Privacy Boundary Determination of Smart Meter Data Using an Artificial Intelligence Adversary

Xiao-Yu Zhang[1,2], Chris Watkins[2], Clive Cheong Took[1], Stefanie Kuenzel[1]

[1] Department of Electronic Engineering, Royal Holloway, University of London, UK

[2] Department of Computer Science, Royal Holloway, University of London, UK

Correspondence

Xiao-Yu Zhang, Address of the corresponding author

Email: email address of the corresponding author

Summary: The roll-out of the new generation smart meter with artificial intelligence (AI)-based data mining algorithms causes serious privacy issues for consumers. By detecting appliance usages, an adversary can easily monitor the behaviour patterns of residents. In this paper, a privacy-preserving smart metering model is proposed; the system utilizes a data aggregator to aggregate the readings of neighbouring smart meters and a data down-sampler to reduce the sensitive information in the load profiles. An AI-based adversary is introduced to simulate the adversarial process. Four state-of-the-art deep learning/machine learning algorithms (convolutional neural network–long short-term memory (CNN-LSTM); gated recurrent unit (GRU); k-nearest neighbours (KNN); and CNN) are employed as data mining algorithms. By tuning the variables (aggregation size α and interval resolution σ), the detectability boundaries of particular appliances are evaluated. Based on the appliance detectability, a three-level privacy boundary (real-time surveillance, presence/absence detection, and complete protection) is obtained. The result shows that to achieve complete data protection, the aggregation size should exceed 40, and the interval resolution should exceed 8 hours.

KEYWORDS: Smart metering infrastructure, privacy benchmark, privacy-preserving computing, energy disaggregation, deep learning, artificial intelligence.

List of Symbols and Abbreviations: AI, artificial intelligence; LSTM, long short-term memory; GRU, Gated Recurrent Units; RNN, recurrent neural network; KNN, k-nearest neighbours; CNN, convolutional neural network; GAN, generative adversarial network; NILM, non-intrusive load monitoring; MO, microwave oven; STO, stove; AC, air conditioner; FUR, furnace; EV, electric vehicle; REF, refrigerator; WH, water heater; DRY, dryer; DW, dishwasher; HVAC, heating, ventilation, and air conditioning; t, time slice of the smart meter; T, the total number of time slices; j, smart meter series

number; X_t^j, the power consumption of smart meter j at time slice t; X^T, load profile sequence; τ, original interval resolution; i, electricity appliance series number; N, total appliance categories; Y_t^i, power consumption of appliance i at time slice t; $Y^{N\times T}$, appliance profile sequence matrix; M^T, modified load sequence; $\mathcal{A}$, adversary model; $\mathcal{P}$, privacy-preserving model; $f_{\mathcal{P}}$, privacy-preserving functions; $f_{\mathcal{A}}(t)$, adversary function; α, aggregation size; σ, downsampled interval resolution; γ, the ratio of modified interval resolution and original interval resolution; r_t, reset gate; z_t, update gate; h_{t-1}, previous cell state; h_t, current cell state; $\tilde{h}_t$, candidate cell state; g_t, input node, i_t, input gate, c_t, internal gate, f_t, forget gate, o_t, output gate; ρ, Pearson correlation coefficient; ϕ, tanh activation.

1. INTRODUCTION

1.1 Motivation

The smart meter is a new generation electricity measurement device that enables real-time communication between the demand side and the utility. This meter also provides high-granularity electricity data (high-interval resolution data on real-time energy consumption, bills, time-of-use tariffs, etc.) [1]. Moreover, the high granularity of data boosts artificial intelligence (AI) applications in smart grids. AI data analysis and data mining tools (such as machine learning/deep learning) have been widely adopted in smart grid applications, such as short-term load forecasts, renewable energy management, and nonintrusive load monitoring (NILM) [2]. However, smart meter and AI applications are double-edged swords since they introduce severe privacy issues to consumers. By adopting AI mining algorithms on smart meter data (such as NILM), the adversary can easily infer personal information from smart meter data [3].

1.2 Literature review

To protect private information in smart meter data, two categories of approaches are proposed in the literature: demand shaping and data manipulation. Demand shaping techniques mask the ground truth load profiles by utilising extra energy storage facilities (such as a rechargeable battery and renewable energy system). The energy management unit (EMU) controls the energy storage device charge/discharge to fill the gap between the "average daily demand" and "instantaneous demand" to minimise information leakage [4].

Data manipulation modifies the original smart meter data with informatics techniques before sending the data to the utility [3]. Among all informatic techniques, the data aggregation approach, data distortion approach, and data down-sampling approach

are widely discussed in the literature. The data aggregation approach (or spatial aggregation) envisages sending aggregate power measures for a group of smart meters to prevent the utility from distinguishing individual power consumption [5]. The data aggregation scheme introduces a data aggregator (DA) with/without a trusted third party (TTP). To guarantee security during data communication, encryption mechanisms such as homomorphic encryption (HE) [6] and multiparty computation (MPC) [7] are introduced. These advanced encryption algorithms enable third parties to operate the data without knowing the details of the data. The data down-sampling approach (or temporal aggregation) aggregates the data from neighbouring timestamps [8]. As the interval spans, the sensitive information in the load profile also decreases [9].

Empirical methods to quantify the privacy boundary are discussed by N. Buescher et al. [10] and EA Technology [11]. A naïve statistical analysis is implemented in [11], and three privacy metrics, visual inspection, correlation analysis, and clustering analysis, are proposed in this work to determine the optimum aggregation size. Their result shows that two houses are enough to achieve high-level anonymity. However, another study by N. Buescher shows that challengers can still obtain an advantage with a minimum aggregation size of 100 houses [10]. These conventional methods can only measure the similarity between the individual power consumption and aggregated power consumption rather than privacy leakage; the adversarial model is also not introduced.

Relevant work that utilises AI adversaries to protect privacy includes the differential privacy NILM algorithm, generative adversarial privacy model, and NILM adversarial model. In differential privacy NILM, a differential private stochastic gradient descent (DP-SGD) mechanism is employed [12]. Random Gaussian noise is added to the gradient of every training step, achieving (ε, δ) differential privacy [13]. M. Shateri et al. [14] introduce an adversarial modelling framework that consists of a data releaser and an adversary. Both the releaser and the adversary utilise recurrent neural networks against each other. The privacy performance of the releaser is improved because of competition. G. Eibl and D. Engel [15] discuss the relationship between interval resolution and privacy in edge detection-based NILM technology. They find that with intervals under 15 min, which is the sampling frequency adopted by most EU manufacturers, most appliances are still detectable.

Although many works have proposed different privacy-preserving smart metering schemes, few studies demonstrate the process of how the adversary obtains valuable information from the load profile. Moreover, there is a lack of information on the correlation of data granularity (e.g., interval resolution, aggregation size) with the sensitivity information.

1.3 Contributions

Inspired by the generative adversarial network (GAN) proposed by I. Goodfellow in 2014 [16], this work trains an artificial intelligence adversarial model to improve the performance of the privacy-preserving model and further detect the boundary of the privacy-preserving model. The main novelties of this paper are listed as follows:

(1) A privacy-preserving smart metering system that combines a data aggregation approach and a data down-sampling approach is proposed. The system enables functionalities (billing, grid management and operation) and simultaneously protects private information.

(2) This work employs an AI-based adversary model to demonstrate the adversarial process. The adversary can use state-of-the-art convolutional neural network–long short-term memory (CNN-LSTM), gated recurrent unit (GRU), CNN, and k-nearest neighbours (KNN) deep neural networks to detect appliance usages and further infer the behaviour patterns of the residents.

(3) The influence of two parameters, aggregation size α and interval resolution σ, on the appliance detectability is investigated by simulation. Nine typical appliances that represent three load categories (continuous load, intermittent load, and active load) are included in the study.

