Privacy Protection of Synthetic Smart Grid Data Simulated via Generative Adversarial Networks

Kayode S. Adewole[1,2,3][a] and Vicenç Torra[2][b]

[1]Department of Computer Science and Media Technology, MalmöUniversity, Malmö, Sweden

[2]Department of Computing Science, Umeå University, Umeå, Sweden

[3]Department of Computer Science, University of Ilorin, Ilorin, Nigeria

Keywords: Smart Grid, Non-Intrusive Load Monitoring, Generative Adversarial Networks, Data Privacy, Microaggregation, Discrete Fourier Transform.

Abstract: The development in smart meter technology has made grid operations more efficient based on fine-grained electricity usage data generated at different levels of time granularity. Consequently, machine learning algorithms have benefited from these data to produce useful models for important grid operations. Although machine learning algorithms need historical data to improve predictive performance, these data are not readily available for public utilization due to privacy issues. The existing smart grid data simulation frameworks generate grid data with implicit privacy concerns since the data are simulated from a few real energy consumptions that are publicly available. This paper addresses two issues in smart grid. First, it assesses the level of privacy violation with the individual household appliances based on synthetic household aggregate loads consumption. Second, based on the findings, it proposes two privacy-preserving mechanisms to reduce this risk. Three inference attacks are simulated and the results obtained confirm the efficacy of the proposed privacy-preserving mechanisms.

1 INTRODUCTION

The management of power grid has recently advanced through the proliferation of smart technologies. Over the years, smart grid has enjoyed rapid advancement in smart solutions deployment due to the need for managing the grid in an economical, reliable, and sustainable manner (Fekri et al., 2019). One of these technologies is the introduction of advanced metering infrastructure (AMI). AMI enables smart meters to measure and communicate energy consumption within the configured intervals (Fekri et al., 2019; El Kababji and Srikantha, 2020). These measurements have the potential to offer great insights, increase smart grid flexibility, improve decision-making as well as grid reliability (Adewole and Torra, 2022b). Non-intrusive load monitoring (NILM) is one of the technologies that provides the insights.

NILM separates a building's aggregated energy into constituent fine-grained energy demands by the individual household appliances (Batra et al., 2019; Kelly and Knottenbelt, 2015). NILM research has produced techniques for increasing energy efficiency through generation of appliance-level consumption details that can guide consumers to adopt better energy usage habits. Considering the load consumption disaggregation and based on the increasing energy awareness of individual equipment, consumers may adapt consumption behaviours, replace equipment or install management systems focusing on energy optimization (Hart et al., 1989).

Research in NILM focused on energy disaggregation, however, privacy issues in energy disaggregation are a major concern as individual household lifestyles can be inferred from their consumption (Adewole and Torra, 2022a; Adewole and Torra, 2022b). The information inferred can be used by malicious third parties. For instance, cases of cyber-attacks on smart grid have been reported in the recent years (BBCNews, 2017).

This paper targets to solve two issues in smart grid domain. In the first issue, we extend a methodology for synthetic data generation to simulate smart grid data with significant number of households aggregate consumption. This enables us to check the privacy risk associated with the synthetic households

[a] https://orcid.org/0000-0002-0155-7949

[b] https://orcid.org/0000-0002-0368-8037

Adewole, K. and Torra, V.
Privacy Protection of Synthetic Smart Grid Data Simulated via Generative Adversarial Networks.
DOI: 10.5220/0011956800003555
In *Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT 2023)*, pages 279-286
ISBN: 978-989-758-666-8; ISSN: 2184-7711

Copyright © 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0)

aggregated consumption. In the second issue, we propose two privacy-preserving mechanisms to protect smart grid data from household privacy leakage. More specifically, the paper: **(1)** extends the existing data simulation framework to generate smart grid data useful for data analysis and privacy-preserving studies based on generative adversarial networks (GAN) **(2)** investigates the performance of deep learning Sequence-to-Sequence (Seq2Seq) NILM disaggregation algorithm alongside three events extraction methods to ascertain the privacy leakage in publishing smart grid data **(3)** investigates the performance of two privacy-preserving methods for privacy protection of individual household appliances.

The organization of the remaining sections of this paper is as follows: Section 2 discusses the related works in energy disaggregation, synthetic data generation and privacy-preserving methods. Section 3 highlights the proposed methods in this paper. Section 4 focuses on experimental and evaluation procedures. Section 5 presents the results and discusses the findings, and finally, Section 6 concludes the paper and offers future directions.

2 RELATED WORKS

2.1 Energy Disaggregation

Energy disaggregation or Non-intrusive load monitoring (NILM) research has spanned more than two decades starting with the significant contributions from (Hart et al., 1989). With the use of a single smart meter to measure aggregate consumption of a household, the NILM system provides the opportunity to mine information regarding the consumption details of individual household appliances.

