LEARNING DEFENSIVE POLICIES AGAINST DIVERSE INFERENCE ATTACKS FOR SMART METER PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

Smart meter (SM) data provides fine-grained visibility into household energy consumption, but also exposes users to privacy risks. Inference attacks, known as non-intrusive load monitoring (NILM), can perform appliance-level inference from aggregate signals and recover sensitive behavioral patterns. In practice, attacker models are unknown and heterogeneous, making robust defense challenging. We formulate SM privacy protection as a black-box inference defense problem, aiming to reduce the recoverability of appliance-level information while generalizing across diverse and unseen attackers. We propose a proxy-guided hierarchical reinforcement learning framework that learns battery-based load-shaping policies to inject realistic but misleading appliance-level signatures into the aggregate signal, thereby disrupting the structured patterns exploited by NILM. A self-supervised aggregate-structure privacy probe provides a reconstruction-error-based surrogate reward for disrupting recoverable load structure, while a signature library makes the perturbations appliance-relevant and physically realizable through battery control. We provide theoretical rationale showing that proxy-guided optimization improves inference robustness under attacker diversity. Experiments on real-world datasets UK-DALE and REDD demonstrate strong cross-model and cross-appliance generalization. Across six unseen NILM attackers, covering four appliances on UK-DALE and five on REDD, our proposed defense increases average appliance-level RMSE by 107% and 166%, respectively, while reducing F1 score by 79% and 80%.

1 INTRODUCTION

Smart meters (SMs) record high-frequency, real-time electricity consumption time series, where aggregate load measurements encode structured appliance-level usage patterns that can be exploited by non-intrusive load monitoring (NILM) models (Hart, 1992). Prior work Kolter & Jaakkola (2012); Kelly & Knottenbelt (2015a); Chen et al. (2018); Zhang et al. (2018); Yue et al. (2020); Sykiotis et al. (2022) has shown that diverse NILM methods can recover fine-grained appliance activities from aggregate signals without direct access to individual devices. Such appliance-level information reveals latent household behaviors and therefore raises serious privacy concerns.

Existing SM privacy defenses largely optimize proxy objectives not explicitly aligned with the NILM disaggregation task—encouraging curve smoothness Erdemir et al. (2020); Zhang et al. (2024b), reducing aggregate variance Li et al. (2023); Zhang et al. (2024a), other load-level proxies Farokhi & Sandberg (2019); Farokhi (2020), or information-theoretic criteria Shateri et al. (2021); Shateri (2023). These reshape aggregate trajectories without explicitly targeting the inference structure NILM attackers exploit, so their effect on appliance-level inference is only indirectly controlled, creating a privacy–task mismatch. The key challenge is therefore to learn defender policies that disrupt NILM-relevant load structures while remaining robust to diverse, unseen attackers.

To address this challenge, we propose a proxy-guided appliance-signature mimicry framework combining aggregate-structure disruption with a NILM-relevant signature-level intervention mechanism. We use the reconstruction error of a self-supervised aggregate-level sequence probe—requiring no appliance-level labels or attacker access—as the training signal, and instead of injecting arbitrary perturbations, the defender inserts realistic appliance signatures from a curated library through feasible battery control. We formulate SM privacy protection as a black-box inference defense problem under attacker uncertainty, and learn a hierarchical reinforcement learning (HRL) policy

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

that decides when and which signature to mimic, while a low-level battery executor realizes it under physical constraints, reshaping the reported load to reduce the recoverability of true appliance-level consumption. Experiments on UK-DALE and REDD show that the resulting framework consistently degrades NILM reconstruction performance across diverse unseen attackers and remains effective across datasets, households, and seasonal conditions.

Contributions. (1) A privacy defense that injects realistic appliance signatures into aggregate load via a curated signature library to mask appliance-level consumption. (2) A proxy-guided HRL framework learning a battery control policy from an aggregate-level privacy probe signal. (3) A black-box inference-defense formulation under attacker uncertainty, with a theoretical justification via the unified task formulation of NILM inference and the transferability of surrogate-based attacks. (4) Empirical evidence of consistent degradation across multiple NILM models, datasets, and unseen attackers, with strong generalization across households and seasonal conditions.