(4) A three-level privacy boundary (real-time surveillance, presence/absence detection, complete protection) is presented based on the simulation results. This benchmark would either benefit consumers to better understand how safe smarts are installed in their homes or contribute to policymakers in regulating smart meter markets.

1.4 Organisation of the paper

The remainder of the paper is organised as follows: The problem formulation is demonstrated in Section 2. In Section 3, the privacy-preserving model, as well as the AI adversary model, is introduced. In Section 4, the implantation process, which

includes dataset construction, data preprocessing, and privacy metrics, is illustrated. Three case studies are designed in Section 5 to determine the privacy boundaries of smart meter data, including aggregation size, interval resolution, and the combined effect of these two factors. The conclusion and future works are drawn in the last section.

2. PROBLEM FORMULATION

Referring to X. Zhang et al. [17], the privacy intrusion issues raised by smart meters include data sensitivity and algorithm sensitivity. For data sensitivity, real-time high-resolution data (active/reactive power, voltage, time-of-use tariff, etc.) collected by the new generation smart meter provides rich information for adversaries. The adversaries can access the collected smart meter data (e.g., purchase from the energy suppliers or hack into the smart metering system). State-of-the-art data-driven deep learning-based NILM algorithms enable adversaries to extract behaviour patterns based on high granularity data (refer to Figure 1).

In this paper, we denote the power consumption recorded by the smart meter at time slice $t \in \mathcal{T} := \{1,2,\cdots,T\}$ as X_t, and the original interval resolution is denoted as τ. In conventional smart metering systems, X_t can be decomposed into individual appliance signals via the NILM algorithm implemented by a third party:

$$X_t = \sum_1^N Y_{i,t} \quad (i \in \{1,2 \cdots N\}) \tag{1}$$

where Y_t^i is the power consumption of electrical appliance i (ranging from 1 to N) at time slice t. The load profile sequence is denoted as X^T. The appliance profile sequence matrix is denoted as $Y^{N\times T}$:

$$Y^{N\times T} = \begin{bmatrix} Y_{1,1} & \cdots & Y_{1,T} \\ \vdots & \ddots & \vdots \\ Y_{N,1} & \cdots & Y_{N,T} \end{bmatrix} \tag{2}$$

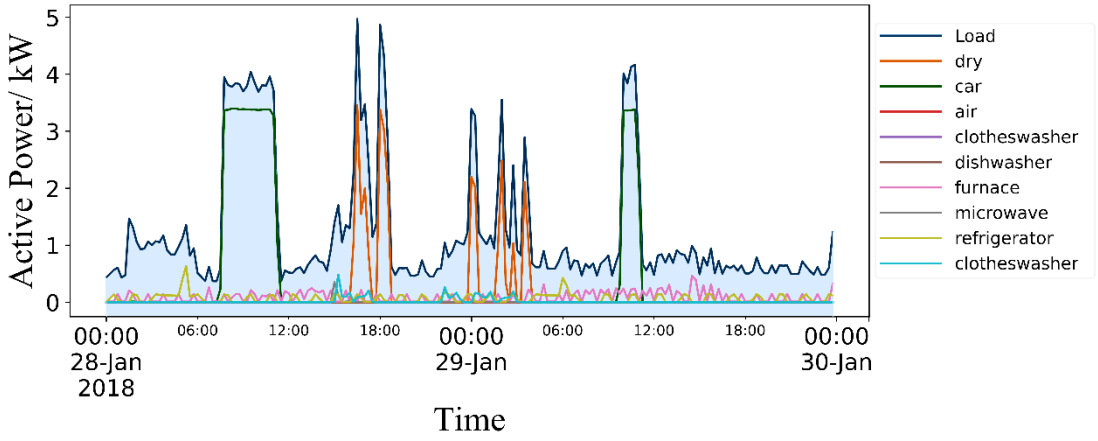

Figure 1 Example of household load profile, with detailed appliance usages (Data source: Pecan Street

Dataport)

The mathematical model that shows the data privacy preservation and adversary inference process is presented in Figure 2. Since $Y^{N\times T}$ contains sensitive information that can be used for behaviour pattern identification, the purpose of the privacy-preserving model $\mathcal{P}$ is to modify the original load profile X^T into a modified load sequence M^T to hide sensitive information $Y^{N\times T}$. In this paper, two privacy-preserving functions $f_{\mathcal{P}}$ are thoroughly investigated: the data aggregation function and the data down-sampling function, as shown in Section 3. Moreover, the difference between X^T and M^T is measured by mutual information (MI). In contrast, the purpose of the adversary model $\mathcal{A}$ is to infer information about $Y^{N\times T}$ from M^t as much as possible ($p(Y^{i,t}|M^t)$) at the real-time base, and the adversary function $f_{\mathcal{A}}(t)$ is expressed as:

$$f_{\mathcal{A}}(t) = argmax\, p(Y^{i,t}|M^t) \tag{3}$$

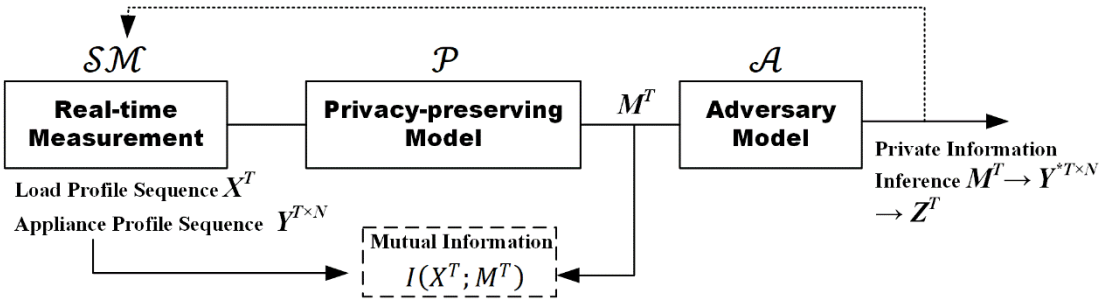

Figure 2 Mathematical model of the privacy-preserving and adversary inference process

3. PRIVACY-PRESERVING SMART METERING FRAMEWORK

3.1 Privacy-preserving model

The conventional smart meter has a fixed sampling frequency and directly sends the power consumption data to the utility without any modification. This single-channel smart metering system has a high risk of revealing private information to the energy supplier or third parties. To overcome the drawbacks of the existing smart metering system, a two-channel smart metering system is proposed (refer to Figure 3). The main structure of the proposed system is an aggregator and a data down-sampler. The purpose of the data aggregator is to concurrently aggregate the smart meter data of neighbouring smart meters and send the aggregated data to the grid operator; then, the grid operator sends commands to manage and operate the grid. The data down-sampler channel down-samples the data for billing purposes only.

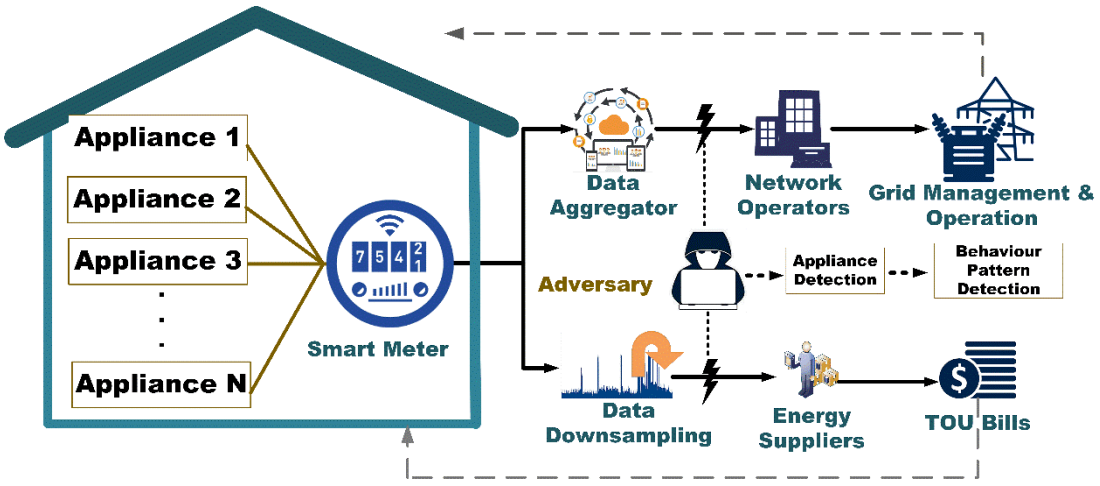

Figure 3 Privacy-preserving smart metering framework

3.1.1 Data aggregation scheme

In the data aggregation scheme, the privacy-preserving function $f_{\mathcal{P}}$ is the data aggregation function. In this scheme, a data aggregator that aggregates all smart meters under the aggregator is constructed. It is meaningful to quantify the aggregation size that can satisfy the privacy requirement to minimise investments. Encryption methods such as HE [18], zero-knowledge protocols [19], and MPC [20] are applied to guarantee communication between smart meters and the aggregator. Detailed encryption algorithms are beyond the scope of this paper.