While this approach is non-intrusive, its noticeable benefit to providing consumers with personalized services has become widespread (Batra et al., 2019). With the introduction of clustering-based method using transient and steady-state characteristics (Hart et al., 1989), several approaches have also been studied for energy disaggregation tasks. These include factorial hidden Markov model (FHMM), combinatorial optimization (CO) as well as deep learning algorithms for energy disaggregation (Batra et al., 2019; Kelly and Knottenbelt, 2015). One of the notable challenges in smart grid research is the lack of robust public datasets with different characteristics and significant number of households. Recently, a number of studies have addressed this gap through the simulation of synthetic load data for both aggregate and appliance level (El Kababji and Srikantha, 2020).

2.2 Synthetic Load Data Simulation

Synthetic data simulation for smart grid has been addressed from two major directions, which are model-based and data-driven approaches. Model-based approaches such as the one presented in (Lopez et al., 2018) heavily relied on the underlying physical characteristics of the simulated load. In other words, this approach is highly parametric which limits the flexibility of modeling load consumption of different devices. The limitation is addressed by data-driven approaches which do not make prior assumptions about the physical characteristics of a load. For instance, (El Kababji and Srikantha, 2020) used GAN to generate synthetic smart grid data. However, this approach has implicitly introduced privacy issues in the simulated data since the data were generated from real load consumption with inherent privacy concerns. In our study, we assess the privacy leakage associated with these data-generating frameworks. Particularly, we focus on the framework described by Kababji & Srikantha (El Kababji and Srikantha, 2020).

2.3 Privacy-Preservation of Smart Grid Data

Although there are several approaches for privacy-preserving smart grid data publishing, the most prominent are the studies on data anonymization (Adewole and Torra, 2022a; Sangogboye et al., 2018) and differential privacy (Soykan et al., 2019). Although each study addresses privacy protection of smart grid data from different perspectives, investigation of the privacy leakage of individual appliance signatures in the energy signals is not addressed. As shown in (Adewole and Torra, 2022b), the privacy leakage related to the individual appliance is high. Therefore, one of the objectives of this paper is to investigate the privacy level of the synthetic data. We add additional privacy guarantees considering this type of disclosure risk utilizing two privacy protection mechanisms.

3 PROPOSED APPROACH

Figure 1 shows the proposed framework for smart grid synthetic data generation, disclosure risk assessment and privacy protection. The framework includes three modules to achieve the objectives of the research. The first module (M1) relies on the use of GAN to learn the underlying distribution of load operations from the real appliance datasets that are publicly available (see §3.1). This module is based on the framework described by Kababji & Srikantha (El Kababji and Srikantha, 2020) as previously stated. Module (M2) is responsible for checking the disclosure risk associated with the individual appliances in the simulated households aggregate data. This is achieved through the simulation of three inference attack scenarios that are discussed in §3.2. The third module (M3) applies two privacy protection mechanisms (MDAV and DFTMicroagg) previously studied in our work (Adewole and Torra, 2022a).

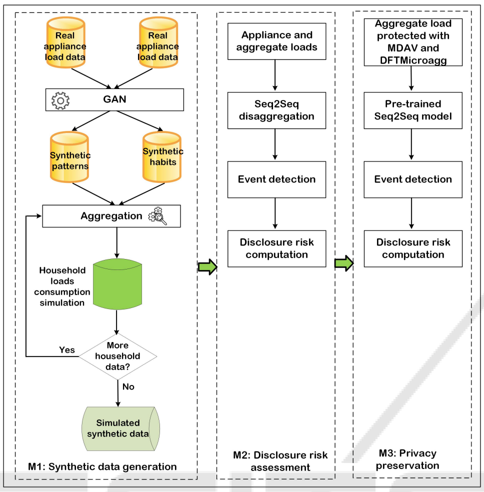

Real appliance load data
Real appliance load data
GAN
Synthetic patterns
Synthetic habits
Aggregation
Household loads consumption simulation
More household data?
Yes
No
Simulated synthetic data
M1: Synthetic data generation
Appliance and aggregate loads
Seq2Seq disaggregation
Event detection
Disclosure risk computation
M2: Disclosure risk assessment
Aggregate load protected with MDAV and DFTMicroagg
Pre-trained Seq2Seq model
Event detection
Disclosure risk computation
M3: Privacy preservation

Figure 1: Proposed framework for smart grid data generation and privacy protection.