2 Related work

NILM as inference threat. NILM methods recover appliance-level consumption from aggregate SM signals, ranging from combinatorial optimization Hart (1992) and factorial HMMs Kolter & Jaakkola (2012) to convolutional, recurrent, and transformer-based sequence models Kelly & Knottenbelt (2015a); Zhang et al. (2018); Yue et al. (2020); Sykiotis et al. (2022); Petralia et al. (2025). Despite this architectural diversity, recent surveys Rafiq et al. (2024); Kaselimi et al. (2022) note that modern NILM models share a common task formulation: estimating appliance-level power or on/off state from a window of aggregate readings, evaluated under a shared family of reconstruction losses. This task-level commonality, combined with architectural diversity, motivates the black-box defense setting in this work, where the defender does not know the deployed NILM architecture but can exploit the shared structure of the inference task.

SM privacy defenses. Existing SM privacy approaches fall into data-centric and physical-layer categories. Data-centric mechanisms such as aggregation Gomez-Vilardebo & Gündüz (2014) and differential privacy Zhang et al. (2025) perturb or post-process reported readings, but implicitly assume that adversaries cannot tap the physical power line, which limits their applicability under realistic deployment. Physical-layer approaches reshape household consumption with a rechargeable battery before measurement Erdemir et al. (2019); Shateri et al. (2020), providing local and real-time protection against attackers that exploit the structure of real load trajectories. Early battery-based methods rely on rule-based or heuristic strategies Erdemir et al. (2019); Farokhi & Sandberg (2019); Farokhi (2020), while more recent work introduces RL to automate battery scheduling Shateri et al. (2020; 2021); Erdemir et al. (2020); Li et al. (2023); Shateri (2023); Zhang et al. (2024a;b). These RL-based methods typically optimize either load-level proxies such as smoothness Erdemir et al. (2020); Zhang et al. (2024b) or variance reduction Li et al. (2023); Zhang et al. (2024a), or information-theoretic measures such as mutual information and information leakage Shateri et al. (2021); Shateri (2023). Load-level proxies do not explicitly model how attackers exploit appliance-level temporal patterns, leading to a privacy–inference gap when attackers rely on different representations. Information-theoretic objectives are conceptually principled but cannot be directly attributed to individual control actions, and existing schemes therefore rely on auxiliary estimators without explicitly optimizing transfer to unseen attackers.

Adversarial robustness and transferability. Our formulation also connects to transfer-based black-box adversarial attacks, where perturbations crafted on a surrogate model transfer to unseen targets (Papernot et al., 2017; Liu et al., 2016; Gu et al., 2023; Demontis et al., 2019), including in time-series regression (Mode & Hoque, 2020; Tokgoz et al., 2026). This motivates using a surrogate privacy probe to measure load-structure recoverability and guide the learning of transferable, physically constrained battery perturbations structured as appliance-signature mimicry.

3 Problem Formulation

We consider a black-box SM privacy setting in which the defender does not know the deployed NILM attacker (see Fig. 1). At time t, let $P_{\text{user}}(t)$ denote the original household aggregate load and $P_{\text{batt}}(t)$ the battery power controlled by the defender (> 0: charging, < 0: discharging),

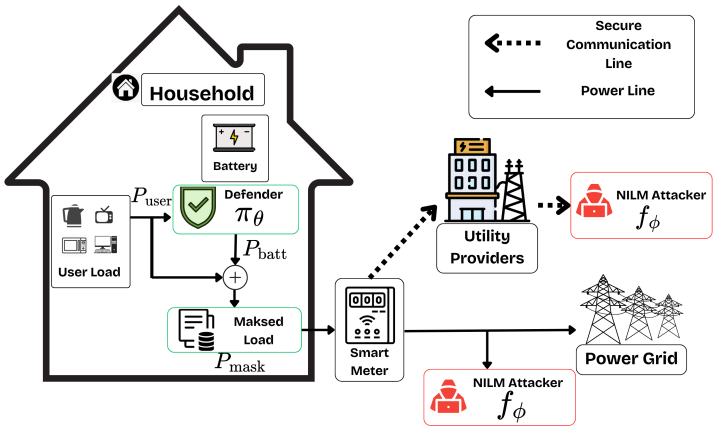

Figure 1: Defender π_θ injects battery-driven mimic signatures P_{batt} into user load signal P_{user}, aiming to deceive an NILM attacker f_ϕ who attempts to infer appliance patterns from aggregated SM data.

