future internet

Article

Security Threats and Promising Solutions Arising from the Intersection of AI and IoT: A Study of IoMT and IoET Applications

Hadeel Alrubayyi †, Moudy Sharaf Alshareef †, Zunaira Nadeem †, Ahmed M. Abdelmoniem and Mona Jaber *

School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK; h.s.alrubayyi@qmul.ac.uk (H.A.); m.s.a.alshareef@qmul.ac.uk (M.S.A.); z.nadeem@qmul.ac.uk (Z.N.); ahmed.sayed@qmul.ac.uk (A.M.A.)

* Correspondence: m.jaber@qmul.ac.uk

† These authors contributed equally to this work.

Abstract: The hype of the Internet of Things as an enabler for intelligent applications and related promise for ushering accessibility, efficiency, and quality of service is met with hindering security and data privacy concerns. It follows that such IoT systems, which are empowered by artificial intelligence, need to be investigated with cognisance of security threats and mitigation schemes that are tailored to their specific constraints and requirements. In this work, we present a comprehensive review of security threats in IoT and emerging countermeasures with a particular focus on malware and man-in-the-middle attacks. Next, we elaborate on two use cases: the Internet of Energy Things and the Internet of Medical Things. Innovative artificial intelligence methods for automating energy theft detection and stress levels are first detailed, followed by an examination of contextual security threats and privacy breach concerns. An artificial immune system is employed to mitigate the risk of malware attacks, differential privacy is proposed for data protection, and federated learning is harnessed to reduce data exposure.

Keywords: IoT; ML/AI; security threat; IoMT; energy; artificial immune system; federated learning; differential privacy; hyperconnected intelligent world

check for updates

Citation: Alrubayyi, H.; Alshareef, M.S.; Nadeem, Z.; Abdelmoniem, A.M.; Jaber, M. Security Threats and Promising Solutions Arising from the Intersection of AI and IoT: A Study of IoMT and IoET Applications. *Future Internet* **2024**, *16*, 85. https://doi.org/10.3390/fi16030085

Academic Editor: Xuebin Ren

Received: 29 December 2023
Revised: 24 February 2024
Accepted: 28 February 2024
Published: 29 February 2024

1. Introduction

The *hyperconnected intelligent world* (HIW), which has been described as the technology trend of 2023, was founded on the Internet of Things (IoT), which is the network of connected devices that gather the data we need to build intelligent systems (https://www.forbes.com/sites/bernardmarr/2022/09/26/the-5-biggest-technology-trends-in-2023-everyone-must-get-ready-for-now/, accessed on 28 December 2023). The HIW is shaped by IoT data and empowered by artificial intelligence (AI); it is essential for creating data-driven actionable insights. In this work, we closely examine two of such systems: energy theft detection enabled by the Internet of Energy Things (IoET) and remote stress detection enabled by the Internet of Medical Things (IoMT) [1,2]. IoET-based systems are key in the transition toward clean energy; however, the penetration rate of smart meters (basic IoET devices) is very slow (e.g., only 57% in 2023 in the UK (https://assets.publishing.service.gov.uk/media/646f7ff47dd6e70012a9b3f6/Q1_2023_Smart_Meters_Statistics_Report.pdf, accessed on 28 December 2023)). For this reason, IoET systems have been selected as pivotal use cases within HIW, highlighting their significance in achieving a sustainable future. IoMT-based systems, which have a direct impact on health and human lives, are selected as another critical HIW use case, representing systems with no tolerance for security threats, data leakages, or tampering with life-critical IoMT devices.

The potential of marrying IoT and AI in these two applications has been well established but the uptake of related technology is still slow. While connected IoT devices are the key components of such systems, they also create security risks and introduce vulnerabilities that are targeted by attackers to gain access to these systems. Regarding IoET and IoMT, compromising the data transmitted (e.g., man-in-the-middle—MITM) or stored (e.g., malware attack) in the systems represents a serious threat. For instance, the malicious acquisition of energy consumption data would reveal patterns indicating when residents are not at home, potentially inviting targeted robbery attempts [3]. Similarly, unlawful access to biometric data generated by an IoMT device could be used to identify the individual, track their location, and correlate their stress level to targeted marketing, stalking, or worse.

It follows that improving the IoT systems' capabilities to thwart these security attacks is a challenge and a priority in any HIW-related research. In this work, we address IoT security from two perspectives. The *first* work examines malware attack detection mechanisms and proposes an artificial immune system (AIS) that is suitable for the limited computational power of IoT devices in combatting these types of attacks. Malware attacks often lead to the leakage of stored data in the IoT device, gateway, or server. Servers benefit from high computational power and, therefore, can be programmed with complex malware detection mechanisms that mitigate this risk. IoT devices and gateways on the other hand are often mounted on limited energy, low cost, low memory, and low computational power equipment and, thus, are incapable of hosting elaborate malware detection algorithms. For example, deep learning and conventional machine learning techniques are effective at detecting the behaviour of unseen malware attacks [4]. However, these are computationally expensive to implement in IoT devices. AIS is proven to be compatible with IoT device limitations and, hence, is selected in this work to mitigate the data leakage risk at the edge (i.e., IoT devices/gateway) [5]. The *second* work examines an IoMT use case concerned with energy theft detection (ETD) using smart meter data. Prior work [2] proposed a deep learning approach for detecting ETD with promising results, albeit using a centralised server, thus being vulnerable to MITM attacks. This work explores the utilisation of federated learning (FL), which enables efficient learning over decentralised private data without the need to transfer to the central server [6]; thus, the data are less exposed to MITM attacks and related risks of data leakage. In this context, this work presents one of the first FL-based AI methods for energy theft detection based on IoET-enabled energy consumption data. The *third* work investigated differential privacy, an alternative effective countermeasure against MITM attacks, to unlock the potential of IoMT-based tress detection reported in [1]. Differential privacy is a mechanism in which signals carrying private data are contaminated with additive noise to alleviate the risk of data leakage in the event of an attack.

We first present two leading HIW solutions: IoET and IoMT in Section 2, where we demonstrate the unmatched potential of combining forces between IoT and AI to enable intelligent, scalable, and affordable solutions. Next, we present an overall summary of security threats in the IoT Section 2.3, in particular, malware and MITM attacks are discussed. Section 3 exposes the HIW privacy risks under malware attacks and presents AIS, a state-of-the-art IoT-specific malware detection solution. Section 4 elaborates on the HIW privacy risks that might result from MITM attacks on data transferring to the server and presents two solutions. The first is based on an IoET use case and is concerned with ETD using federated learning to avoid sensitive data transfer. The second examines an IoMT use case of remote stress detection and leverages differential privacy for data protection whilst being transferred to the server. In this work, we argue that security threats and adequate measures are not blanket solutions that apply to any IoT system. Instead, a contextual and application-specific solution is needed to address security and data privacy concerns. To this end, in Section 5, we propose an evaluation framework that comprises generic performance metrics as well as application-specific metrics to assess the efficacy of AI for IoT data. In Section 6, we discuss the remaining challenges related to leveraging IoT systems in HIW, particularly related to data privacy concerns, and we present emerging trends to mitigate these risks.

2. Role of AI and IoT in the Hyperconnected Intelligent World

Two leading HIW applications are discussed in this work. The first is presented in Section 2.1, ETD, which falls under the IoET paradigm and uses connected smart meters as the embodiment of IoET devices [2]. In Section 2.2, we present the second application within the ambit of IoMT, which is focused on remote intelligent stress monitoring and uses connected sensors as IoMT devices to measure physiological signals [1].