3.1 Generative Adversarial Networks

GAN is a widely used algorithm for generating synthetic datasets. This algorithm has two neural networks parts: the generator (G) and the discriminator (D). The generator samples random noise z as input from the probability distribution $p_g(z)$ to produce synthetic patterns. The discriminator can take input from the generator or from the real training data as x drawn from the data distribution $p_d(x)$. Both G and D are trained simultaneously with opposing objectives. G aims to ensure its outputs are indistinguishable by D, that is, to *minimize* the probability of D predicting its output as synthetic. Conversely, the goal of D is to *maximize* the probability of predicting its outcome as either real or synthetic.

In this study, GAN is utilized to generate synthetic load patterns and usage habits based on the cost function in eq. 1. As observed in (El Kababji and Srikantha, 2020), this cost function works better with time series data as opposed to the original GAN cost function that was evaluated on image datasets. During the training phase, the discriminator is updated by ascending its stochastic gradient using eq. 2 and the generator is updated by descending its stochastic gradient using eq. 3. In §4, we provide the detail of the experimental settings for generating synthetic datasets used in this study.

$$C(D,G) = \mathbb{E}_{x\sim p_d(x)}[\log(1-D(x))] +$$

$$\mathbb{E}_{z\sim p_g(z)}[\log D(G(z))] \qquad (1)$$

$$\nabla_{\theta d}\frac{1}{n}\sum_{i=1}^{n}[\log(1-D(x^{(i)}))] + [\log D(G(z^{(i)}))] \qquad (2)$$

$$\nabla_{\theta g}\frac{1}{n}\sum_{i=1}^{n}[\log D(G(z^{(i)}))] \qquad (3)$$

3.2 Inference Attacks Simulation

In this study, we simulate three inference attacks scenarios to check the privacy leakage associated with the individual appliances in the synthetic aggregate data at the household level. Thereafter, we implement two privacy protection mechanisms to ascertain if the privacy risk is reduced for a specific appliance. The three scenarios are briefly discussed as follows:

Scenario 1: Attacker inferring household consumption from the same household in the same dataset. In the first case, we train a Seq2Seq disaggregation algorithm for each appliance using household data. The resulting model was tested with aggregate energy data drawn from the same household.

Scenario 2: Attacker inferring household consumption from different households in the same dataset. In the second case, we train a Seq2Seq disaggregation algorithm for each appliance using particular household data. The resulting model was tested with aggregate energy data selected from another household in the same dataset. This is to check if the attacker can use a model trained with a particular household data to attack another household from the same dataset.

Scenario 3: Attacker inferring household consumption from different households in different datasets. In the third scenario, we train a Seq2Seq disaggregation algorithm for each appliance using particular household data in one dataset. The resulting model was tested with aggregate energy data selected from another household in a different dataset. This is to check if the attacker can use a trained model on one household to attack another household from a different dataset.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

3.3 Seq2Seq Algorithm

Seq2Seq (Kelly and Knottenbelt, 2015) is a deep learning NILM algorithm based on different convolutional neural networks layers and a fully connected layer. It takes a sequence of aggregate energy data $Y_{t:t+W-1}$ and maps it to another sequence corresponding to the target appliance load signature $X_{t:t+W-1}$. The regression is defined as $X_{t:t+W-1} = F_s(Y_{t:t+W-1}, \theta_s) + \varepsilon$, where ε is W-dimensional Gaussian random noise and θ_s are the parameters of the neural networks.

3.4 Event Detection

To compute the activation period of a specific appliance at time t from the output of Seq2Seq algorithm, we adopt three event detection methods. These methods are utilized to compute the activation threshold λ that determines when a particular appliance is switched ON at time t. Any value that is less than λ signifies the OFF state.

3.4.1 Activation Time Extraction

Activation Time Extraction (ATE) is an event detection method proposed by (Kelly and Knottenbelt, 2015). It was specifically tunned on UK-DALE NILM dataset. The method extracts activations of appliances such as light bubs and toasters by considering consecutive activations above a certain threshold. The algorithm is included in NILMTK (Batra et al., 2019), which is a toolkit for energy disaggregation.

3.4.2 Middle-Point Thresholding

Middle-Point Thresholding (MPT) (Precioso and Gomez-Ullate, 2020) computes the threshold λ for each appliance ℓ from the distribution of power measurement of individual appliances in the training data based on clustering analysis. The training data are split into two clusters and the centroids of each cluster are utilized to fix the threshold value according to eq. 4. The first centroid is denoted as $m_0^{(\ell)}$ representing the OFF state and the second centroid is $m_1^{(\ell)}$ for ON event. These two values are used to fix the event detection threshold in the case of MPT. In this paper, we used K-means algorithm for the clustering.