As shown in Figure 4, X_t^j is the reading of smart meter j $(1 \ll j \ll \alpha)$, where α is the total number of smart meters under the aggregator. At each timestep t, an aggregator synchronously aggregates readings from all smart meters:

$$f_{\mathcal{P}}^{agg}(t) = \sum_{j=1}^{\alpha} X_t^j; t = 1,2,\cdots,T \tag{4}$$

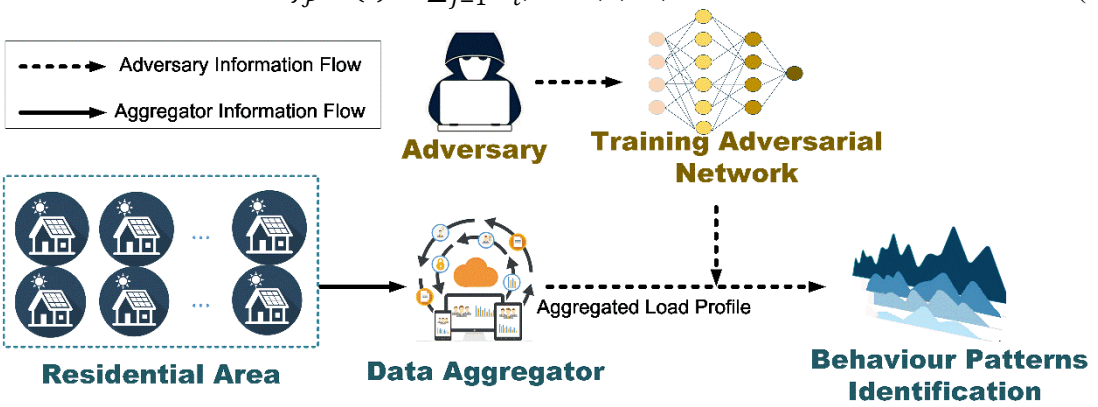

Figure 4 Privacy-preserving data aggregation channel

3.1.2 Data down-sampling scheme

The interval resolution τ of the existing smart meter ranges from 5 seconds to 15 minutes depending on the manufacturer [21]. Current NILM algorithms achieve high accuracy even with a low sampling rate [22]. Hence, as a vital variable that influences private information leakage, the privacy boundary of τ should be quantified. The down-sampling channel aims to reduce sensitive information by reducing the interval resolution of the metered data. A simplified down-sampling scheme is shown in Figure 5; the original curve is flattened by taking the average power consumption of several sampling points. We define a down-sampled interval resolution σ, which is an integer multiple of the original interval resolution τ ($\gamma = \sigma/\tau$). At the end of each γ time slice, the down-sampler takes the average value of all data within the time window:

$$f_{\mathcal{P}}^{down}(j,t) = \frac{\sum_{t=1}^{\gamma} X_t^j}{\gamma}; j = 1,2,\cdots,\alpha \tag{5}$$

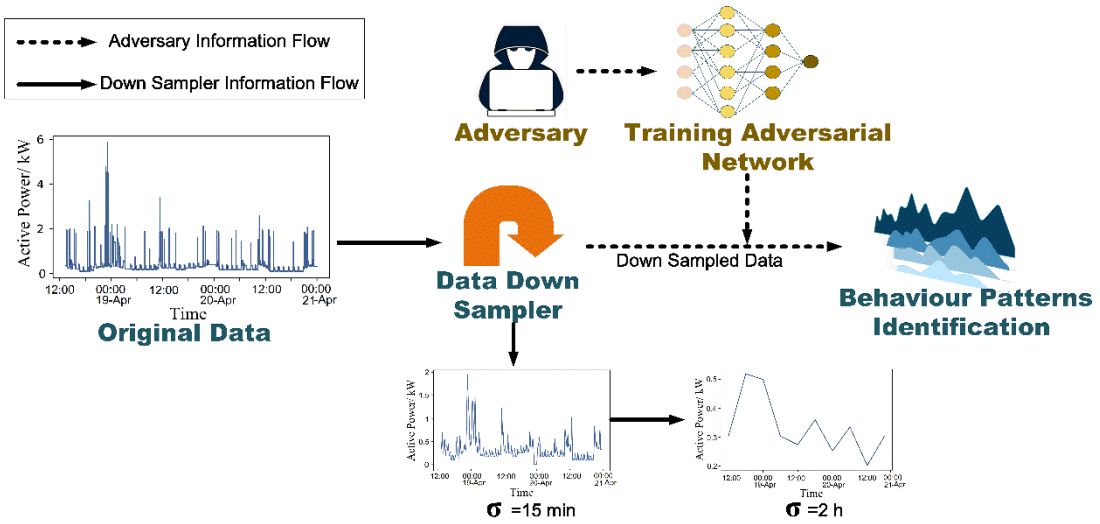

Figure 5 Privacy-preserving down-sampling channel

3.2 Deep Learning Adversary Model

3.2.1 Long Short-Term Memory (LSTM)

Unlike conventional recurrent neural networks (RNNs), which are designed for short-term memory and have poor performance for long sequences (vanishing gradient), LSTM retains both long-term and short-term information without much loss by introducing a memory cell. Moreover, LSTM has gates to help memory cells regulate information from the past.

LSTM has recurrent edges that connect adjacent time steps, enabling LSTM to selectively pass information across sequence steps. The structure of a typical LSTM

block is shown in Figure 6. As demonstrated in (6-8), the components inside the block include an input node g_t, input gate i_t, internal gate c_t, forget gate f_t, output gate o_t, and output h_t. The gate's nature is a sigmoid unit (output range between [0, 1]); it can recognise and pass important information and block unimportant information. Once the input and output gates are closed, the flow will be blocked inside the memory cell and will not affect the following time steps until the gate reopens.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$f_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

$$o_t = \sigma\,(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

Both g_t, i_t, f_t and o_t are the functions of data in the current time step input x_t and the output of the previous time step h_{t-1}, and b_i, b_f, b_o are the bias parameters of nodes. After the values of the gates are determined, the candidate value $\tilde{c}_t$ is calculated and compared with the previous cell state c_{t-1}. With the gate status i_t and f_t, the memory cell determines whether to update its value. By regulating the current cell state c_t with the tanh activation function ϕ and multiplying by the output gate, the output value is calculated; refer to (9-11):

$$\tilde{c}_t = \phi\,(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{9}$$