2.1. Energy Theft Detection

IoET is an emerging archetype that is gaining more attention given the world energy crisis and its relation to climate change. In this context, IoET offers unprecedented leverage over controlling energy consumption (EC) and production, aiming to maximise the energy system efficiency and minimise energy wastage. For instance, IoET is proposed to control energy consumption and reduce energy demand peaks in [7]. The latter is a prime cause for energy wastage as it causes the system to temporarily overproduce outside peak periods. Other IoET applications in the context of smart buildings and smart homes include energy management of smart grids, battery storage systems, electric mobility, and renewable energy sources, as detailed in [8]. An optimal energy management system of smart grids that leverages IoET is proposed in [9]. The system employs private blockchain technology for secure data transactions and accounts for the high uncertainties of renewable energy sources and the charging demands of plug-in hybrid electric vehicles. This work is concerned with energy theft that harms energy providers in terms of revenue loss and system inefficiency. It significantly impacts EC costs and reduces the chances of new powerhouses due to induced fluctuation in demand and supply. The annual revenue loss by energy theft is estimated to be approximately USD 96 billion worldwide (https://www.prnewswire.com/news-releases/96-billion-is-lost-every-year-to-electricity-theft-300453411.html, accessed on 28 December 2023) and is expected to increase. Conventional methods for ETD, including physical onsite inspections to detect and analyse the cause of energy losses, are time-consuming, costly, and labour-intensive. The rise of IoET and the spread of smart meters, together with the advances in AI, render access to and mining of EC data possible at the right time for early ETD. Deep learning methods using convolutional neural networks (CNNs) have been proven to successfully extract latent features in time series EC data for accurate representation and classification [10]. In the context of ETD, theft occurrences are less than 10% of EC data, a major challenge to data-driven CNN methods, which tend to replicate and even reinforce the bias in a skewed dataset. In this case, maximising ETD whilst reducing false alarms (i.e., honest EC samples misclassified as theft) are competing goals that necessitate an energy-aware model for accurate representation. The authors in [2] present a state-of-the-art ETD and demonstrate the superiority of the proposed model compared to previous works, where they reduce the revenue loss due to ETD by 30% by improving the ETD rate and reducing false alarms. However, they propose a centralised learning approach in which granular EC data need to be sent to the server at the risk of overloading the communication network and exposing the data to leakage and privacy breaches.

2.2. Remote Intelligent Stress Monitoring

The spread of IoMT is a key enabler for pervasive and affordable e-healthcare services anywhere and at anytime [11,12]. An intelligent stress monitoring assistant is a leading IoMT application that uses affective computing to detect stress levels based on physiological signals. Detecting stress becomes critically important when three-quarters of people in our society report feeling unable to cope due to stress [13]. To this end, the authors in [1] propose a deep-learning model to detect stress using a self-attention transformer model applied to a public dataset for *wearable stress and affect detection* (WESAD) [14], containing multimodal physiological data collected with medical sensors. The model proposed in [1] successfully detects stress in unseen subjects with 96% accuracy; however, the model is centralised and requires data to be transferred to the central server for training at the risk of MITM attacks. Furthermore, training a model on IoMT-sensitive data, such as the WESAD dataset, involves significant privacy risks as the model might potentially memorise or learn certain characteristics of individual patients. As a result, it is critical to integrate strong security and privacy measures throughout the development and deployment of machine learning models for IoMT-based healthcare applications.

2.3. Security Threats in IoT

IoT networks consist of interconnected devices with unique identifiers that provide real-time interactions; examples include smart meters in IoET and medical sensors in IoMT. These devices often have limited computational capacity and small memory. IoT devices are interconnected, which means they are connected to the cloud and/or other IoT devices for data exchange. Despite the constrained capabilities of connected IoT devices, these can exchange services without the control of a central processor. Such communication is vulnerable to cyber attacks, where a malicious device could be disguised as an accepted IoT device and inject malware files. Moreover, connected devices are heterogeneous as they might run on different platforms and have different specifications. Finally, the increasing number of connected IoT devices generates an enormous scale of data across a large-scale network. The IoT system architecture consists of three main layers: the **perception layer,** which is the physical layer for the connected devices, the **network layer,** which is responsible for securing the connection between the device and the cloud, and the **application layer,** which is the front-end layer to deliver the service. Each layer is a target for different types of attacks. For instance, malicious node injection and battery draining are forms of security threats to the physical layer. MITM and denial of service are threats to the network layer. Malware attacks consist of malicious files (e.g., viruses) and are major threats to the application layer. For more information about the IoT system architecture and corresponding layers, readers are encouraged to refer to [5].

Malware forms a major security threat to IoT systems and is difficult to effectively detect using malware attack detection algorithms. There are three known methods for analysing incoming files and detecting malware: static, dynamic, and hybrid analyses, which use signature and behavioural-based techniques. A signature-based technique reads a unique part of the file for signature detection; thus, it is efficient for detecting known malware files but is unable to detect unknown malware. On the other hand, the behavioural-based technique simulates the behaviour of a malicious or benign file for detection and, hence, is efficient in detecting unknown malware files; however, it is computationally expensive to run. Based on the IoT characteristics and malware detection techniques presented, the best way to secure the IoT is by implementing a lightweight (i.e., not computationally expensive) and adaptive method (i.e., capable of detecting unseen malware).

3. Malware Attack and Effective Detection Method

Malware is targeted malicious software that aims to disrupt a system. If it is undetected and allowed to run, it would cause harm in many ways. In the IoT context, these include disabling the functionality of a device, stealing/deleting/encrypting IoT data, and controlling IoT devices to attack connected systems and organisations. In this section, we discuss the privacy risks that might result from an undetected malware attack and a malware detection mechanism that is suitable and effective in IoT systems. In this context, suitable means that the algorithm does not require a large amount of memory or high computational power to run and, hence, can be hosted by an IoT device. Both AIS and the state-of-the-art malware detection method are implemented using Amazon Web Services (AWS) with different memory sizes. By measuring the processing requirement of each method, we are able to demonstrate that our proposed method is indeed the most suitable for IoT.

3.1. Privacy Risks under Malware Attack

Most IoT devices are purposely designed to be lightweight, which means that they have restricted processing power, storage, and energy. Lightweight IoT devices are affordable and easy to plug and play, thus allowing anywhere–anytime access to data for enabling various HIW solutions. However, being lightweight limits the choice of malware detection schemes for protecting these against malware attacks. In this context, such an attack may lead to unlawful access to potentially sensitive information as in the case of IoET and IoMT.

Malware attacks present a serious risk to smart grids and the IoET paradigm, as revealed in [15]. Malware has been shown to infect the control system of smart grids, which could compromise data privacy and grant unauthorised control over the grid. According to a recent study [16], smart meters are a prime vulnerable target for hackers seeking to disrupt smart grids. Recently, a malware attack targeted the National Health Services (NHS) in the United Kingdom, affecting hospitals and medical staff all over the country [17]. According to the authors in [18], IoT applications are vulnerable points that expose a system to targeted malware attacks, which might enable hackers to obtain access to critical data [19]. Similarly, the authors in [20] investigated how sensitive medical IoMT data collected are exposed to tampering by malware attacks and how the vulnerabilities of IoMT devices could be exploited to deplete their energy or flood the system [21].

3.2. Artificial Immune Systems (AIS) for Detecting Malware Attacks in IoT

AIS is a leading malware detection method that has been proven effective and suitable for IoT systems [5]. It is a digital security method that mimics the human immune system mechanisms in response to antigens. Specifically, AIS replicates how the B and T cells defend the human body at the time of an attack. B cells are activated when an attack (antigen) enters the body, upon which, the B cells mark the antigen by attaching to it and providing antibodies. When an infection occurs, T cells are activated to kill the infected body cells and keep a record of the attack using memory T cells (refer to [5] for more details on B/T cell functionality). There are four main AIS methods used in security applications (such as malicious process detection, anomaly detection, intrusion detection, scan and flood detection, and fraud detection). The *first* is the negative selection method, which is a supervised learning method that replicates the B cell technique in marking an antigen. The *second* is the positive selection method, which replicates the T cell technique in defending the human body. The *third* is the clonal method, which is based on the B cell cloning antibody technique for specific antigens. The *fourth* is AIS, which is an unsupervised classification method based on B memory cells for identifying an attack. AIS methods are adaptive, distributed, and not computationally expensive. Thus, they are a good fit to secure the ever-changing IoT environment with lightweight IoT devices.