subject to standard battery feasibility constraints (Appendix A.2). The smart meter reports the masked aggregate load $P^{\pi_\theta}_{\text{mask}}(t) = P_{\text{user}}(t) + P_{\text{batt}}(t)$. For window length τ, the clean aggregate window is $\mathbf{x}^0_t = [P_{\text{user}}(t-\tau+1), \ldots, P_{\text{user}}(t)]$, the masked window induced by policy π_θ is $\mathbf{x}^{\pi_\theta}_t = [P^{\pi_\theta}_{\text{mask}}(t-\tau+1), \ldots, P^{\pi_\theta}_{\text{mask}}(t)]$, and $\mathbf{y}_t$ denotes the appliance-level target in the same window.

An NILM attacker f_ϕ maps the reported aggregate to appliance-level predictions: $\hat{\mathbf{y}}_t = f_\phi(\mathbf{x}^{\pi_\theta}_t)$. For a given attacker f_ϕ and defender policy π_θ, we define the attacker's inference risk as

$$R(f_\phi, \pi_\theta) = \mathbb{E}_t \left[\mathcal{L}\left(f_\phi(\mathbf{x}^{\pi_\theta}_t), \mathbf{y}_t \right) \right], \tag{1}$$

where $\mathcal{L}(\cdot)$ is appliance-level reconstruction loss (RMSE/MAE/SAE), with F1 reduction reported for activation-state inference; a larger $R(f_\phi, \pi_\theta)$ means stronger privacy protection. The deployed attacker is unknown, so we treat it as drawn from a heterogeneous family $f_\phi \in \mathcal{F}$ spanning optimization-based, probabilistic, and neural sequence models, and maximize the expected inference risk over this family:

$$\max_\theta \quad \mathbb{E}_{f_\phi \sim \mu} \left[R(f_\phi, \pi_\theta) \right], \tag{2}$$

where μ denotes an unknown distribution over possible NILM attackers. Direct optimization of Eq. (2) is infeasible because the deployed attacker and the distribution μ are not accessible during training. We therefore introduce a self-supervised aggregate-structure privacy probe in Section 4 to provide a proxy learning signal.

4 Theoretical Rationale

The black-box defense problem in Section 3 is a physically constrained adversarial attack against NILM inference models, where the perturbation set Δ_{batt} is determined by feasible battery actions rather than norm-bounded digital noise. Direct optimization of the robust objective $\mathcal{R}_{\text{avg}}(\pi_\theta) = \mathbb{E}_{f_\phi \sim \mu}[R(f_\phi, \pi_\theta)]$ (Eq. (2)) is infeasible, so we develop a tractable surrogate resting on two design choices: an aggregate-level privacy probe as the training signal, and signature mimic as the perturbation mechanism.

Proxy-guided adversarial optimization. We introduce an aggregate-level privacy probe f_η, trained via identity self-supervision to reconstruct a clean aggregate window $\mathbf{x}^0$ and kept fixed during defender training. The proxy inference risk is

$$R_p(\pi_\theta) = \mathbb{E}_t \left[\mathcal{L}\left(f_\eta(\mathbf{x}^{\pi_\theta}_t), \mathbf{x}^0_t \right) \right], \tag{3}$$

and the defender is trained to maximize $R_p(\pi_\theta)$: a larger value means the masked window deviates further from the clean aggregate patterns the probe was trained on, i.e., the defender disrupts aggregate-level temporal regularities rather than directly estimating appliance-level NILM error. $R_p(\pi_\theta)$

substitutes for $\mathcal{R}_{\text{avg}}(\pi_\theta)$ by replacing the appliance-level target $\mathbf{y}_t$ with the label-free aggregate $\mathbf{x}_t^0$ and the unknown attacker distribution μ with the fixed probe f_η, on the premise that NILM attackers share a common dependence on aggregate temporal patterns.