$$c_t = \tilde{c}_t \odot i_t + c_{t-1} \odot f_t \tag{10}$$

$$h_t = \phi(c_t) \odot o_t \tag{11}$$

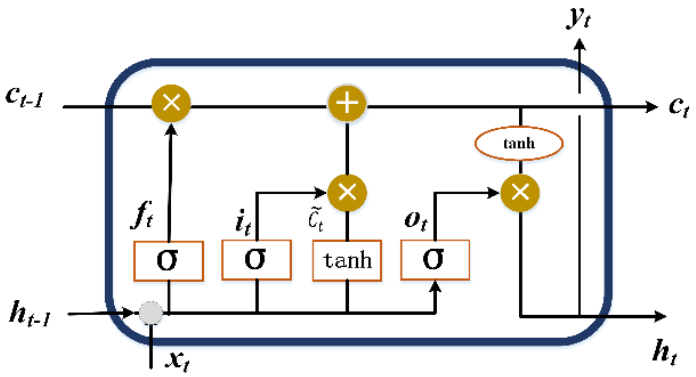

Figure 6 Structure of LSTM-RNN

3.2.2 NILM-based adversary model

In this paper, a 1DCNN-LSTM NILM model is adopted as the adversary $\mathcal{A}$. Based on the previous formulation, a deep neural network is constructed; detailed hyperparameter settings are listed in Table 1. The adversary takes the modified data sequence M^T from the privacy-preserving model $\mathcal{P}$ as input, and the target of $\mathcal{A}$ is to identify the behaviour patterns $Y^{N\times T}$; refer to Figures 4 and 5. The performance of

$\mathcal{A}$ on a single house and original interval resolution is shown in Table 2, which shows that the AI adversary achieves an average accuracy of 83%, which shows that the adversary has a high computation ability in detecting behaviour patterns. The aggregation size α and down-sampling resolution σ increase steadily until the adversary action of $\mathcal{A}$ fails to M^T.

(1) Input Data: The collected data are pre-processed and fed into the model.

(2) 1st 1D Convolutional Layer: The 1D convolutional layer is effective in extracting features from one-dimensional data, especially time-series data. Sixteen filters with a kernel size of 4 are designed to allow the first layer to learn 16 different features from the inputs.

(3) 2nd 1D Convolutional Layer: The output of the first 1D CNN layer is then fed to the second 1D CNN layer. Thirty-two filters with a kernel size of 4 are defined.

(4) Max Pooling Layer: A max-pooling layer is typically employed after CNN layers to avoid overfitting problems raised by the CNN layer. The size of the pooling layer is chosen as 3 to reduce the output matrix size to one-third of the input matrix, and the complexity of the output is reduced as a result.

(5) 1st Bidirectional LSTM Layer: Compared to the conventional LSTM layer, bidirectional LSTM has better performance for time-series data since it can learn the inputs from both the forward direction and backward direction. In this layer, we define 512 LSTM units.

(6) 2nd Bidirectional LSTM Layer: In this layer, we define 512 LSTM units.

(7) Dropout Layer: Dropout is an effective regularization method that is employed in neural networks to avoid overfitting. The dropout layer will randomly set the weight of neurons to zero during the training process. In this model, we set the dropout rate to 0.5, which means that 50% of neurons obtain a zero weight.

(8) Fully Connected Layer: The final layer will reduce the output matrix to a single value between 0 and 1, which is the estimated active power of the targeted household appliance.

Table 1 Adversary network settings.

Hyperparameters	**Value**	**Description**
Learning rateϵ	0.05	Steps to minimise error.
Optimiser	Adam	
Number of LSTM/GRU layers	4	
LSTM/GRU units per RNN layer	512	
Number of 1D CNN layers	1	Extracting features from time-series data
Kernel size of 1D CNN layer	5	Sliding window size of the 1D CNN
Batch size B	128	Number of training examples utilised in one iteration.
Activation function for hidden layers	ReLU	$f_{ReLU} = \max[0, z]$.

Activation function for the output layer	ReLU	Positive Output.
Epoch number	100	One cycle through the entire training dataset.
Loss function	MSE	Minimise the error between ground truth and prediction
Dropout	0.5	Reduce overfitting

4 IMPLEMENTATION

4.1 Dataset construction

The data adopted in this paper are The Reference Energy Disaggregation Data Set (REDD) [23] and Pecan Street Dataport (Dataport) [24]; refer to Table 3. Both datasets contain appliance-level and house-level power consumption data. Hence, not only the load profiles but also the appliance signatures can be obtained from the datasets. We select nine typical household appliances for this research: air conditioner (AC), microwave oven (MO), electric vehicle (EV), water heater (WH), dishwasher (DW), dryer (DRY), stove (STO), furnace (FUR), and refrigerator (REF). Three variables related to the appliance—the power rating, minimum duration, and power threshold—are described in Table 2. The power threshold in the table represents the minimum power to operate the appliance. The threshold is the minimum power to start the device; when the power is larger than the power threshold, we regard the appliance as "on". Minimum duration represents the minimum operating hours of a particular appliance throughout the day. The rated power is the highest power input allowed through a particular device.

Aggregation Size Dataset: Referring to Section 3.1.1 and (4), the houses inside an aggregation group are selected randomly from two datasets to make up the new dataset. The new dataset is split into training/testing datasets (90% for training and 10% for testing). The input data of the model are the aggregated power consumption $f_{\mathcal{P}}^{agg}(t)$, and the output of the model is the power consumption of a particular appliance $Y^{i,t}$ in house i.

Interval Resolution Dataset: The original dataset is down-sampled using the down-sampling method mentioned in the previous section. The new dataset is generated by taking the average values of γ time slices. The new dataset is also divided into training/testing datasets; both the input (household power consumption) and the output (appliance consumption) are obtained from the same house i.

Table 2 Property of appliances

Appliance	Rating (kW)	Threshold (kW)	Min Drn (h)	Adversary Acc. (%)
Microwave Oven (MO)	1.5	0.30	0.025	0.77
Stove (STO)	1.2	0.24	2	0.89

Air Condition (AC)	2.0	0.40	12	0.85
Furnace (FUR)	1.0	0.20	8	0.91
Electric Vehicle (EV)	3.0	0.50	4	1.00
Refrigerator (REF)	0.055	0.01	24	0.94
Water Heater (WH)	3.5	1.00	2.5	0.75
Dryer (DRY)	2.1	0.7	1	0.76
Dishwasher (DW)	1.2	0.15	2	0.90

Table 3 Dataset description

Dataset	**Interval Resolution**	**NUM. of Houses**	**NUM. of Submeters**	**Duration**
Dataport[24]	1 min	$\gg 1000$	75	4 years
REDD[23]	3 s	6	20	30 days

4.2 Data preprocessing

The purpose of data pre-processing is to make the input data more amendable to the model. Typically, data pre-processing consists of vectorization, missing data detection, and data normalisation. Since the input data are already vectorised, only normalisation and missing data detection are required.

4.2.1 Missing data detection

There are some missing values in the original data for some reason; these missing values will influence the performance of the model. In this study, we replace all missing values with '0'.

4.2.2 Data normalisation

Normalisation is vital to the neural network to prevent it from converging. In this work, max-min normalisation is adopted to guarantee that all input values range between 0 and 1. The equation of max-min normalisation is shown in (12):

$$x_{normalized} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{12}$$

where max (x) and min (x) represent the maximum value of the data and minimum value of the data, respectively.

4.3 Hardware & software platform

The simulation and computation are implemented on a Dell laptop equipped with a Core i7-7700HQ CPU, NVIDIA GTX 1060 GPU, and 8 GB RAM. The deep learning algorithm runs on Python 3.7, and the TensorFlow 2 framework is adopted to train the DNN model.

4.4 Privacy metrics for appliance detection

Once the adversary model is designed, the performance of the adversary should be evaluated and quantified. In this section, we introduce two performance metrics that assess the performance of DNNs.

4.5.2 F-measure (F1 score)

The F-measure is a performance measurement for classification adopted in NILM works and privacy measures [15]. As shown in Table 4, there are four combinations of the confusion matrix (TP, FP, FN, and TN); each element represents one estimation condition (whether the estimation is correct or incorrect).