$$\lambda^{(\ell)} = \frac{m_0^{(\ell)} + m_1^{(\ell)}}{2} \tag{4}$$

3.4.3 Variance-Sensitive Thresholding

This method improves on MPT method by introducing standard deviation $\sigma_k^{(\ell)}$ estimated from the data points in each cluster according to eq 5:

$$d = \frac{\sigma_0^{(\ell)}}{\sigma_0^{(\ell)} + \sigma_1^{(\ell)}}$$

$$\lambda^{(\ell)} = (1-d)m_0^{(\ell)} + dm_1^{(\ell)} \tag{5}$$

3.5 Privacy Protection Algorithms

We investigate the performance of two privacy preserving mechanisms: Maximum Distance to Average Vector (MDAV) microaggregation (Domingo-Ferrer and Torra, 2005; Samarati, 2001) and DFTMicroagg (Adewole and Torra, 2022a). These two algorithms have been studied in our recent work (Adewole and Torra, 2022a) to protect daily energy consumption data. DFTMicroagg combines MDAV and discrete Fourier transform (DFT) to provide an additional level of privacy protection. In this paper, we investigate the performance of these algorithms for protecting the privacy of individual appliance signatures in synthetic energy signals.

4 EXPERIMENTAL SETUP

All experiments have been conducted using Python. We extend the framework in (El Kababji and Srikantha, 2020) for synthetic dataset simulation. Energy disaggregation experiment is based on NILMTK and NILKTK-Contrib (Batra et al., 2019) environment.

4.1 Synthetic Dataset Simulation

In this paper, we simulate two synthetic datasets, named *Synthetic AMPd* and *Synthetic REFIT*. The first dataset (Synthetic AMPd) contains 200 households load consumption data with seven appliances (Cloth Dryer (CDE), Dishwasher (DWE), Fridge (FRE), Heat Pump (HPE), Cloth Washer (CWE), Instant hot water unit (HTE), and Kitchen wall oven (WOE)). The data was simulated from two publicly available datasets, *Almanac of Minutely Power dataset (AMPd)* and *Rainforest Automation Energy dataset (RAE)* using GAN. We combine the synthetic patterns and usage habits generated by GAN into aggregate-level data for household consumption. The generated synthetic dataset covers a period of one year spanning from January 2021 to December 2021 with 3 minutes granularity.

The second dataset (Synthetic REFIT) contains 200 households load consumption with five appliances (Washing machine (WME/CWE), Dishwasher

(DWE), Fridge (FRE), Microwave (MWE) and Kettle (KTE)). The data was simulated from REFIT public dataset using GAN. REFIT was collected from 20 households with a sampling interval of 8 seconds. The generated synthetic dataset covers a period of two years spanning from January 2020 to December 2021 with 6 minutes granularity.

4.2 Training and Testing

For the privacy protection experiment, we check when k is 40 and 50 for microaggregation. The coefficient values of the DFTMicroagg algorithm used for the Synthetic AMPd dataset is 80 while that of Synthetic REFIT is 40 (see (Adewole and Torra, 2022a) for details). Table 1 shows the experimental settings used to train and test Seq2Seq disaggregation algorithm during privacy evaluation of the two synthetic datasets.

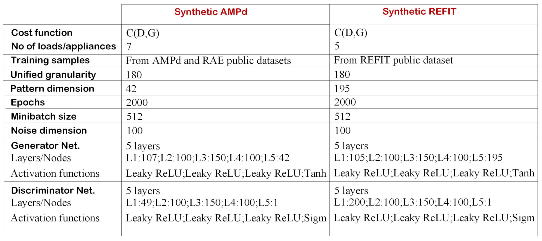
Table 1: Experimental settings for the two datasets.

Dataset	No of Appliances	Training Periods	Testing Periods
Synthetic AMPd dataset	7	8 months	4 months
Synthetic REFIT dataset	5	17 months	7 months

4.3 Evaluation

4.3.1 GAN Evaluation

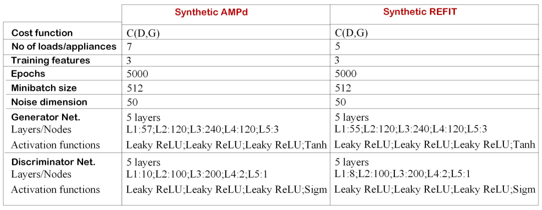
Figure 2 shows the architecture of the GAN system used for load pattern synthesis for the two datasets discussed in §4.1. As an extension to the architecture in (El Kababji and Srikantha, 2020), this paper simulated seven loads for Synthetic AMPd and five loads for Synthetic REFIT datasets. Figure 3 presents the architecture of GAN system used for the load habit synthesis for the two datasets.