Task commonality across NILM attackers. NILM attackers differ in formulation—optimization-based (Hart, 1992), probabilistic (Kolter & Jaakkola, 2012), neural (Kelly & Knottenbelt, 2015a)—but share a premise: appliance outputs are inferred from aggregates assumed consistent with training-time aggregate–appliance co-occurrence statistics, and inference degrades once the reported aggregate is pushed outside that region. Our probe, reconstructing clean aggregates from the same distribution, summarizes these statistics, so a large $R_p(\pi_\theta)$ certifies the masked window has left this region—without the probe itself performing appliance-level inference—which is expected to raise inference risk broadly across $\mathcal{F}$, consistent with the surrogate-transfer principle discussed in Section 2 (Papernot et al., 2017; Demontis et al., 2019; Mode & Hoque, 2020). The structural condition under which this transfer provably increases $\mathcal{R}_{\text{avg}}$ is given in Appendix B.

Signature mimic as a task-aligned perturbation. Since NILM attackers recover appliance-level consumption from aggregate load, an effective perturbation interferes with the appliance-level evidence used for disaggregation rather than merely adding aggregate noise. We construct perturbations from a library $\mathcal{S}$ of real appliance signatures:

$$\mathbf{x}_t^{\pi_\theta} = \mathbf{x}_t^0 + \delta_{\text{sig}}, \qquad \delta_{\text{sig}} \in \Delta_{\text{sig}}, \tag{4}$$

where δ_{sig} is induced by feasible battery actions that replay a selected signature trajectory.

The perturbation is therefore physically realizable while retaining appliance-like temporal structure. By restricting the policy to signature-induced perturbations, the framework converts aggregate-structure disruption into plausible but misleading appliance-level evidence, thereby creating ambiguous disaggregation explanations for NILM attackers. This view is consistent with a broader trend in time-series adversarial attacks: informed, selective perturbations that exploit task structure outperform unstructured ones under fixed budgets (Tokgoz et al., 2026). In SM data privacy tasks, signature-based interventions provide a structured perturbation class better matched to the NILM inference task. The next section instantiates these principles in a hierarchical RL framework operationalizing proxy-guided signature mimic under realistic battery constraints.

5 Proxy-Guided HRL Policy Training

We instantiate the proxy-guided signature mimicry principle of Section 4 as a hierarchical RL framework: the fixed privacy probe provides the reward signal, the signature library defines the high-level action space, the defender policy π_θ decides when and which signature to mimic, and all battery actions execute under hard physical feasibility constraints.

Proxy-based privacy reward. We optimize $R_p(\pi_\theta)$ with PPO by deriving a per-step privacy reward from its instantaneous form. Since the probe f_η has nonzero residual reconstruction error even on clean aggregate windows in practice, we measure the defender's contribution as the *increase* in probe error caused by signature insertion, relative to the probe's residual on the unmodified window $\mathbf{x}_t^0$:

$$r_t^{\text{priv}} = \mathcal{L}\left(f_\eta(\mathbf{x}_t^{\pi_\theta}), \mathbf{x}_t^0\right) - \mathcal{L}\left(f_\eta(\mathbf{x}_t^0), \mathbf{x}_t^0\right). \tag{5}$$

Thus, the policy receives positive reward when the selected signature insertion makes the reported aggregate harder to reconstruct as the original clean aggregate structure under the fixed privacy probe. Battery operation also incurs both monetary energy cost $C_{\text{energy}}(t)$ and battery degradation cost $C_{\text{deg}}(t)$. Hence we use the combined reward

$$r_t = \lambda r_t^{\text{priv}} - (1-\lambda)\left(C_t^{\text{energy}} + C_t^{\text{deg}}\right), \tag{6}$$

where $\lambda \in [0,1]$ controls the privacy-cost trade-off. The main privacy experiments use $\lambda = 1$, in which Eq. (6) reduces to the privacy reward in Eq. (5). The cost model is detailed in Appendix A.1.

Table 1: Overall black-box transfer performance averaged over five random seeds. Sig. columns report the Holm–Bonferroni-corrected significant fraction at family-wise $\alpha = 0.05$ (Appendix F.2).