Table 4 Confusion Matrix.

	Actual Positive	Actual Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Based on the matrix, the F-measure can be calculated (refer to (13)). Usually, when the F-measure is smaller than 0.5, the classifier is inadequate.

$$F-measure = \frac{1}{1+(FN+FP)/2TP} \tag{13}$$

4.5.3 Correlation analysis

The Pearson correlation coefficient ρ is used to measure whether two continuous variables are linearly associated. The value of ρ ranges from -1 to 1 (a positive value indicates a positive correlation, while a negative value indicates a negative correlation); the larger is ρ, the stronger is the correlation between two variables. The expression of the Pearson correlation coefficient is shown in (14):

$$\rho = \frac{\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t-\bar{x})^2 \sum_{t=1}^{n}(y_t-\bar{y})^2}} \tag{14}$$

where n is the sample size, x_t is the appliance power consumption at time t and y_t is the power consumption generated by the adversary; $\bar{x}$ and $\bar{y}$ are the mean values of x_t and y_t, respectively. A benchmark is presented for the following analysis process; refer to Table 5. We define an appliance as measurable when two metrics, the F-measure and ρ, are lower than 0.2.

Table 5 Benchmarks of privacy metrics in appliance detection

Performance	F-measure	Pearson correlation coefficient (ρ)
Poor privacy protection	0.5–1	0.5–1
Fine privacy	0.2–0.49	0.2–0.49
Good privacy	0.01–0.19	0.01–0.19
Perfect privacy	<0.01	<0.01

5 RESULTS AND DISCUSSION

This section quantifies the privacy boundary influenced by aggregation size α and interval resolution σ. Two case studies are designed for each parameter; both the detectability of particular appliances and algorithm sensitivity in two privacy-preserving schemes are thoroughly investigated. A discussion based on the results is also presented to demonstrate the proposed three-level privacy benchmarks.

5.1 Privacy boundary level based on electrical events

Household appliances can be divided into three categories, loads depending on the characteristics and operating duration of the loads [25]. Detailed classifications are described as follows:

Continuous load: A continuous load means that the device consumes energy throughout the day, such as the refrigerator and freezer, as well as the computer and printer in "standby" mode. Since continuous loads are not influenced by residents' activities, these loads contain minimal sensitive information.

Intermittent load: These appliances are not always on but are active enough to be recorded by the lowest hourly smart meters, such as air conditioners, electric vehicles, furnaces, and water heaters.

Active load: Power use of appliances in an active house, such as microwave ovens, dishwashers, stoves, and dryers.

Based on the previously introduced load categories, three-level privacy boundaries are defined:

Level I (Real-Time Surveillance): All loads, including continuous loads, intermittent loads, and active loads, are detected by $\mathcal{A}$. $\mathcal{A}$ knows the entire life cycles of all residents (sleeping pattern, number of residents, when people leave their home, etc.). Private information of residents is at high risk at this level.

Level II (Presence/Absence Detection): Both continuous loads and intermittent loads are detected by $\mathcal{A}$. At this privacy level, $\mathcal{A}$ knows whether residents are inside/outside the house, but $\mathcal{A}$ cannot monitor all electrical activities inside a house.

Level III (complete protection): No event is detected by the adversary, or only continuous loads are detected by $\mathcal{A}$. At this level of protection, $\mathcal{A}$ cannot infer any sensitive information from given data.

5.2 Privacy boundary of aggregation size α

This case study focuses on the privacy-preserving aggregation channel illustrated in Section 3.1.1. Recalling $f_{\mathcal{P}}^{agg}(t)$ in (4), the aggregation size α is an essential variable that influences the performance of $\mathcal{P}$. The purpose of $\mathcal{A}$ is to detect appliance usage given M^T. As demonstrated in Figures 3 and 4, the precision of detection is evaluated when α increases steadily.

5.2.1 Detectability of particular appliances in an aggregation scheme

$\mathcal{A}$ has high accuracy in appliance detection in a single house, which raises privacy issues related to smart meters. Recall the threshold identified in Table 5; an appliance is defined as detectable when both the F-measure and ρ are higher than 0.2. In this case, study, nine typical appliances, which are introduced in Table 2, are investigated. These appliances represent continuous load (REF), intermittent load (AC, EV, WH, and FUR), and activate load (MO, DW, STO, and DRY), respectively. A 1DCNN-LSTM model with four LSTM layers is adopted as $f_{\mathcal{A}}(t)$. The model achieves high efficiency in detecting appliances in a single house (refer to Table 2). By steadily increasing α from 1 to 100, the number of smart meters inside an aggregator is enlarged. The mutual information between M^T and X^T also decreases with an increase in α, which increases the difficulty of $\mathcal{A}$'s inference process.

Figure 7 presents a heat map to show the performance of $\mathcal{A}$ in appliance detection given different α. As expected, the detectability of $\mathcal{A}$ is high with a small aggregation size ($\alpha < 5$). By a continuously increasing α, both the F-score and ρ consequently decrease, which means that the appliance detectability is also reduced. Appliances such as EV, DW, and WH become undetectable when α reaches 10. Most of these appliances operate during peak periods, and load components under the aggregation scheme are extremely complex during this duration, so the inference process of $\mathcal{A}$ is easily blocked. As α reaches 20, MO, STO, DRY, and REF become undetectable. It should be noted that heating, ventilation, and air conditioning (HVAC) devices, such as AC and FUR, remain detectable even for $\alpha = 40$ as HVAC devices have a long operational duration (8–12 hours per day) and high power rating (1–2 kW). To blind $\mathcal{A}$ for these HVAC devices, a minimum number of 50 houses is required. Figure 8 takes MO, DW, REF, and AC as examples to compare the information inferred by $\mathcal{A}$ and the ground truth data under the aggregation scheme with $\alpha = 1, 2, 5, 50$.

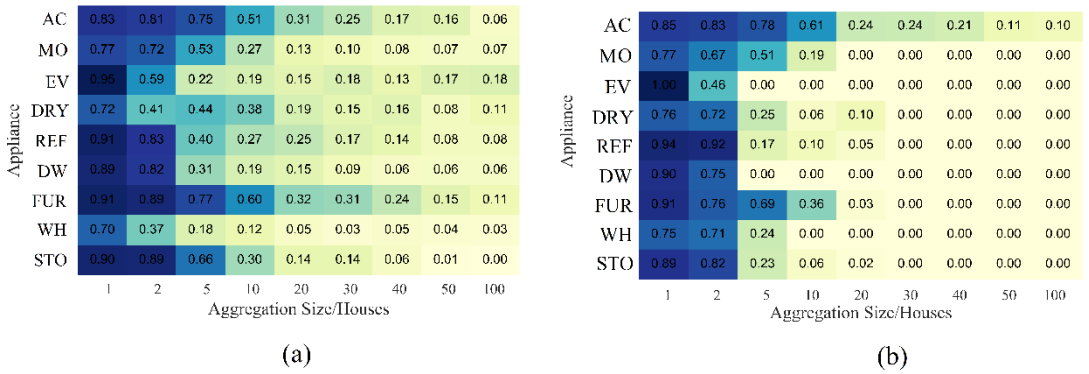

Figure 7 Heatmap of the adversary on particular appliances with different aggregation sizes: (a) Pearson correlation coefficient and (b) F-measure.