	Synthetic AMPd	Synthetic REFIT
Cost function	C(D,G)	C(D,G)
No of loads/appliances	7	5
Training samples	From AMPd and RAE public datasets	From REFIT public dataset
Unified granularity	180	180
Pattern dimension	42	195
Epochs	2000	2000
Minibatch size	512	512
Noise dimension	100	100
Generator Net. Layers/Nodes Activation functions	5 layers L1:107;L2:100;L3:150;L4:100;L5:42 Leaky ReLU;Leaky ReLU;Leaky ReLU;Tanh	5 layers L1:105;L2:100;L3:150;L4:100;L5:195 Leaky ReLU;Leaky ReLU;Leaky ReLU;Tanh
Discriminator Net. Layers/Nodes Activation functions	5 layers L1:49;L2:100;L3:150;L4:100;L5:1 Leaky ReLU;Leaky ReLU;Leaky ReLU;Sigm	5 layers L1:200;L2:100;L3:150;L4:100;L5:1 Leaky ReLU;Leaky ReLU;Leaky ReLU;Sigm

Figure 2: GAN architecture for load patterns synthesis for the two datasets.

To evaluate the performance of GAN system for load pattern and habit synthesis based on the architecture in Figures 2 and 3, we train a neural network with multilayer perceptron as described in (El Kababji and Srikantha, 2020). This provides a background evaluation of the accuracy achieved by the proposed GAN systems. Section §5.1 shows the evaluation results.



	Synthetic AMPd	Synthetic REFIT
Cost function	C(D,G)	C(D,G)
No of loads/appliances	7	5
Training features	3	3
Epochs	5000	5000
Minibatch size	512	512
Noise dimension	50	50
Generator Net. Layers/Nodes Activation functions	5 layers L1:57;L2:120;L3:240;L4:120;L5:3 Leaky ReLU;Leaky ReLU;Leaky ReLU;Tanh	5 layers L1:55;L2:120;L3:240;L4:120;L5:3 Leaky ReLU;Leaky ReLU;Leaky ReLU;Tanh
Discriminator Net. Layers/Nodes Activation functions	5 layers L1:10;L2:100;L3:200;L4:2;L5:1 Leaky ReLU;Leaky ReLU;Leaky ReLU;Sigm	5 layers L1:8;L2:100;L3:200;L4:2;L5:1 Leaky ReLU;Leaky ReLU;Leaky ReLU;Sigm

Figure 3: GAN architecture for load habits synthesis for the two datasets.

4.3.2 Disclosure Risk Evaluation

To check the disclosure risk associated with each appliance in the energy data, we adopt the formula proposed in our recent study (Adewole and Torra, 2022b). The disclosure risk (DR) is given as,

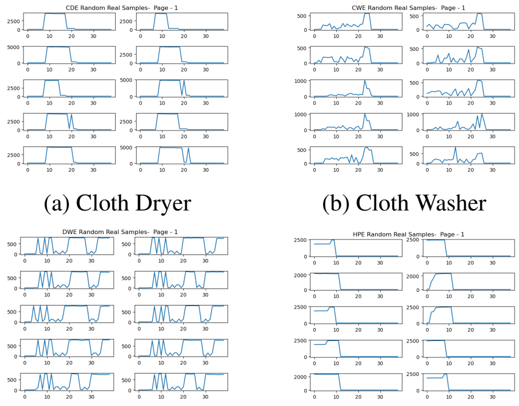
$$DR^{(\ell)} = TP^{(\ell)} / (TP^{(\ell)} + FN^{(\ell)}) \qquad (6)$$

where $TP^{(\ell)}$ represents the proportion of correctly predicted ON events for appliance ℓ, $FN^{(\ell)}$ represents the proportion of ON events for appliance ℓ which was mistakenly predicted as OFF events, and $DR^{(\ell)}$ represents the disaggregation risk of appliance ℓ that takes a value in the interval [0,1]. The higher the value of DR, the higher the disclosure risk.

5 RESULTS AND DISCUSSION

5.1 Synthetic Datasets

It can be seen in Figures 4 and 5 that GAN was able to model the real patterns for the individual loads. These figures show the sample results for the simulated Synthetic AMPd loads patterns for some appliances due to the space constraint.