Dataset	Cases	RMSE Inc. ↑	MAE Inc. ↑	SAE Inc. ↑	F1 Red. ↑	RMSE Sig.	MAE Sig.	SAE Sig.	F1 Sig.
UK-DALE	24	107.12%	118.92%	107.36%	79.32%	22/24	21/24	15/24	**24/24**
REDD	30	165.90%	127.35%	196.33%	80.10%	26/30	22/30	25/30	19/30

Hierarchical signature mimicry policy. We use a signature mimicry mechanism based on a library $\mathcal{S} = \{s_k\}_{k=1}^{K}$, where each signature $s_k = [p_1^{(k)}, \ldots, p_{L_k}^{(k)}]$ is extracted from real appliance usage and L_k denotes its duration. The hierarchical policy separates high-level signature selection from low-level battery execution. At each decision point, the manager observes the state s_t, including the current load, time encoding, time-of-use price, battery state of charge, and remaining throughput budget, and outputs a discrete action $a_t^{\text{high}} \in \{0, 1, \ldots, 2K\}$. Here, $a_t^{\text{high}} = 0$ denotes no mimicry; actions $1, \ldots, K$ replay the selected signature as a charging profile, while actions $K+1, \ldots, 2K$ replay the corresponding signature as a discharging profile. The low-level executor then replays the selected signature trajectory subject to battery feasibility constraints, including charge/discharge power limits, daily throughput budget, battery capacity, and the state-of-charge range. Detailed battery dynamics and constraint enforcement are provided in Appendix A.2. If a replay trajectory violates these constraints, the executor terminates the action. During replay, the manager remains inactive and resumes control only after the selected signature finishes or is stopped by battery constraints.

Policy training and deployment. Only the high-level manager is optimized; the privacy probe is fixed and the low-level executor deterministic. We train π_θ with PPO using the reward in Eq. (5). After training, only the manager policy, signature library, and deterministic executor are deployed—the privacy probe is needed only for offline reward computation, and appliance-level labels are used only to build the signature library and for evaluation, so the deployed policy runs online from aggregate load, battery state, time, price, and remaining budget alone, without access to the attacker's model (full procedure in Appendix C).

6 Experiments

We evaluate whether a defender trained with a fixed privacy probe generalizes to heterogeneous black-box NILM attackers: whether it degrades unseen NILM models, whether the effect persists across datasets and test conditions, and which components of the signature-mimicry mechanism drive the privacy gain.

Experimental setup. *Dataset.* UK-DALE Kelly & Knottenbelt (2015b) is the primary benchmark (consistent minute-level preprocessing and train–test protocols); we repeat the full pipeline on REDD Kolter & Johnson (2011) to test a different data distribution (Appendix E.1). *Signature library.* A trajectory-level library of 10 short-burst/medium-duration patterns from intermittent high-power appliances, built once from UK-DALE House 1 training data and fixed across all experiments, including unseen households and REDD, with no target-environment appliance-level data used to adapt it (Appendix E.2). *Privacy probe.* A Transformer Seq2Seq aggregate reconstruction model (Appendix E.3); CNN and LSTM variants are compared in Sec. 6 and perform comparably. *Evaluation attackers.* Six independently trained NILM attackers spanning five paradigms: CO (Hart, 1992), FHMM (Kolter & Jaakkola, 2012), S2P (Zhang et al., 2018), DAE (Kelly & Knottenbelt, 2015a), ELECTRIcity (Sykiotis et al., 2022), and BERT4NILM Yue et al. (2020), trained independently on their original settings with dataset-specific adaptations (Appendix E.6) to reflect black-box transfer rather than overfitting to one architecture. *Defender training.* PPO via Stable-Baselines3 Raffin et al. (2021); main experiments use $\lambda = 1$ (Eq. (5)); hyperparameters in Appendix E.4. *Evaluation metrics.* With π_0 the no-defense policy, $\Delta_{\text{priv}}(f_\phi, \pi_\theta) = R(f_\phi, \pi_\theta) - R(f_\phi, \pi_0)$ quantifies privacy improvement (positive = masked signal harder to invert); we report RMSE increase and F1 reduction as primary metrics (power reconstruction and activation detection), with MAE/SAE increases as complementary.