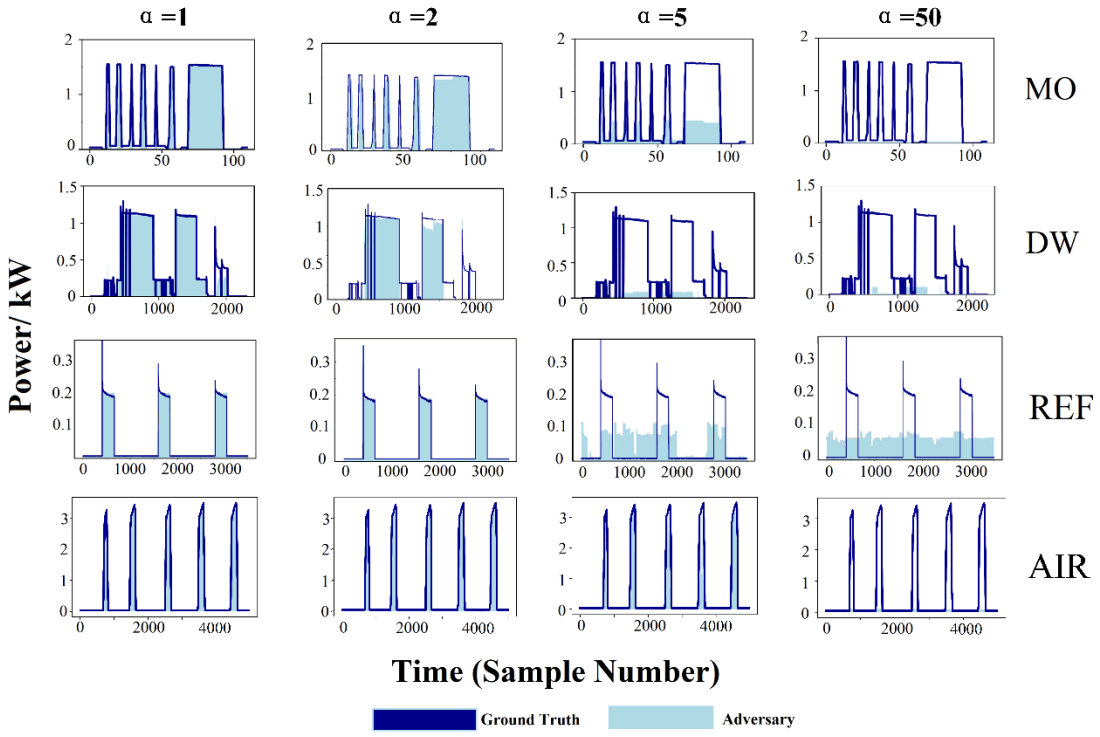

Figure 8 Examples of information inferred from the adversary and ground truth data in the aggregation scheme

Since different appliances have different characteristic properties (rating, threshold, and minimum duration), the performance of $\mathcal{A}$ on different appliances varies greatly. Based on the results shown in Figure 7, a correlation analysis between appliance characteristic properties and adversary detectability is implemented (shown in Table 6). It is observed that the three characteristics show almost equal correlations with adversary detectability (0.44 for Rating, 0.50 for Threshold, and 0.53 for Minimum Duration). To summarise, appliances with high ratings, high threshold, and long duration (such as AC, FUR, and DRY) require larger α to blind $\mathcal{A}$.

Table 6 Correlation between appliance characteristic properties and adversary detectability

	Rating	Threshold	Minimum Duration
α	0.44	0.50	0.53
σ	0.34	0.26	0.74

5.2.2 Sensitivity of algorithms in an aggregation scheme

Rather than the CNN-LSTM algorithm adopted in the previous sections, $\mathcal{A}$ can also adopt different deep learning algorithms. In this case, the sensitivity of the algorithms in an aggregation scheme is discussed. Apart from the proposed algorithm, three $\mathcal{A}$s that adopt state-of-the-art algorithms, such as(GRU) [26], CNN [27], and the neighbour KNN [28] NILM algorithms, are well analysed, referring to previous works. In Figure 9, each bar represents the average values of the F-measure/ρ of all appliances with a particular algorithm. All algorithms have desirable detectability on a single house (F-measure>0.77, and ρ>0.78), and CNN-LSTM has the best performance among all algorithms, followed by the GRU, while CNN and KNN have similar performances. The machine learning algorithm, KNN, is the most sensitive to the parameter α, as $\mathcal{A}$ with KNN becomes blind when α>10, while the other three $\mathcal{A}$s can still infer private information with high accuracy at this level. Moreover, CNN-LSTM and GRU have similar characteristics throughout the whole simulation. Both $\mathcal{A}$ with CNN-LSTM and $\mathcal{A}$ with GRU lose general detectability when α>30 (it should be noted that the general detectability only represents the average privacy metrics of all appliances, and some specific appliances are still detectable, as stated in Section 5.2.1). In summary, the proposed aggregation scheme is efficient for all algorithms discussed in this section, as the detectability of the four algorithms decreases to nearly zero with high aggregation sizes (α>50).

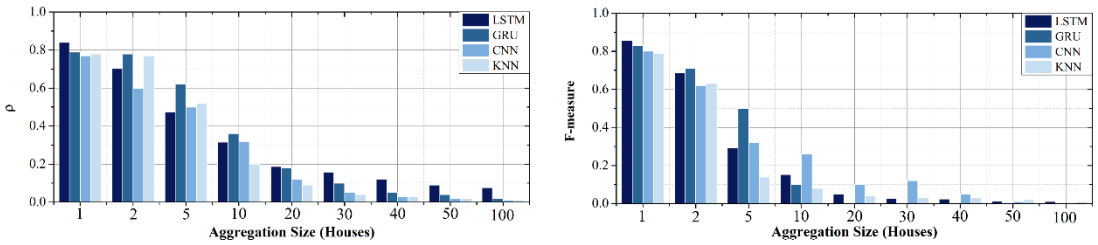

Figure 9 Comparison of different adversary algorithms in the aggregation scheme: (a) Pearson correlation coefficient and (b) F-measure

5.3 Identifying the boundary of interval resolution

The privacy boundary of another critical parameter, interval resolution σ, is discussed in this section. Similar to Section 5.2, two case studies are implemented to investigate the appliance detectability and algorithm sensitivity. The original interval resolution of the dataset is 3 s. By implementing the down-sampling formula in (5), a new dataset with a larger σ is obtained.

5.3.1 Detectability of particular appliances in a data down-sampling scheme

Similar to Section 5.2.1, the detectability of α on appliances regarding different σ is discussed in this section. As shown in Figure 5, high granularity smart meter data with a small α contain more detailed features of the load profile, and $\mathcal{A}$ can easily apply the NILM algorithm and infer private information. From Figure 10, all appliances are highly detectable when σ<5 min, with the exception of MO. Appliances such as MO have a very high rating (1.5 kW), but the operation duration is short (0.025 hours). Hence, when the interval resolution increases, MO becomes challenging to detect. Referring to Table 6, appliance detectability in the data down-sampling scheme has a high correlation with a minimum duration (0.72), followed by a rating (0.34). Appliances with long operation durations require significant σ values to hide sensitive information. For instance, AC requires at least 1 h interval resolution to blind $\mathcal{A}$, and σ>5 h is required by EV. For continuous loads, such as REF, which operates all day, σ should be larger than 10 h. In summary, σ>10 h is required to provide complete privacy. Figure 11 takes MO, DW, and REF as examples to compare the information inferred by $\mathcal{A}$ and the ground truth data under the data down-sampling scheme with $\sigma = 3s, 5min, 0.5h, 2h$.