(a) Cloth Dryer

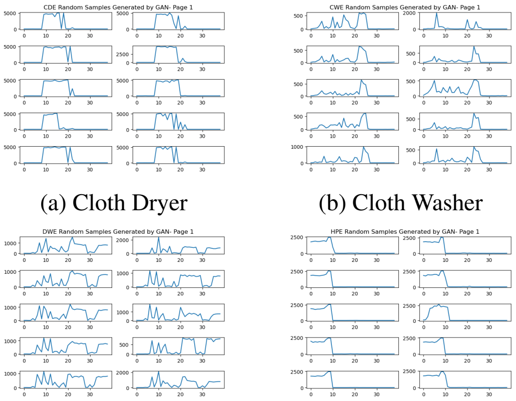
(b) Cloth Washer

(c) Dishwasher

(d) Heat Pump

Figure 4: Sample random real patterns of appliances in Synthetic AMPd.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS



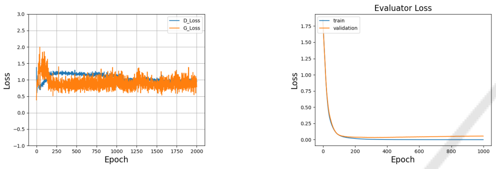
(a) Cloth Dryer

(b) Cloth Washer

(c) Dishwasher

(d) Heat Pump

Figure 5: Sample random synthetic patterns of appliances in Synthetic AMPd.



(a) GAN Evolution

(b) EvaluatorNet loss

Figure 6: Performance evaluation of pattern generation for the seven loads in Synthetic AMPd.

Figure 6(a) shows the evolution of Generator and Discriminator during the training phase of GAN. It can be seen that the two networks show good convergence patterns during the training phase. Figure 6(b) depicts the loss of the neural network that evaluated the generated patterns.

5.2 Attack Scenario 1

5.2.1 Results Based on Synthetic AMPd Dataset

For this inference attack scenario, House 2 data were used to train and test the Seq2Seq NILM algorithm. The goal is to ascertain if the NILM disaggregation algorithm can effectively disaggregate the signature of each appliance. The result of this energy disaggregation represents a disclosure risk that reveals the level in which attackers can predict the lifestyles of the individual households. Table 2 and 3 show the disaggregation risk associated with Synthetic AMPd dataset. It can be seen that except for Fridge, Washing machine and Electric water heater, the disaggregation risk of the other devices is on the high side. To reduce the risk associated with individual appliances, we utilized MDAV and DFTMicroagg algorithms. Based on our findings, DFTMicroagg lowers the disclosure risk when compared with MDAV especially, when the value of $K = 50$ according to Table 3. These results confirmed that both MDAV and DFTMicroagg are promising privacy-preserving mechanisms for smart grid data protection and for reducing the disaggregation risk associated with individual appliances. Thus, the privacy leakage of the individual household lifestyles is reduced.

Table 2: Disaggregation risk for each appliance in Synthetic AMPd dataset for Scenario 1. A = Original, B = MDAV, C = DFTMicroagg. K = 40 and Coeff = 80.

	Disaggregation risk - Synthetic AMPd								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
HPE	0.90	0.03	0.01	0.96	0.00	0.00	0.92	0.02	0.01
DWE	0.99	0.57	0.43	0.90	0.11	0.03	0.98	0.44	0.27
CDE	0.99	0.00	0.00	0.99	0.00	0.00	0.98	0.00	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WOE	0.97	0.08	0.03	0.84	0.00	0.00	0.97	0.07	0.00
HTE	0.01	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.00

Table 3: Disaggregation risk for each appliance in Synthetic AMPd dataset for Scenario 1. A = Original, B = MDAV, C = DFTMicroagg. K = 50 and Coeff = 80.

	Disaggregation risk - Synthetic AMPd								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
HPE	0.90	0.04	0.03	0.96	0.00	0.00	0.92	0.04	0.04
DWE	0.99	0.46	0.25	0.90	0.03	0.04	0.98	0.27	0.13
CDE	0.99	0.00	0.00	0.99	0.00	0.00	0.98	0.00	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WOE	0.97	0.17	0.06	0.84	0.00	0.00	0.97	0.17	0.00
HTE	0.01	0.01	0.01	0.00	0.00	0.01	0.06	0.03	0.00

5.2.2 Results Based on Synthetic REFIT Dataset

Similarly, we use House 2 data in Synthetic REFIT to train and test Seq2Seq NILM algorithm for this scenario. Tables 4 and 5 show the disaggregation risk associated with this dataset and the results we obtained for the two privacy protection mechanisms. Similar to the results obtained with Synthetic AMPd, DFTMicroagg outperformed MDAV on average, especially, when the value of K was increased to 50. This result offers better privacy protection than $K = 40$.

Table 4: Disaggregation risk for each appliance in Synthetic REFIT dataset for Scenario 1. A = Original, B = MDAV, C = DFTMicroagg. K = 40 and Coeff = 40.

	Disaggregation risk - Synthetic REFIT								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
CWE	0.31	0.14	0.12	0.79	0.18	0.17	0.61	0.20	0.18
DWE	0.74	0.22	0.13	0.87	0.09	0.05	0.88	0.09	0.07
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KTE	0.93	0.19	0.00	0.97	0.19	0.00	0.98	0.37	0.00

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 5: Disaggregation risk for each appliance in Synthetic REFIT dataset for Scenario 1. A = Original, B = MDAV, C = DFTMicroagg. K = 50 and Coeff = 40.