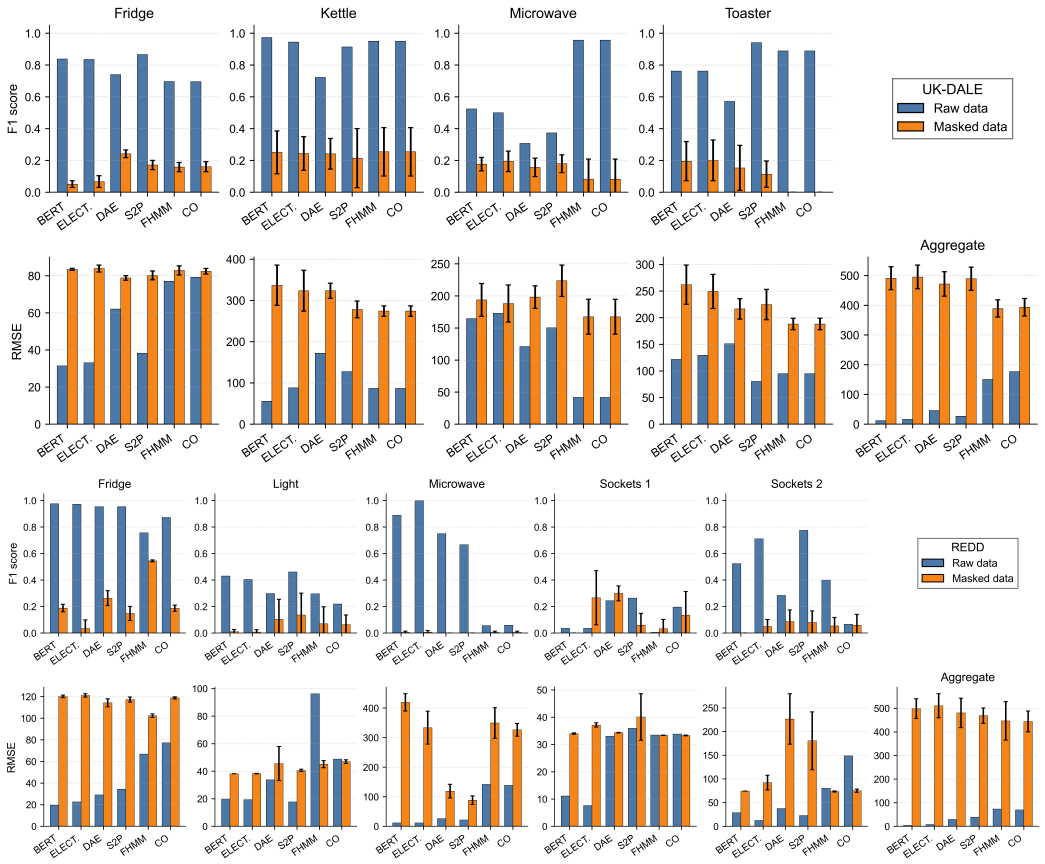

Figure 3: Main black-box transfer results on UK-DALE and REDD, averaged over five random seeds. Top: UK-DALE results over four target appliances and the aggregate channel. Bottom: REDD results over five target appliances and the aggregate channel. The proposed defense consistently increases NILM inference error and reduces appliance detection performance across six unseen attackers.

Main results: Black-box transfer across NILM attackers. Figure 2 shows the learned defender selectively inserting realistic signature trajectories under battery constraints. Figure 3 reports attacker–appliance-level results (six unseen attackers, five seeds): masked signals consistently raise NILM reconstruction error (RMSE) and lower detection accuracy (F1) on both UK-DALE and REDD, despite their different appliance composition and distribution.

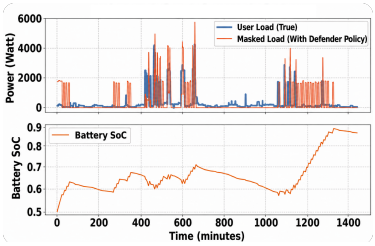

Figure 2: Illustration of the learned defender policy strategy and battery SoC on test day.