The AIS work presented in [5] focuses on an AIS-based method for unknown malware detection in the IoT. The method is based on a negative and positive selection (NPS) technique that works in two stages: a detector generation stage and a detection stage. During the detector generation stage, two different sets of detectors are generated. The first set is a positive detector set containing detectors that match self-data based on the positive selection concept. The other set is a negative detector set, which contains detectors that do not match self-data based on the negative selection concept. The performance of NPS outperforms that of the state-of-the-art by achieving a 21% increase in the malware detection rate while reducing the number of detectors by 65% (directly related to a reduction in memory and computational power requirements). The work in [22] builds on this theoretical study and presents a performance analysis of the NPS method when implemented in realistic IoT systems using Amazon Web Services (AWS). Different IoT system architectures are studied with IoT devices and gateways with differing memory sizes (see Figure 1). This work investigates the feasibility of running the AIS algorithm on individual IoT devices for malware detection. While the approach demonstrates effectiveness on constrained devices, the authors suggest deploying the algorithm on the IoT gateway for improved protection of all connected devices under the same hub. This centralised approach reduces the computational burden on individual devices and potentially offers broader protection.

As shown in Table 1, which uses NSL-KDD [23], NPS outperforms the state-of-the-art MNSA method [24] and effectively detects malware attacks. It should be noted that the MNSA results in Table 1 are produced by implementing the method from [24] in the AWS systems described in Figure 1. To evaluate the proposed approach, four performance metrics are used, i.e., *accuracy,* representing the rate of correctly classified samples (malware and benign), *precision* and *recall*, which measure the rate of correctly detecting malware with respect to all malware or all samples, respectively, and the *F1 score*, which combines the last two metrics. In the case of an unbalanced dataset, such as the NS-KDD, the detection rate is better represented by the F1 score, which best reflects the rate of correctly classified malware and benign files. The results in Table 1 indicate that the four performance metrics improve when the IoT device memory size increases. A larger memory size enables a larger number of detectors to be stored and, therefore, leads to a better chance of detecting malware (refer to [5] for a detailed performance analysis). Moreover, a larger memory size enables a larger training dataset and, thus, yields a better detection rate (see [22]). Table 1 lists the CPU utilisation metrics of both AIS and MNSA methods across four systems, serving as indicators of the computational requirements of each. It is observed that an increased memory size decreases CPU utilisation systematically, which might be due to faster model training convergence when more detectors and a larger dataset are available. More importantly, the results show that the proposed NPS method is indeed lightweight and less computationally expensive than the prior approach, as shown in [5] within a simulation environment and in [22] through an AWS implementation. In [25], the computational cost of malware detection algorithms is estimated based on the required time to train/detect malware using the same NSL-KDD dataset. It is shown that the decision tree approach is the fastest but the proposed semi-supervised double-deep Q-network (SSDDQN) method yields the best results. Given that the running time of an algorithm is tightly related to the processing power of the host, it is not possible to compare the complexities of the AIS methods in Table 1 to the results in [25] without implementing each of these models in the AWS framework.

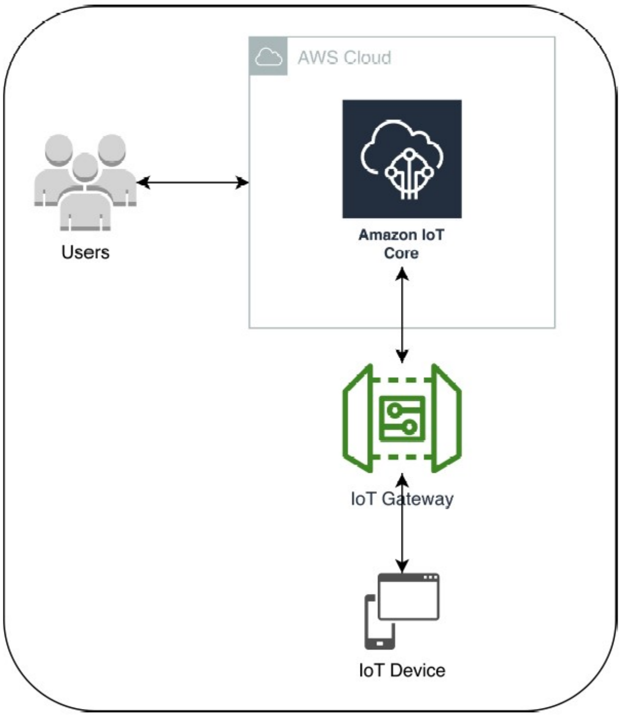


Figure 1. The implemented IoT system architecture using AWS [22].

Table 1. AIS implementation in realistic IoT systems.

AIS Method	IoT Device Memory Size	Accuracy	Precision	Recall	F1 Score	CPU Utilisation
NPS	30 GB	85.13%	91.89%	90.28%	91.08%	55.70%
	32 GB	88.00%	91.52%	94.77%	93.11%	53.20%
	64 GB	92.20%	94.22%	97.35%	95.76%	52.80%
	128 GB	96.80%	97.94%	98.75%	98.34%	52.30%
MNSA [24]	30 GB	68.33%	78.91%	73.17%	75.95%	85.20%
	32 GB	71.00%	80.00%	78.87%	79.43%	83.95%
	64 GB	74.00%	83.10%	80.82%	81.94%	80.35%
	128 GB	76.40%	85.71%	81.52%	83.57%	78.62%

4. Man-in-the-Middle (MITM) Attacks

MITM is a common type of cybersecurity attack that allows malicious users to eavesdrop on the communication between two targets that are both legitimately communicating hosts. In an IoT context where an IoT device is communicating with the server, a malicious user could intercept the communication. Thus, the malicious user can pretend to be the IoT device when communicating with the server or pretend to be the server when communicating with the IoT device. In this case, neither the IoT device nor the server are aware of this interception. Hijacking the communication allows the malicious user to inspect packets (sniffing), which might expose private information in an IoT system. Once the malicious user learns to mimic the communication between these two legitimate entities, they may inject malicious packets blended with valid data communication streams (packet injection). Similarly, the malicious user may sniff the session token and hijack the session (session hijacking). Given the sensitivity of the data transmitted over the IoT (e.g., IoET and IoMT) and the critical role of the IoT device (e.g., controlling a pacemaker or an autonomous vehicle), sniffing, packet injection, and session hijacking can cause significant harm, including malicious remote control and privacy risks.

In [26], MITM attacks that intercept the communication between the central server and IoET edge devices for malicious remote control are studied. These are proven to successfully inject false data and perform data corruption such as sending false commands to IoET edge devices. Similarly, the authors in [27] demonstrate the risk of MITM attacks in an emulated cyber-physical power system in both false data and false command injections.

Similarly, an MITM attack could be caused by hackers that target weak protocols and insert themselves between IoT devices and the server to steal data (sniffing), thus resulting in data leakage. In centralised AI, i.e., a system that requires data to be transferred from the IoT devices to the server, the risk of data leakage due to MITM attacks becomes a threat. This is further aggravated when the data in question carry personal or private information, such as IoET or IoMT applications.