	Disaggregation risk - Synthetic REFIT								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
CWE	0.31	0.10	0.10	0.79	0.11	0.12	0.61	0.12	0.13
DWE	0.74	0.15	0.10	0.87	0.00	0.00	0.88	0.00	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KTE	0.93	0.07	0.00	0.97	0.19	0.00	0.98	0.34	0.00

5.3 Attack Scenario 2

5.3.1 Results Based on Synthetic AMPd Dataset

For this experiment, we trained Seq2Seq NILM algorithm with data from households 1 and 2, and the model was tested using households 41 and 42 data. The results as shown in Tables 6 and 7 reveal that Seq2Seq algorithm and the activation extraction methods can successfully disaggregate Heat pump, Dish washer, Cloth dryer, Oven and Electric water heater for Scenario 2. As opposed the result obtained in Scenario 1, when Seq2Seq was trained and tested with more data from different households, the disaggregation risk of Electric water heater increased. This shows that a pre-trained model from a particular household data can be used to reveal the lifestyles of another household.

We further check if the proposed privacy-preserving mechanisms can lower these disaggregation risks. The results obtained show that DFTMicroagg still outperformed MDAV on average (see Tables 6 and 7) for Synthetic AMPd dataset. This pattern of result can also be seen in Tables 8 and 9 for Synthetic REFIT. Both MDAV and DFTMicroagg provide promising privacy protection guarantees.

Table 6: Disaggregation risk for each appliance in Synthetic AMPd dataset for Scenario 2. A = Original, B = MDAV, C = DFTMicroagg. K = 40 and Coeff = 80.

	Disaggregation risk - Synthetic AMPd								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
HPE	0.86	0.01	0.00	0.98	0.00	0.00	0.95	0.01	0.00
DWE	0.99	0.49	0.35	0.91	0.05	0.09	0.99	0.31	0.22
CDE	0.99	0.00	0.00	0.99	0.00	0.00	0.98	0.00	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WOE	0.97	0.00	0.00	0.88	0.00	0.00	0.97	0.00	0.00
HTE	0.48	0.20	0.15	0.37	0.05	0.04	0.75	0.53	0.35

5.3.2 Results Based on Synthetic REFIT Dataset

Tables 8 and 9 show the disaggregation risk associated with this dataset and how the two privacy protection algorithms reduced this risk.

Table 7: Disaggregation risk for each appliance in Synthetic AMPd dataset for Scenario 2. A = Original, B = MDAV, C = DFTMicroagg. K = 50 and Coeff = 80.

	Disaggregation risk - Synthetic AMPd								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
HPE	0.86	0.00	0.00	0.98	0.00	0.00	0.95	0.00	0.00
DWE	0.99	0.30	0.25	0.91	0.03	0.02	0.99	0.13	0.09
CDE	0.99	0.00	0.00	0.99	0.00	0.00	0.98	0.00	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WOE	0.97	0.18	0.00	0.88	0.00	0.00	0.97	0.16	0.00
HTE	0.48	0.19	0.10	0.37	0.02	0.01	0.75	0.28	0.21

Table 8: Disaggregation risk for each appliance in Synthetic REFIT dataset for Scenario 2. A = Original, B = MDAV, C = DFTMicroagg. K = 40 and Coeff = 40.

	Disaggregation risk - Synthetic REFIT								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
CWE	0.33	0.06	0.06	0.78	0.09	0.05	0.64	0.07	0.04
DWE	0.41	0.12	0.10	0.94	0.11	0.09	0.80	0.14	0.11
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KTE	0.95	0.13	0.00	1.00	0.13	0.00	0.97	0.22	0.00

5.4 Attack Scenario 3

For this scenario, Households 1 and 2 data from Synthetic REFIT were used for training. This covers a period of 17 months. The testing was done using Household 1 and 2 data from Synthetic AMPd. Test data covers a period of 8 months. The experiment was carried out using only Washing Machine, Dish Washer and Fridge since they are the common appliances in the two simulated synthetic datasets. The disaggregation risk for this scenario is minimal since only the dish washer was disaggregated based on ATE and VST activation methods. Nevertheless, the proposed privacy-preserving methods further lower this risk, particularly when $K = 50$. Similarly, DFTMicroagg outperformed MDAV based on the results obtained. See Tables 10 and 11.

Table 9: Disaggregation risk for each appliance in Synthetic REFIT dataset for Scenario 2. A = Original, B = MDAV, C = DFTMicroagg. K = 50 and Coeff = 40.