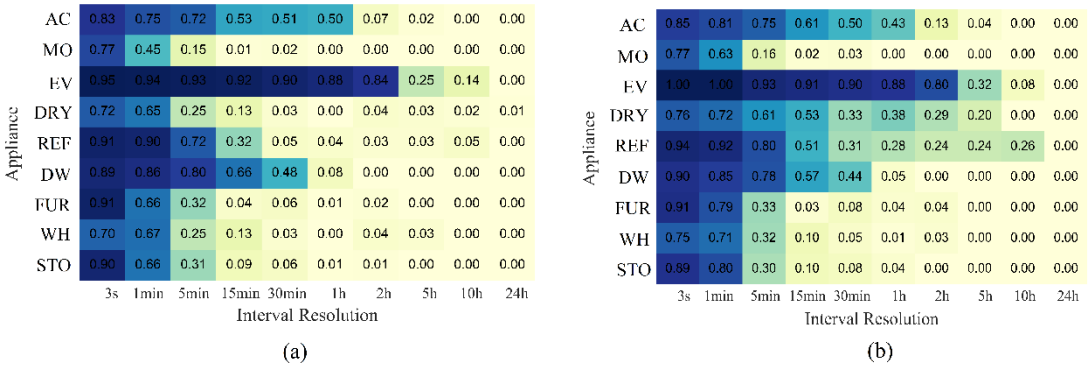

Figure 10 Performance of the adversary on particular appliances with different interval resolutions: (a) Pearson correlation coefficient and (b) F-measure

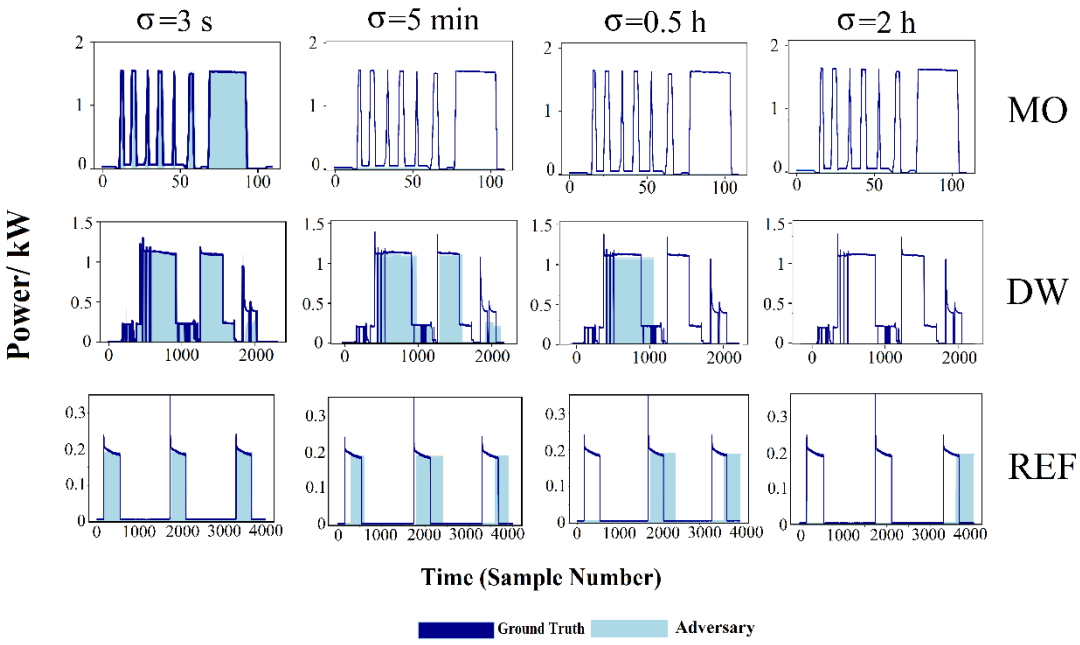

Figure 11 Examples of information inferred by the adversary and ground truth data in the data down-sampling scheme

5.3.2 Sensitivity of algorithms in a data down-sampling scheme

Similar to Section 5.2.2, four adversaries with different algorithms (CNN-LSTM, GRU, CNN, and KNN) are introduced to determine the sensitivity of algorithms in a data down-sampling scheme. As shown in Figure 12, the increase in σ substantially weakens the detectability of all four adversaries. It is essential to note that all adversaries still maintain a high inference ability when σ ranges from 15 to 30 min, while the sample frequencies of most smart meters in the UK are in this scope. This result demonstrates our argument that the current smart metering system in the UK is highly vulnerable and can be abused by $\mathcal{A}$. A benchmark of σ=10 h is a safe threshold for the privacy-preserving model against the attack from $\mathcal{A}$.

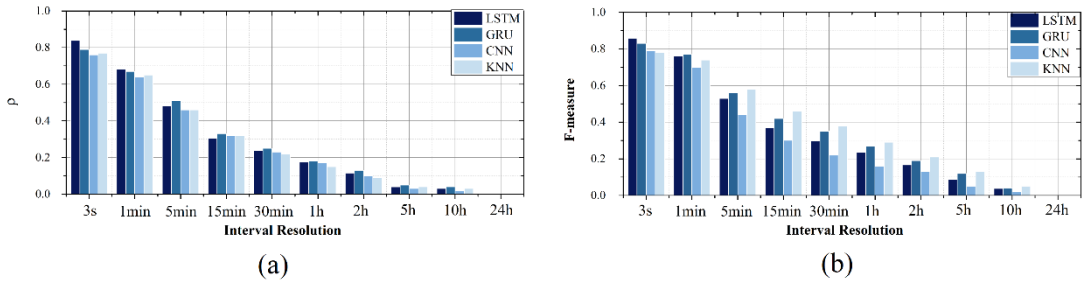

Figure 12 Comparison of different adversary algorithms in the data down-sampling scheme: (a) Pearson correlation coefficient and (b) F-measure

5.4 Combined Effect of Interval Resolution and Aggregation Size

In this section, the combined effect of two parameters, α and σ, on the adversary computing ability is demonstrated. The aggregation size α and interval resolution σ are changed synchronously, and the dynamic variation of two privacy metrics, the F-measure and ρ, are observed. The simulation results are presented in Figure 13, which uses 3D models to show dynamic changes. The detectability recedes rapidly, and both the F-measure and ρ decrease to zero given $\alpha > 10$, and $\sigma > 30\ min$.

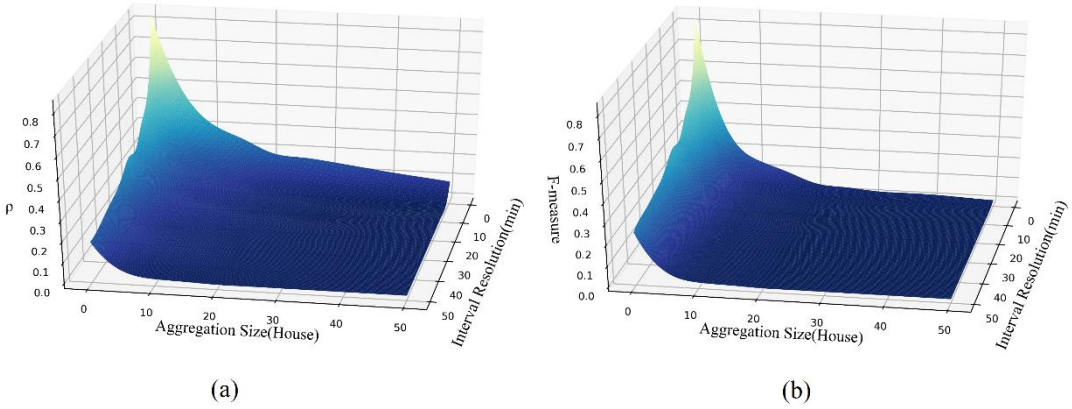

Figure 13 3D model of the privacy performance of the adversary with two parameters: (a) Pearson correlation coefficient and (b) F-measure

5.5 Discussion

Based on the simulation results and quantification of appliance detectability obtained in the previous sections, three-level privacy boundaries are concluded in Table 7. When $\alpha < 20$ or $\sigma < 5h$, consumers are at privacy level I, which represents consumers under real-time surveillance at this level. By detecting appliance signatures of active loads (MO, DW, STO, and DRY), $\mathcal{A}$ can have knowledge of detailed behaviour patterns of residents inside the house. When $20 \leq \alpha < 40$ or $5h \leq \sigma < 8h$, the consumers are at privacy level II, and $\mathcal{A}$ can infer presence/absence information from intermittent loads (AC, EV, WH, and FUR) but cannot understand complex behaviours inside the house. When $40 \leq \alpha$ or $8h \leq \sigma$, the consumers are at privacy level III; at this level, consumers are protected entirely and free of privacy concern. In addition, when we take the co-effects of the two parameters, the detectability of $\mathcal{A}$ decreases dramatically compared to a single parameter. When $10 \leq \alpha$ and $30min \leq \sigma$, privacy level III is already achieved.