In this section, we discuss two solutions for mitigating these risks. The first employs federating learning, which enables data-driven learning without the need to transmit data from clients to the server [6,28], thus protecting the data from MITM. The FL method is implemented using the Flower framework (https://flower.dev/docs/framework/tutorial-series-use-a-federated-learning-strategy-pytorch.html, accessed on 28 December 2023) using PyTorch [28], and is validated using the open-access dataset, the State Grid Corporation of China (SGCC) (https://github.com/henryRDlab/ElectricityTheftDetection, accessed on 28 December 2023). The second employs differential privacy to protect data before they are sent over the communication channel, therefore, mitigating the risk of exposing personal data under MITM sniffing. This experiment was implemented in *TensorFlow Privacy* and validated using the open-access WESAD dataset. It should be noted that the complexity analysis of both solutions is not discussed here, given that the devices hosting the model training are not constrained. Distribution centres are used in the first and servers in the second, with ample computational power and memory. The computational costs of these solutions will be investigated in future work.

4.1. Federating Learning to Mitigate MITM Attacks: An IoET Case

In response to the significant threat of MITM and the related risk of exposing transmitted data to malicious users, we propose a privacy-preserving federated learning (FL)-based solution with ETD as a use case. FL is an emerging distributed AI paradigm where several clients work together to build a global model federated by a federated server (FS) while keeping their datasets local (see Figure 2). In ETD's case, the EC data of each residential home are processed locally at the *distribution substation*, which acts as a client in an FL system. As such, each *distribution substation* trains an individual AI model using the local data, whilst a global model is federated at the FS. This is conducted by exchanging the model parameters instead of the data. Hence, it mitigates the MITM attacks on EC data and alleviates the burden on the network by avoiding the overload of transmitting extensive data from the edge to the cloud (where the FS is located).

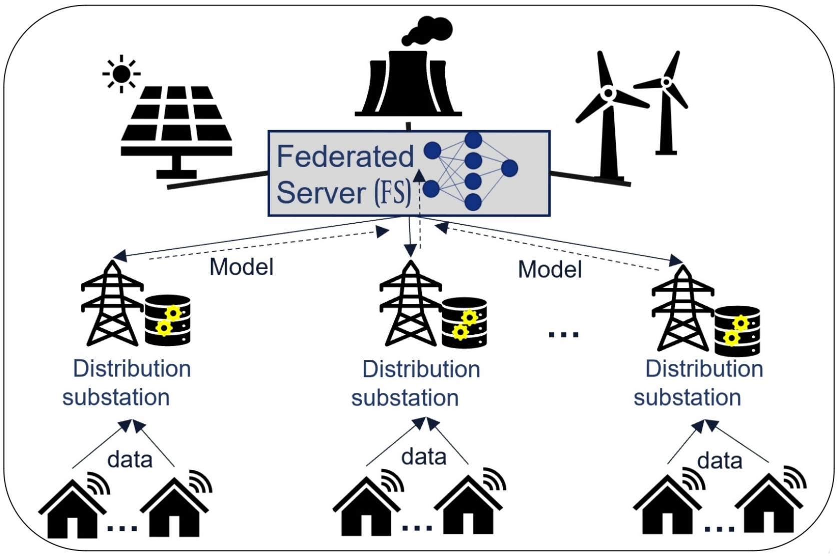


Figure 2. Overview of the system model with smart electricity distribution.

An FL-based ETD method (FedDetect) is presented in [29] with encouraging accuracy results; however, when examined under the energy-aware framework in [2], it was revealed that the effective ETD rate is unacceptable and that the reported results may be misleading due to the highly imbalanced data (see Table 2). In contrast, the proposed FL ETD method (FL-CNN) is energy-aware; it effectively reduced the revenue loss attributed to energy theft by 41% for ten clients (see Table 2).

Table 2. The results of the proposed FL-CNN method in comparison with the state-of-the-art FL result [29] and centralised methods showing both conventional and ETD-specific metrics (%) for different numbers of clients, *C*. († The FedDetect [29] results are reproduced using the pre-processed data in Table 3 and the authors' public GitHub code).

	Clients	Accuracy	Precision	Recall	F1 Score	TPR	FPR	Cost (K£)
Central	CNN-B [2]	89.60	72.19	89.93	30.20	65.58	17.73	2890
	CNN-B †	91.82	94.19	91.83	63.10	**81.75**	7.23	**1415**
FL-CNN	2	87.75	91.90	87.75	49.29	69.64	10.55	2277
	5	80.53	90.34	80.53	37.11	67.22	18.23	2818
	10	72.47	88.09	72.47	24.39	64.43	27.44	3460
Fed Detect	2	91.98	91.00	91.98	40.18	34.77	3.21	3871
	5	92.58	91.03	92.58	36.07	27.03	1.91	4242
	10	92.55	90.53	92.55	25.39	16.37	1.0	4799

Table 3. FL-CNN parameters used in the comparative analysis [2].

Parameters	Values
Number of epochs E	10
Communication rounds R	10
Number of clients C	2, 5, 10
Learning ratio λ	0.001

Although FL mitigates the data leakage between IoT devices and FS, information may still be inferred by sniffing the model parameters using back-engineering, as discussed in [30]. To this end, differential privacy [28,31] is integrated with FL in [29] to further protect the model parameters and reduce the possibility of back-engineering the actual data.

In this work, we propose FL-CNN, which adopts the federated averaging (FedAvg) methodology of the Flower framework (https://flower.dev/docs/framework/tutorial-series-use-a-federated-learning-strategy-pytorch.html, accessed on 28 December 2023) using PyTorch [28]. The process is adapted wherein *distribution substations* are clients that undergo model training on their local data and transmit the model parameters to the FS, where these parameters are aggregated and averaged to facilitate the update of the global model. Subsequently, the revised global model parameters are sent back to the *distribution substation* for subsequent training iterations (see Figure 2). The inherent design of FedAvg aligns well to mitigate cybersecurity risks, like MITM attacks.

The dataset used in this work is the State Grid Corporation of China (SGCC) (https://github.com/henryRDlab/ElectricityTheftDetection, accessed on 28 December 2023) which comprises real-time EC records of 42,372 residential consumers with 10 times more honest data points than theft (see Table 4 for details).

Table 4. Overview of the SGCC dataset [2].

Attributes	Raw Data	Clean Data
Total Customers	42,372	41,897
Honest Customers	38,757	38,321
Dishonest Consumers	3615	3576
Outliers	475 (39 theft)	

The dataset is first pre-processed using linear interpolation as an imputation method to replace the missing values. Next, the three-sigma rule method is used to identify and remove outliers (see clean data in Table 4). To address the data imbalance issue in the training dataset, we implement a hybrid approach, combining undersampling and oversampling techniques using the synthetic minority oversampling technique (SMOTE); see [2] for more details.

Similar to [2], an energy-aware approach for model training and ETD evaluation is adopted in this work, which estimates the actual monetary loss due to (1) undetected theft and (2) honest users falsely classified as dishonest. In addition, this work proposes an FL-CNN system to mitigate the risk of data leakage in the centralised system in [2]. We assume that C *distribution substations* are deployed, where each carries E epochs of learning during which the adaptive moment estimation (Adam) is used in the training phase with a learning ratio of 0.001 λ. At the end of the training epochs, the model weights are communicated to the FS in what is referred to as a communication round. At each communication round, given r_i for $i = \{1 \ldots R\}$ (where R is the maximum number of rounds), the local model weights θ_m^i of *distribution substation* m are forwarded to the FS, which then calculates the averaged weights (based on FedAvg) θ_m^i and pushes these back to the clients. This process repeats until the convergence of the global model, which is similarly trained using Adam. The parameter settings adopted in this work are listed in Table 3. The mathematical representation of FedAvg is represented in Equation (1):

$$\theta_i^t = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\theta_i^{t-1} - \eta\nabla\mathcal{J}_i(\theta_i^{t-1}) & \text{if } t \neq 0 \\ \theta_i^{t-1} - \eta\nabla\mathcal{J}_i(\theta_i^{t-1}) & \text{if } t = 0 \end{cases} \tag{1}$$

The parameters in the equation are as follows:

- θ_i^t: The parameter vector for the i-th client or model at iteration t.
- n: The total number of clients or models in the learning process.
- η: The learning rate, a positive scalar that determines the step size in the direction of the negative gradient.
- $\nabla J_i(\theta_i^{t-1})$: The gradient of the loss function J_i with respect to the parameter vector θ_i^{t-1} for the i-th client or model at iteration $t-1$.