Table 1 summarizes dataset-level relative changes (averaged over all attacker–appliance cases and seeds, excluding the aggregate channel); Sig. columns give the fraction of statistically significant cases. Both datasets show large error increases and substantial F1 reductions, and the REDD results confirm the pipeline remains effective under a different dataset distribution rather than being specific to UK-DALE. Full numerical results are in Appendix F.1.

Comparison with existing privacy defenses. Table 2 compares our method against random/rule-based battery baselines and five strong SM privacy defenses—DDQL-flat, DDQL-MI (Shateri et al., 2021; Shateri, 2023), TCN (Li et al., 2023), PLS-DQN, and PLS-DDPG (Zhang et al., 2024a;b), covering representative smoothness/variance-based, information-theoretic, and generic DRL load-shaping reward designs. All baselines and our defender are fully black-box to the six evaluation attackers (no access to their parameters, architecture, gradients, or predictions), and all are tuned

Table 2: Comparison with existing privacy defenses on UK-DALE.

Method	RMSE Inc. ↑	F1 Red. ↑	RMSE Sig.	F1 Sig.	Energy Cost	Deg. Cost	Total Cost	Train Time
No defense	0.00%	0.00%	–	–	0.000	0.000	0.000	–
Random baseline	18.58%	7.25%	7/24	8/24	0.036	**0.029**	0.065	–
Rule-based baseline	32.44%	24.56%	10/24	10/24	0.177	0.207	0.384	–
DDQL-flat	74.03%	70.06%	19/24	20/24	-0.686	0.456	-0.230	3.2h
DDQL-MI	93.51%	72.67%	15/24	20/24	-0.482	0.600	0.118	6.6h
TCN (Li et al., 2023)	76.18%	70.84%	18/24	19/24	-0.641	0.438	-0.203	3.7h
PLS-DQN (Zhang et al., 2024a)	96.42%	74.91%	19/24	20/24	-0.827	0.472	-0.355	4.8h
PLS-DDPG (Zhang et al., 2024b)	98.07%	75.76%	20/24	20/24	-0.936	0.476	-0.460	4.4h
Ours	**107.12%**	**79.32%**	**22/24**	**24/24**	**-1.224**	0.460	**-0.764**	1.2h

Table 3: Component isolation on UK-DALE: signature library × probe reward.

Variant	Signature Library	Probe Reward	RMSE Inc. ↑	F1 Red. ↑	RMSE Sig.	F1 Sig.
Flat + conventional	No	No	63.8%	47.1%	14/24	16/24
Signature + conventional	Yes	No	80.5%	59.6%	18/24	20/24
Flat + probe	No	Yes	87.6%	64.8%	20/24	22/24
Full (Ours)	Yes	Yes	**107.1%**	**79.3%**	**22/24**	**24/24**

under the same battery capacity, power limit, SoC range, throughput budget, tariff, and cost model. Our method is strongest among all seven learned defenses, improving over PLS-DDPG (the best baseline) by 9.1 points RMSE increase and 3.6 points F1 reduction, with more significant cases (22/24 RMSE, 24/24 F1), the lowest energy/total cost, and less training time than every DRL baseline, since privacy reward is computed only when a mimic signature is inserted; the negative energy cost reflects time-of-use arbitrage (charging off-peak, discharging on-peak).

Ablation studies. We conduct ablation studies to examine how the privacy probe architecture and signature library size affect black-box transfer performance. Table 4 shows that Seq2Seq, LSTM, and CNN probes all yield strong black-box transfer, with RMSE increases above 94% and F1 reductions above 77%. Seq2Seq performs best overall, suggesting that a more expressive surrogate probe provides a stronger learning signal, but the defense is not tied to a specific probe architecture: all three variants transfer to the same six unseen attackers with comparable effectiveness. Table 5 shows that reducing the library generally weakens privacy protection, with RMSE increase dropping from 107.12% using 10 signatures to 66.45% using 3 signatures. However, the 9-signature setting remains close to the full library, indicating that moderate pruning has limited impact on transfer while richer libraries continue to improve mimicry diversity and, in turn, privacy protection.