Table 7 Quantification of three-level privacy boundaries

Privacy level	Appliance to detect	Quantification (Single parameter)	Quantification (Co-effects of two parameters)

Level I	MO, DW, STO, DRY	$\alpha < 20\ or\ \sigma < 5h$	$\alpha < 2\ and\ \sigma < 5min$
Level II	AC, EV, WH, FUR	$20 \leq \alpha < 40\ or\ 5h \leq \sigma < 8h$	$2 \leq \alpha < 10\ and\ 5min \leq \sigma < 30min$
Level III	All appliances	$40 \leq \alpha\ or\ 8h \leq \sigma$	$10 \leq \alpha\ and\ 30min \leq \sigma$

6 CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this paper, a privacy-preserving smart metering model is proposed; the model adopts a data aggregation scheme and data down-sampling scheme to better protect sensitive information from inference. An AI adversary is then introduced to quantify the privacy boundary (aggregation size and interval resolution) of the smart meter data. The adversary can implement cut edge CNN-LSTM NILM algorithms to detect appliance usage from the demand load curve and further identify the behaviour patterns of consumers. Three case studies are employed to investigate the influence of parameters α and σ and the co-effect of α and σ on the appliance $\mathcal{A}$'s detectability. From the simulation, three-level privacy boundaries are quantified, showing that to achieve Level III privacy (complete protection), the following conditions must be met: (1) $40 \leq \alpha\ or\ 8h \leq \sigma$; (2) 10≤α and 30 min≤σ.

6.2 Implications for Policy

The conclusion obtained in this paper, especially the three-level privacy boundaries, is fundamental to stakeholders in the smart metering system, including consumers, manufacturers, power system operators, and policymakers from the government. As privacy is abstract and hard to quantify, privacy boundaries are easily understandable and provide an insight for people to classify privacy-free and privacy-concerned smart meter data. New generation smart meters can make further improvements based on privacy boundaries. In addition, for smart meter data granularity under privacy boundaries, extra encryption techniques should be adopted by the utility to guarantee the safety of private information.

6.3 Future work

In the future, this work can be extended to the following directions: (1) a combination of the proposed smart metering system with encryption techniques would provide better security and privacy guarantees to consumers; and (2) continuous updating of the privacy boundaries by considering advanced NILM algorithms.

ACKNOWLEDGMENTS

This research is funded by The Leverhulme Trust, UK. Thanks to the anonymous peer-reviewers and editors whose comments helped to improve and clarify this manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

1. Marimuthu KP, Durairaj D, Karthik Srinivasan S. Development and implementation of advanced metering infrastructure for efficient energy utilization in smart grid environment. *International Transactions on Electrical Energy Systems*. 2018;28(3):e2504.
2. Yao R, Li J, Zuo B, Hu J. Machine learning-based energy efficient technologies for smart grid. *International Transactions on Electrical Energy Systems*. 2021/01/04 2021;n/a:e12744.
3. Giaconi G, Gunduz D, Poor HV. Privacy-Aware Smart Metering: Progress and Challenges. *IEEE Signal Processing Magazine*. 2018;35(6):59-78. doi:10.1109/MSP.2018.2841410
4. Yaqub SAR. Artificial Intelligence Assisted Consumer Privacy and Electrical Energy Management. *Global Journal of Computer Science and Technology; Vol 20, No 1-D* 08/24 2020;
5. Farokhi F. Review of results on smart-meter privacy by data manipulation, demand shaping, and load scheduling. *IET Smart Grid*. 2020;3(5):605-613.
6. Li F, Luo B, Liu P. Secure information aggregation for smart grids using homomorphic encryption. presented at: 2010 first IEEE international conference on smart grid communications; 2010;
7. Thoma C, Cui T, Franchetti F. Secure multiparty computation based privacy preserving smart metering system. presented at: 2012 North American power symposium (NAPS); 2012;
8. Engel D, Eibl G. Wavelet-Based Multiresolution Smart Meter Privacy. *IEEE Transactions on Smart Grid*. 2017;8(4):1710-1721. doi:10.1109/TSG.2015.2504395
9. Mashima D. Authenticated down-sampling for privacy-preserving energy usage data sharing. presented at: 2015 IEEE International Conference on Smart Grid Communications (SmartGridComm); 2-5 Nov. 2015 2015;
10. Buescher N, Boukoros S, Bauregger S, Katzenbeisser S. Two is not enough: Privacy assessment of aggregation schemes in smart metering. *Proceedings on Privacy Enhancing Technologies*. 2017;4(2017):198-214.
11. Association EN. Smart meter aggregation assessment final report. 2015;
12. Zhang XY, Kuenzel S. Differential Privacy for Deep Learning-based Online

Energy Disaggregation System. presented at: 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe); 26-28 Oct. 2020 2020;
13. Wang H, Zhang J, Lu C, Wu C. Privacy Preserving in Non-Intrusive Load Monitoring: A Differential Privacy Perspective. *IEEE Transactions on Smart Grid*. 2020;
14. Shateri M, Messina F, Piantanida P, Labeau F. Real-time privacy-preserving data release for smart meters. *IEEE Transactions on Smart Grid*. 2020;11(6):5174-5183.
15. Eibl G, Engel D. Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*. 2014;6(2):930-939.
16. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. *Advances in Neural Information Processing Systems*. 2014;3:2672-2680.
17. Zhang X-Y, Kuenzel S, Córdoba-Pachón J-R, Watkins C. Privacy-Functionality Trade-Off: A Privacy-Preserving Multi-Channel Smart Metering System. *Energies*. 2020;13(12)doi:10.3390/en13123221
18. Garcia FD, Jacobs B. Privacy-Friendly Energy-Metering via Homomorphic Encryption. presented at: International workshop on security and trust management;STM 2010; 2011;
19. Molina-Markham A, Shenoy P, Fu K, Cecchet E, Irwin D. Private memoirs of a smart meter. presented at: Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building; 2010;
20. Mustafa MA, Cleemput S, Aly A, Abidin A. A Secure and Privacy-Preserving Protocol for Smart Metering Operational Data Collection. *IEEE Transactions on Smart Grid*. 2019;10(6):6481-6490. doi:10.1109/TSG.2019.2906016
21. Segovia, Sánchez. - Set of common functional requirements of the SMART METER. *DG INFSO and DG ENER, European Commission, Brussels, Belgium, Tech Rep*. Oct
22. Le TTH, Kim H. Non-Intrusive Load Monitoring Based on Novel Transient Signal in Household Appliances with Low Sampling Rate. *Energies*. 2018;11(12)
23. Kolter J, Johnson M. REDD: A Public Data Set for Energy Disaggregation Research. *Artif Intell*. 01/01 2011;25
24. Street P. Dataport: the world's largest energy data resource. *Pecan Street Inc*. 2015;
25. Delforge P, Schmidt L, Schmidt S. Home Idle Load: Devices Wasting Huge Amounts of Electricity When Not in Active Use(Tech.). NRDC. 2015.
26. Kim J, Kim H. Classification performance using gated recurrent unit recurrent neural network on energy disaggregation. presented at: 2016 international conference on machine learning and cybernetics (ICMLC); 2016;
27. Yang D, Gao X, Kong L, Pang Y, Zhou B. An event-driven convolutional neural architecture for non-intrusive load monitoring of residential appliance. *IEEE Transactions on Consumer Electronics*. 2020;66(2):173-182.
28. Hidiyanto F, Halim A. KNN Methods with Varied K, Distance and Training Data to Disaggregate NILM with Similar Load Characteristic. presented at: Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering 2020; 2020;