Centralised CNN-B results that were published in [2] are listed in Table 2 as a baseline in addition to CNN-B+, an improved version of CNN-B in which the loss function is tailored to reduce the revenue cost due to energy theft and the CNN hyperparameters are optimised. Next, the results of FL-CNN are summarised in Table 2 with different numbers of clients, $C = \{2, 5, 10\}$. It should be noted that FL-CNN is not limited to the C values listed in Table 2; these are selected to enable the comparison with the state-of-the-art work [29] that uses the same dataset SGCC.

As expected, compared with centralised methods, the FL-based methods (both [29] and FL-CNN) result in a degradation of the ETD performance as seen in the true positive rate (TPR) and cost in Table 2. TPR is a metric that quantifies the rate of successfully detected theft samples compared to all theft samples. Moreover, Table 2 highlights an important trend regarding the effect of the number of clients (*distribution substations*) on the performance of FL methods, whereby an increase in clients generally decreases the TPR for both FL-CNN and FedDetect. The trends observed in both methods with regard to the false positive rate (FPR), which represents the rate of misclassified honest data points compared to all honest samples, differ. This is expected since the FedDetect loss function is not designed to account for the data imbalance and, with more clients (smaller local training dataset), the model tends to classify most data points as honest. FL-CNN, on the other hand, is trained to prioritise the detection of theft and, with a smaller training dataset, is more prone to misclassify honest data points as theft. The energy-aware framework proposed in [2] provides a realistic metric for comparing different methods as it portrays the actual revenue loss of an energy provider due to both undetected theft and false theft alarms; both require an in-person visit to the premises but the former results in a prolonged period of unbilled energy consumption before being identified.

Furthermore, Table 2 shows the reproduced results based on our implementation of [29] using the GitHub repository (https://github.com/xierongpytorch/FedDetect/blob/main/FedDetect.py, accessed on 28 December 2023) provided by the authors, together with the preprocessed dataset depicted in Table 3. Similar to our findings in [2], we again reveal the alarming energy-aware performance of published models despite the attractive yet misleading reported results. In this context, the accuracy results (reported in [29] and reproduced in Table 2) seem to outperform the FL-CNN and centralised learning methods. However, upon a closer look at the energy-aware performance represented by TPR and FPR, and the related costs, we can see that [29] has the worst performance as it leads to the highest revenue loss due to energy theft. Indeed, FedDetect [29] resulted in a very poor theft detection rate (16.37%–34.77%) and the highest costs, i.e., $K3871\ GBP$–$K4799\ GBP$ for 2–10 clients, resulting in the highest accuracy values, i.e., ∼92% (see Table 2). This is not surprising since, given the 92.55% share of honest users in the dataset (see Table 4), if all users are classified as honest, then the classification accuracy would be 92.55%. Our proposed FL-CNN method achieves the best results whilst contributing to data privacy for the three common implementations—2, 5, and 10 clients with cost savings of 41%, 33%, and 72% for each scenario, respectively.

4.2. Differential Privacy to Mitigate MITM Attacks: An IoMT Case

IoMT-based remote stress detection is seen as a promising solution for increasing the reach of mental healthcare to vulnerable people with no access to medical staff. Such solutions primarily employ IoMT sensors that measure physiological signals, such as those listed in Table 5. A deep learning method is proposed in [1], where an attention-based transformer network is shown to successfully detect stress in unseen people. Such an implementation requires all data to be transmitted to a central server for model training, thus exposing private information to the MITM attack [32,33]. The same authors employ differential privacy in [34] to protect the IoMT and present a study on the trade-off between privacy and performance.

Table 5. Overview of data streams and their relation to stress.

Data Stream	Description	Relation to Stress
Electrocardiogram (ECG)	Measures heart activity	Indicates stress via heart rate variability
Electrodermal activity (EDA)	Measures skin conductance	Reflects emotional states
Temperature (Temp)	Measures skin temperature	Altered by stress-induced variations
Respiration (Resp)	Measures breathing patterns	Altered by stress levels
Accelerometer (ACC)	Captures movement data	Indicates physical restlessness due to stress
Electromyogram (EMG)	Measures muscle activity	Indicates muscular tension from stress

Differential privacy is an established data protection mechanism, where each data signal is contaminated with additive noise before being sent to the central server for processing, as shown in Figure 3. In [34], a noise value ν is sampled from a zero-mean Gaussian distribution with standard deviation $\sigma = [0 \dots 1]$. In differential privacy terminology, this standard deviation is referred to as the noise multiplier, where $\sigma = 0$ indicates that the original signal is not altered and $\sigma = 1$ offers the highest protection with the highest noise level. Based on the setting of the noise multiplier, a privacy budget ε is calculated, estimating the potential risk of privacy leakage as follows in Equation (2) [35]:

$$\varepsilon = \sqrt{2I \log(1/\delta)}\sigma + I(e^{\sigma^2/2} - 1), \tag{2}$$

where I is the number of training iterations, δ is a hyperparameter usually set to a small value, and (σ) is the noise multiplier.

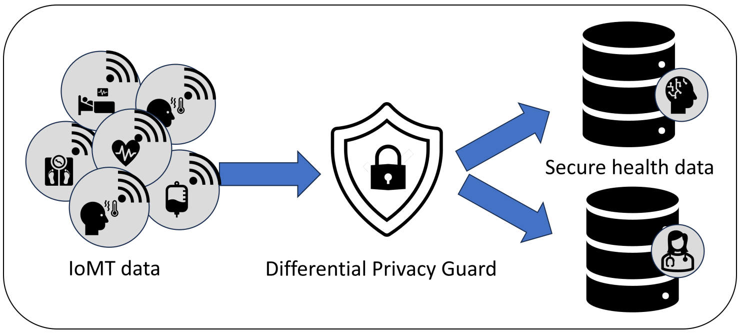
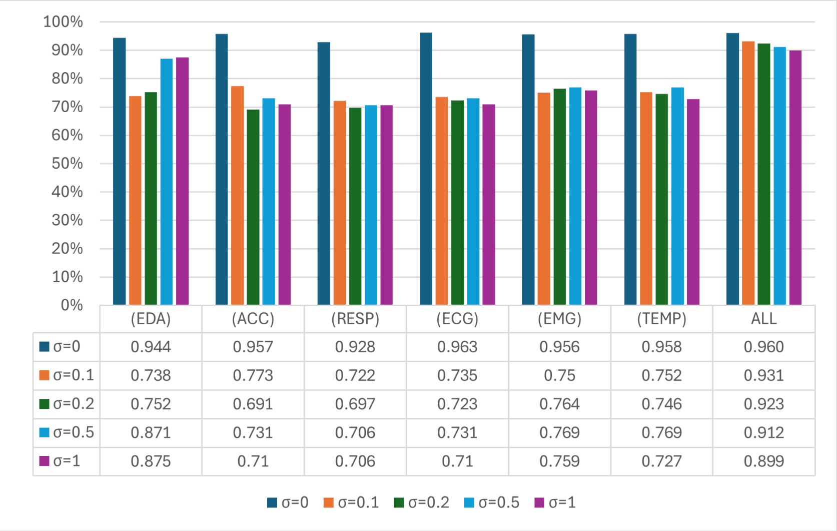


Figure 3. Differential privacy for securing IoMT data privacy.

In Figure 4, we present the results obtained with the proposed method using WESAD, a publicly available dataset. As can be seen, adding differential privacy ($\sigma \neq 0$) to the transformer model for any of the data streams negatively impacts the performance of the model. The drop in accuracy when σ increases from 0 to 0.1 for each data stream is, on average, 20.6%, with a minimum of 18.4% in the case of the accelerometer data stream and a maximum of 22.8% for the electrocardiogram data stream. In contrast, the multimodal model (referred to as ALL in Figure 4) is significantly more robust when differential privacy is implemented, experiencing a drop in accuracy of 2.9% for $\sigma = 0.1$. As expected, the degradation in the stress detection rate increases with higher noise multipliers, but the highest drop of 6.1% is recorded with maximum privacy (i.e., $\sigma = 1$), thus yielding an acceptable accuracy of 89.9% for the multimodal model (ALL).



	(EDA)	(ACC)	(RESP)	(ECG)	(EMG)	(TEMP)	ALL
■ σ=0	0.944	0.957	0.928	0.963	0.956	0.958	0.960
■ σ=0.1	0.738	0.773	0.722	0.735	0.75	0.752	0.931
■ σ=0.2	0.752	0.691	0.697	0.723	0.764	0.746	0.923
■ σ=0.5	0.871	0.731	0.706	0.731	0.769	0.769	0.912
■ σ=1	0.875	0.71	0.706	0.71	0.759	0.727	0.899

Figure 4. Accuracy results of the differential privacy implementation using the WESAD dataset (taken from [34]).

Finding the trade-off between privacy and accuracy is delicate and tightly related to the level of harm that might be caused if there were a breach of privacy. It follows that there is no blanket solution or optimum setting that always works best; instead, the trade-off needs to be adjusted for different scenarios and IoMT systems. As an example, the potential harm that might result from leaking data related to skin temperature is lower than that of data related to ECG. ECG patterns uniquely identify a person and, therefore, ECG data do not pose any security risks. This is confirmed by the analysis of the privacy level and potential harm for each IoMT data stream used in stress detection, as presented in [34]. In other IoMT applications, as systems for remote monitoring and control of diabetes, the detected sugar level is key for controlling insulin levels and, therefore, the detection accuracy cannot be compromised for the sake of privacy [36]. The optimum trade-off for each of the data streams in Table 5 is shown in Figure 5.

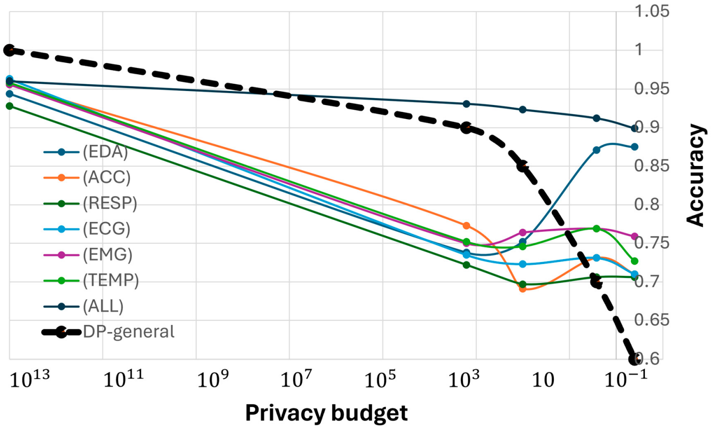


Figure 5. The generic trade-off curve between data privacy and the effectiveness of machine learning in the case of differential privacy is shown in the dashed line. The blue star indicates the ideal but impossible-to-achieve point. Each of the coloured doughnuts indicates the optimum point between privacy and accuracy for each of the data streams in Table 5 with differing privacy requirements.

5. Privacy Evaluation Methods

Two different types of metrics are frequently used to evaluate how well privacy-preserving methods perform: (1) privacy metrics to determine how much privacy a dataset has lost/gained, and (2) utility metrics to determine how useful the protected data are for data analysis. A comprehensive survey of more than 80 privacy metrics is presented in [37], where two categories for privacy evaluation are identified: adversary models and data sources. Methods that employ adversary models evaluate an adversary's skills and intentions (e.g., MITM), whereas those that examine the data sources investigate potential privacy leakage risks (e.g., Malware attacks). Both are key elements in establishing the privacy measure for a given scenario and numerous metrics are often used to accurately analyse the related degree of privacy loss.

Privacy-preserving measures often lead to compromising utility metrics, which are used to quantify the usefulness of data being protected. This is seen in both applications (FL for ETD) in Section 4.1 and differential privacy for stress detection in Section 4.2. Information loss measures are established to quantify the similarity between the original data and protected data for general analysis purposes. Usually, the usefulness of protected data is determined by how well the statistical details of the original data are preserved. By comparing the accuracy of the evaluation carried out on the protected data with that of the original data, the data utility is evaluated to determine its suitability for particular analytical goals, such as machine learning and statistical analysis [38]. Differential privacy is a rigorous mathematical definition quantifying privacy leakage in which noise is added to the original signal to reduce the probability of privacy leakage in case the data are intercepted. As shown in Figure 5, a higher noise level improves data privacy at the cost of reduced utility. It can also be seen that the proposed differential privacy's multimodal approach (ALL) is closest to the ideal point as it delivers high-accuracy with an acceptable privacy budget. As highlighted in [39], differential privacy has been extensively investigated in various AI domains, including machine learning, deep learning, and multi-agent learning. The study highlights the crucial role of differential privacy in surmounting significant AI challenges, including enhancing privacy stability, fairness, and security. It facilitates composition and utility in AI systems while effectively mitigating risks like membership and attribute inference attacks. These aspects not only ensure robust privacy guarantees but also have a profound impact on the performance and reliability of AI models. One of the fundamental aspects of differential privacy is the trade-off between utility and privacy, especially given the challenges in quantifying utility loss. In differential privacy, the ϵ parameter plays a key role in defining the level of privacy. It quantifies the privacy budget, with lower values indicating stronger privacy. Larger ϵ values are chosen to ensure utility, but this often leads to a gap between theoretical privacy guarantees and actual privacy loss observed in practice. Prior studies suggest that while choosing a higher ϵ value to maintain data utility might weaken privacy guarantees, the actual leakage under inference attacks remains low [40,41].

The authors in [38] provided a thorough overview of privacy and utility metrics for privacy-preserving machine learning applications, ranging from broad to targeted data analysis, including several privacy-preserving FL techniques. The authors examined the types of attacks that can be performed against FL systems [6] to assess user data privacy. As evaluation criteria, the attacker's success rate and the amount of information demonstrated in the attacks were used. A membership inference attack (MIA) is a typical privacy metric used to assess the vulnerability of FL models to inference attacks, in which an attacker attempts to identify whether a specific data point was included during model training. In this context, the attack success rate (ASR) is a measure of an attacker's correct predictions of MIA.

In [42], the authors examine privacy leakage under MIA between FL and corset-based learning, as well as accuracy and communication costs. They demonstrate the accuracy–privacy–cost trade-off of each technique using real-world datasets and advanced attack models. In this work, we find a similar trend between model accuracy in a centralised learning framework and data privacy in the FL framework (see Section 4.1).

6. Remaining Challenges

In this section, we discuss some of the key open challenges yet to be addressed.

6.1. Risk of Model Stealing

In this work, to protect the privacy of the users' data, we advocate a distributed ML paradigm and propose a federated learning approach to collaboratively train a global model. This approach entails the exchange of trained local and global models over the network. As a consequence, there is a potential risk for the system to be subjected to model stealing. ML model stealing, also known as model extraction, is a type of attack where an adversary attempts to extract or steal the trained ML model used by another party [43,44]. In IoT scenarios, this is quite concerning because the ML models are often deployed on less-capable devices (i.e., resource-constrained with limited computational power and memory) without proper security measures, making them more vulnerable to attacks. Typically, adversaries can steal the ML models that are deployed on devices or exchanged during training to extract private information, such as user behaviour or personal data.