	Disaggregation risk - Synthetic REFIT								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
CWE	0.33	0.10	0.09	0.78	0.05	0.04	0.64	0.11	0.10
DWE	0.41	0.04	0.08	0.94	0.00	0.00	0.80	0.00	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KTE	0.95	0.04	0.00	1.00	0.09	0.00	0.97	0.24	0.00

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 10: Disaggregation risk of the common appliances in the two synthetic datasets for Scenario 3. A = Original, B = MDAV, C = DFTMicroagg. K = 40 and Coeff = 80.

	DR - Synthetic REFIT and AMPd								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
CWE	0.06	0.09	0.02	0.00	0.07	0.00	0.07	0.10	0.03
DWE	0.21	0.11	0.10	0.07	0.03	0.01	0.15	0.06	0.05
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 11: Disaggregation risk of the common appliances in the two synthetic datasets for Scenario 3. A = Original, B = MDAV, C = DFTMicroagg. K = 50 and Coeff = 80.

	DR - Synthetic REFIT and AMPd								
	ATE			MPT			VST		
Load	A	B	C	A	B	C	A	B	C
CWE	0.06	0.02	0.00	0.00	0.00	0.00	0.07	0.02	0.00
DWE	0.21	0.07	0.02	0.07	0.01	0.00	0.15	0.05	0.00
FRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

6 CONCLUSIONS

In this paper, we assessed the privacy levels of synthetic households aggregate smart grid data. We investigated the performance of Seq2Seq energy disaggregation algorithm and three activation extraction methods. The findings revealed that the disclosure risk associated with a significant number of appliances in the synthetic aggregate data is high. Thereafter, we proposed two privacy preserving approaches to lower this disclosure risk. The results show that the two privacy protection methods produced promising results for privacy protection of individual household lifestyles. In future, we would like to investigate the privacy leakage at the top level of the grid hierarchy. This will enable us to have a better understanding of the privacy protection offer by the proposed mechanisms at different levels of smart grid hierarchy.

ACKNOWLEDGEMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. The first author is supported by the Kempe foundation.

REFERENCES

Adewole, K. S. and Torra, V. (2022a). Dftmicroagg: a dual-level anonymization algorithm for smart grid data. *International Journal of Information Security*, pages 1–23.

Adewole, K. S. and Torra, V. (2022b). Privacy issues in smart grid data: From energy disaggregation to disclosure risk. In *International Conference on Database and Expert Systems Applications*, pages 71–84. Springer.

Batra, N., Kukunuri, R., Pandey, A., Malakar, R., Kumar, R., Krystalakos, O., Zhong, M., Meira, P., and Parson, O. (2019). Towards reproducible state-of-the-art energy disaggregation. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation*, pages 193–202. ACM.

BBCNews (2017). Ukraine power cut 'was cyber-attack'. https://www.bbc.com/news/technology-38573074. Accessed: 18th May 2021.

Domingo-Ferrer, J. and Torra, V. (2005). Ordinal, continuous and heterogeneous k-anonymity through microaggregation. *Data Mining and Knowledge Discovery*, 11(2):195–212.

El Kababji, S. and Srikantha, P. (2020). A data-driven approach for generating synthetic load patterns and usage habits. *IEEE Transactions on Smart Grid*, 11(6):4984–4995.

Fekri, M. N., Ghosh, A. M., and Grolinger, K. (2019). Generating energy data for machine learning with recurrent generative adversarial networks. *Energies*, 13(1):130.

Hart, G. W., Kern Jr, E. C., and Schweppe, F. C. (1989). Non-intrusive appliance monitor apparatus. US Patent 4,858,141.

Kelly, J. and Knottenbelt, W. (2015). Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*, pages 55–64. ACM.

Lopez, J. M. G., Pouresmaeil, E., Canizares, C. A., Bhattacharya, K., Mosaddegh, A., and Solanki, B. V. (2018). Smart residential load simulator for energy management in smart grids. *IEEE Transactions on Industrial Electronics*, 66(2):1443–1452.

Precioso, D. and Gomez-Ullate, D. (2020). Nilm as a regression versus classification problem: the importance of thresholding. *arXiv preprint arXiv:2010.16050*.

Samarati, P. (2001). Protecting respondents identities in microdata release. *IEEE transactions on Knowledge and Data Engineering*, 13(6):1010–1027.

Sangogboye, F. C., Jia, R., Hong, T., Spanos, C., and Kjærgaard, M. B. (2018). A framework for privacy-preserving data publishing with enhanced utility for cyber-physical systems. *ACM Transactions on Sensor Networks (TOSN)*, 14(3-4):1–22.

Soykan, E. U., Bilgin, Z., Ersoy, M. A., and Tomur, E. (2019). Differentially private deep learning for load forecasting on smart grid. In *2019 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS