. Mach. Learn. Res.* **2017**, *18*, 1558–1590.
185. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Wang, R. Efficient knn classification with different numbers of nearest neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 1774–1785. [CrossRef] [PubMed]
186. Bamakan, S.M.H.; Wang, H.; Yingjie, T.; Shi, Y. An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization. *Neurocomputing* **2016**, *199*, 90–102. [CrossRef]
187. Zhang, T.; Zhu, Q. Distributed privacy-preserving collaborative intrusion detection systems for VANETs. *IEEE Trans. Signal Inf. Process. Netw.* **2018**, *4*, 148–161. [CrossRef]
188. Zhu, H.; Liu, X.; Lu, R.; Li, H. Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 838–850. [CrossRef] [PubMed]
189. El Sayed, M.S.; Le-Khac, N.A.; Azer, M.A.; Jurcut, A.D. A Flow Based Anomaly Detection Approach with Feature Selection Method Against DDoS Attacks in SDNs. *IEEE Trans. Cogn. Commun. Netw.* **2022**, *8*, 1862–1880. [CrossRef]