To prevent model stealing in IoT settings, methods such as model obfuscation, data perturbation, and model watermarking can be used [43,45]. Model obfuscation adds noise or complexity to mitigate model extraction attacks. Another approach leverages data perturbation by adding random noise to the training data, making it harder to infer from the model's parameters. In this work, we propose AIS, which enhances the security measures of distributed ML deployments in IoT scenarios and can be leveraged as a complementary mitigation mechanism against ML model-stealing attacks.

6.2. Assumptions on the Threat Model

In this work, it is assumed that 3rd parties (or external adversaries) are the main risk sources of federated learning systems. However, it is important to recognize the various types of thread models that FL systems are subjected to [46].

For example, it might not be the case that the attackers are foreign agents, as the vulnerabilities in a federated learning setup might emerge from within the group of participating clients that contribute data during the learning phase. Attacks originating from these clients pose a significant risk, potentially exerting greater influence than those during the inference phase. This is because adversarial clients in federated learning possess the ability to not only exploit the model service boundaries in production but also actively manipulate and alter the model's boundaries during its development stage [46].

Guaranteeing resilience against model backdoors is a critical concern. Within the realms of machine learning and computer security, various backdoor attack methods have emerged, targeting neural network models. Thus, questions such as "Can we ensure robustness against these back doors by strategically designing backdoor-proof models?" and "Does the challenge of avoiding backdoors become self-referential, considering that models inherently learn from the dataset's distribution?" arise. It is anticipated that approaches that gauge or maintain resilience [46] or leverage robust knowledge representations against such attacks display great potential in this direction.

However, privacy is of lesser concern in federated learning (FL) because the data are distributed across multiple devices, and the model training process is performed locally on the devices. There are methods proposed to enhance privacy, such as differential privacy [31], homophobic encryption (HE) [47], and secure aggregation (SA) [48].

For instance, similar to data obfuscation, differential privacy adds random noise to the data to mask user data and protect their privacy [31]. This ensures that the model's output does not depend on any single user's data by adding sufficient noise to the training data to protect the users' data, while still maintaining the overall quality of the model. In FL, the model is trained by aggregating the model updates from multiple devices, which can comprise privacy and potentially expose the users' data to the federated server (FS). Some approaches leverage homomorphic encryption, allowing computations to be performed on encrypted data without decrypting it. This is achieved in FL by encrypting the model before sharing it with the FS, and the FS can perform aggregation on encrypted models [47]. Another approach leverages secure aggregation methods based on multi-party computations. In FL, the FS computes the sums of model parameter updates from individual user devices in a secure manner, without revealing the individual user's data [48].

7. Conclusions

In this work, we examine the potential of the Internet of Things (IoT) in enabling the *hyperconnected intelligent world* (HIW), with emphasis on two use cases: the Internet of Energy Things (IoET) and the Internet of Medical Things (IoMT). We also expose the security risks of these systems and the related threat models and data privacy concerns. Two types of cyber security attacks are discussed: malware and man-in-the-middle (MITM) attacks. An artificial immune system is proposed to address the former and is proven to be suitable for IoT systems, given their limited processing power and memory space. For each approach, federated learning and differential privacy are suggested to mitigate the MITM attacks, and the trade-off between the efficacy of data-driven models and data privacy protection is discussed. We conclude that IoT systems are indispensable for advancing HIW applications, and while the proposed measures for combating cyberattacks are promising, a few challenges remain that require further analysis.

Author Contributions: Methodology, M.S.A., Z.N., A.M.A. and M.J.; Investigation, H.A. and M.J.; Writing—original draft, H.A., M.S.A., Z.N., A.M.A. and M.J. All authors have read and agreed to the published version of the manuscript.

Funding: Moudy Sharaf Alshareef PhD studies are funded and sponsored by University of Umm AI-Qura university, Saudi Arabia, and the Saudi Cultural Bureau.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alshareef, M.S.; Alturki, B.; Jaber, M. A transformer-based model for effective and exportable IoMT-based stress detection. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 1158–1163. [CrossRef]
2. Nadeem, Z.; Aslam, Z.; Jaber, M.; Qayyum, A.; Qadir, J. Energy-aware Theft Detection based on IoT Energy Consumption Data. In Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 20–23 June 2023; IEEE: Florence, Italy, 2023; pp. 1–6. [CrossRef]
3. De, S.J.; Métayer, D.L. Privacy Harm Analysis: A Case Study on Smart Grids. In Proceedings of the 2016 IEEE Security and Privacy Workshops (SPW), San Jose, CA, USA, 22–26 May 2016; pp. 58–65. [CrossRef]
4. Tayyab, U.e.H.; Khan, F.B.; Durad, M.H.; Khan, A.; Lee, Y.S. A Survey of the Recent Trends in Deep Learning Based Malware Detection. *J. Cybersecur. Priv.* **2022**, *2*, 800–829. [CrossRef]
5. Alrubayyi, H.; Goteng, G.; Jaber, M.; Kelly, J. Challenges of Malware Detection in the IoT and a Review of Artificial Immune System Approaches. *J. Sens. Actuator Netw.* **2021**, *10*, 61. [CrossRef]
6. Abdelmoniem, A.M.; Sahu, A.N.; Canini, M.; Fahmy, S.A. REFL: Resource-Efficient Federated Learning. In Proceedings of the Eighteenth European Conference on Computer Systems (EuroSys), Rome, Italy, 8–12 May 2023.
7. Muhsen, D.H.; Haider, H.T.; Al-Nidawi, Y.; Shayea, G.G. Operational Scheduling of Household Appliances by Using Triple-Objective Optimization Algorithm Integrated with Multi-Criteria Decision Making. *Sustainability* **2023**, *15*, 16589. [CrossRef]
8. Afonso, J.A.; Monteiro, V.; Afonso, J.L. Internet of Things Systems and Applications for Smart Buildings. *Energies* **2023**, *16*, 2757. [CrossRef]
9. Wang, B.; Ma, H.; Wang, F.; Dampage, U.; Al-Dhaifallah, M.; Ali, Z.M.; Mohamed, M.A. An IoT-Enabled Stochastic Operation Management Framework for Smart Grids. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 1025–1034. [CrossRef]
10. González-Vidal, A.; Mendoza-Bernal, J.; Niu, S.; Skarmeta, A.F.; Song, H. A Transfer Learning Framework for Predictive Energy-Related Scenarios in Smart Buildings. *IEEE Trans. Ind. Appl.* **2023**, *59*, 26–37. [CrossRef]

11. Abdulmalek, S.; Nasir, A.; Jabbar, W.A.; Almuhaya, M.A.M.; Bairagi, A.K.; Khan, M.A.M.; Kee, S.H. IoT-Based Healthcare-Monitoring System towards Improving Quality of Life: A Review. *Healthcare* **2022**, *10*, 1993. [CrossRef]
12. Alshehri, F.; Muhammad, G. A Comprehensive Survey of the Internet of Things (IoT) and AI-Based Smart Healthcare. *IEEE Access* **2021**, *9*, 3660–3678. [CrossRef]
13. Mental Health Foundation. Stress: Are We Coping? Available online: https://www.mentalhealth.org.uk/explore-mental-health/publications/stress-are-we-coping-report (accessed on 28 December 2023).
14. Garg, P.; Santhosh, J.; Dengel, A.; Ishimaru, S. Stress detection by machine learning and wearable sensors. In Proceedings of the 26th International Conference on Intelligent User Interfaces-Companion, College Station, TX, USA, 14–17 April 2021; pp. 43–45.
15. Bouramdane, A.A. Cyberattacks in Smart Grids: Challenges and Solving the Multi-Criteria Decision-Making for Cybersecurity Options, Including Ones That Incorporate Artificial Intelligence, Using an Analytical Hierarchy Process. *J. Cybersecur. Priv.* **2023**, *3*, 662–705. [CrossRef]
16. Alanazi, F.; Kim, J.; Cotilla-Sanchez, E. Load Oscillating Attacks of Smart Grids: Vulnerability Analysis. *IEEE Access* **2023**, *11*, 36538–36549. [CrossRef]
17. Saleem, M. Brexit Impact on Cyber Security of United Kingdom. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–6. [CrossRef]
18. Outdated Software Leaves NHS 'Vulnerable to Cyber Attack'. 2019. Available online: https://www.digitalhealth.net/2019/04/outdated-software-leaves-nhs-vulnerable-to-cyber-attack-new-research-says/ (accessed on 28 December 2023).
19. Hilt, S.; Kropotov, V.; Mercês, F.; Rosario, M.; Sancho, D. The Internet of Things in the cybercrime underground. *Trend Micro Res.* 2019. Available online: https://media.rbcdn.ru/media/reports/wp-the-internet-of-things-in-the-cybercrime-underground.pdf (accessed on 28 December 2023).
20. Thamilarasu, G.; Odesile, A.; Hoang, A. An Intrusion Detection System for Internet of Medical Things. *IEEE Access* **2020**, *8*, 181560–181576. [CrossRef]
21. Hatzivasilis, G.; Soultatos, O.; Ioannidis, S.; Verikoukis, C.; Demetriou, G.; Tsatsoulis, C. Review of Security and Privacy for the Internet of Medical Things (IoMT). In Proceedings of the 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 29–31 May 2019; pp. 457–464. [CrossRef]
22. Alrubayyi, H.; Goteng, G.; Jaber, M. AIS for Malware Detection in a Realistic IoT System: Challenges and Opportunities. *Network* **2023**, *3*, 522–537. [CrossRef]
23. Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 8–10 July 2009; pp. 1–6. [CrossRef]
24. Pamukov, M.E.; Poulkov, V.K.; Shterev, V.A. Negative Selection and Neural Network Based Algorithm for Intrusion Detection in IoT. In Proceedings of the 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 4–6 July 2018; pp. 1–5. [CrossRef]
25. Dong, S.; Xia, Y.; Peng, T. Role of Internet of things in diabetes healthcare: Network infrastructure, taxonomy, challenges, and security model. *IEEE Trans. Netw. Serv. Manag.* **2021**, *18*, 4197–4212. [CrossRef]
26. Sen, Ö.; van der Velde, D.; Linnartz, P.; Hacker, I.; Henze, M.; Andres, M.; Ulbig, A. Investigating Man-in-the-Middle-based False Data Injection in a Smart Grid Laboratory Environment. In Proceedings of the 2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), Espoo, Finland, 18–21 October 2021; pp. 1–6. [CrossRef]
27. Wlazlo, P.; Sahu, A.; Mao, Z.; Huang, H.; Goulart, A.; Davis, K.; Zonouz, S. Man-in-the-middle attacks and defence in a power system cyber-physical testbed. *Iet-Cyber-Phys. Syst. Theory Appl.* **2021**, *6*, 164–177. [CrossRef]
28. Zahri, S.; Bennouri, H.H.; Chehri, A.; Abdelmoniem, A.M. Federated Learning for IoT Networks: Enhancing Efficiency and Privacy. In Proceedings of the 2023 IEEE World Forum on Internet of Things (WF-IoT), Aveiro, Portugal, 12–27 October 2023.
29. Wen, M.; Xie, R.; Lu, K.; Wang, L.; Zhang, K. FedDetect: A Novel Privacy-Preserving Federated Learning Framework for Energy Theft Detection in Smart Grid. *IEEE Internet Things J.* **2021**, *9*, 6069–6080. [CrossRef]
30. Ibrahem, M.I.; Mahmoud, M.; Fouda, M.M.; ElHalawany, B.M.; Alasmary, W. Privacy-preserving and Efficient Decentralized Federated Learning-based Energy Theft Detector. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; IEEE: Rio de Janeiro, Brazil, 2022; pp. 287–292. [CrossRef]
31. McMahan, H.B.; Ramage, D.; Talwar, K.; Zhang, L. Learning Differentially Private Recurrent Language Models. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
32. Aqajari, S.A.H.; Naeini, E.K.; Mehrabadi, M.A.; Labbaf, S.; Rahmani, A.M.; Dutt, N. GSR analysis for stress: Development and validation of an open source tool for noisy naturalistic GSR data. *arXiv* **2020**, arXiv:2005.01834.
33. Di Martino, F.; Delmastro, F. High-resolution physiological stress prediction models based on ensemble learning and recurrent neural networks. In Proceedings of the 2020 IEEE symposium on computers and communications (ISCC), Rennes, France, 7–10 July 2020; pp. 1–6.
34. Alshareef, M.S.; Jaber, M.; Abdelmoniem, A.M. A Differential Privacy Approach for Privacy-Preserving Multi-Modal Stress Detection. In Proceedings of the CAMAD 2023—2023 International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, Edinburgh, Scotland, 6–8 November 2023.

35. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006; pp. 265–284.
36. Farooq, M.; Riaz, S.; Tehseen, R.; Farooq, U.; Saleem, K. Role of Internet of things in diabetes healthcare: Network infrastructure, taxonomy, challenges, and security model. *Digit Health* **2023**, *9*, 20552076231179056. [CrossRef] [PubMed]
37. Wagner, I.; Eckhoff, D. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.* **2018**, *51*, 1–38. [CrossRef]
38. Yin, X.; Zhu, Y.; Hu, J. A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. *ACM Comput. Surv.* **2021**, *54*, 1–36. [CrossRef]
39. Zhu, T.; Ye, D.; Wang, W.; Zhou, W.; Philip, S.Y. More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 2824–2843. [CrossRef]
40. Jegorova, M.; Kaul, C.; Mayor, C.; O'Neil, A.Q.; Weir, A.; Murray-Smith, R.; Tsaftaris, S.A. Survey: Leakage and privacy at inference time. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 9090–9108. [CrossRef] [PubMed]
41. Chen, D.; Jiang, X.; Zhong, H.; Cui, J. Building Trusted Federated Learning: Key Technologies and Challenges. *J. Sens. Actuator Netw.* **2023**, *12*, 13. [CrossRef]
42. Lu, H.; Liu, C.; He, T.; Wang, S.; Chan, K.S. Sharing models or coresets: A study based on membership inference attack. *arXiv* **2020**, arXiv:2007.02977.
43. 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.
44. Hu, H.; Pang, J. Stealing Machine Learning Models: Attacks and Countermeasures for Generative Adversarial Networks. In Proceedings of the ACSAC'21—Annual Computer Security Applications Conference, Virtual Event, 6–10 December 2021; pp. 1–16. [CrossRef]
45. Lee, J.; Han, S.; Lee, S. Model Stealing Defense against Exploiting Information Leak through the Interpretation of Deep Neural Nets. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, Vienna, Austria, 23–29 July 2022.
46. Jere, M.S.; Farnan, T.; Koushanfar, F. A Taxonomy of Attacks on Federated Learning. *IEEE Secur. Priv.* **2021**, *19*, 20–28. [CrossRef]
47. Wibawa, F.; Catak, F.O.; Kuzlu, M.; Sarp, S.; Cali, U. Homomorphic Encryption and Federated Learning Based Privacy-Preserving CNN Training: COVID-19 Detection Use-Case. In Proceedings of the EICC'22—2022 European Interdisciplinary Cybersecurity Conference, Barcelona, Spain, 15–16 June 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 85–90. [CrossRef]
48. 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 CCS'17—2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 1175–1191. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.