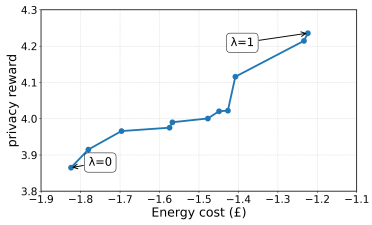

Figure 4: Pareto frontier of privacy reward versus energy cost across the λ sweep.

Source of improvement. To isolate what drives the gain over the baselines in Table 2, Table 3 crosses our signature library against a conventional discrete flat battery action space, and our probe-based reward against a conventional load-leveling objective (squared distance between the masked load and the mean training-set aggregate), in a 2×2 design that keeps the policy, environment, and training protocol otherwise fixed. Relative to flat actions with the conventional objective, the probe reward alone contributes 23.8/17.7 points (RMSE/F1) and the signature library alone contributes 16.7/12.5 points; combining both gives the best performance with a small positive interaction (+2.8/+2.0 points). The probe reward is therefore the primary source of improvement, while the signature library provides an independent, complementary gain rather than being redundant with it.

Table 4: Effect of privacy probe architecture on UK-DALE.

Probe	RMSE Inc.	F1 Red.	RMSE Sig.	F1 Sig.
Seq2Seq	**107.12%**	**79.32%**	**22/24**	**24/24**
LSTM	100.71%	78.75%	**22/24**	22/24
CNN	94.13%	77.46%	21/24	21/24

Table 5: Effect of signature library size on UK-DALE.

Library Size	RMSE Inc.	F1 Red.	RMSE Sig.	F1 Sig.
10 (main)	**107.12%**	**79.32%**	**22/24**	**24/24**
9	103.82%	77.54%	21/24	22/24
7	93.95%	61.87%	19/24	20/24
5	85.03%	68.65%	18/24	19/24
3	66.45%	53.54%	17/24	17/24

Robustness and practical constraints. *Battery configuration sensitivity.* Table 6 studies battery capacity $C_{\max}$, power limit $P_{\max}$, and daily throughput budget $T_{\max}$, averaged over six unseen attackers and four appliances: protection improves substantially with capacity and throughput, with diminishing returns beyond 8 kWh/2 kW, and throughput budget has the strongest influence since it directly limits sustained mimic-signature replay.

Table 6: Battery configuration sensitivity on UK-DALE.

Impact Factor	Battery Config			RMSE Inc. ↑	F1 Red. ↑	Agg. ΔRMSE ↑	RMSE Sig.	F1 Sig.
	$C_{\max}$ (kWh)	$P_{\max}$ (kW)	$T_{\max}$ (kWh)					
Main setting	**8**	**4**	**8**	**107.1%**	**79.3%**	**413.9**	**22/24**	**24/24**
Battery capacity	2	4	8	56.0%	32.6%	239.2	12/24	14/24
	4	4	8	60.2%	42.2%	315.9	17/24	20/24
	12	4	8	107.2%	82.6%	405.4	22/24	24/24
Power limit	8	1	8	85.1%	62.3%	247.9	13/24	14/24
	8	2	8	110.2%	79.3%	393.3	21/24	24/24
	8	6	8	98.0%	80.9%	392.6	22/24	24/24
Throughput budget	8	4	2	35.7%	24.4%	176.0	11/24	11/24
	8	4	4	72.2%	59.0%	288.1	15/24	18/24
	8	4	6	90.2%	76.1%	411.1	21/24	23/24

Privacy-cost trade-off. Figure 4 shows that the privacy-cost trade-off factor λ provides a smooth control knob between privacy and energy cost. The $\lambda = 0$ policy minimizes energy cost through time-of-use arbitrage, whereas $\lambda = 1$ maximizes privacy reward.

Multi-day and multi-household robustness. Tables 7–8 show the same defender generalizes without retraining across seasons, weekdays/weekends, and households on both UK-DALE and REDD: all test days show increased RMSE and reduced F1 (32.1–93.7% on UK-DALE, 42.2–80.1% on REDD). We report both raw and masked metrics, since effect size tracks appliance activity and raw NILM difficulty.

Table 7: Generalization across the main and UK-DALE test days.

Test Day	Active App.	RMSE Raw	RMSE Masked	RMSE Inc. ↑	F1 Raw	F1 Masked	F1 Red. ↓
2013-03-18 H1 (Spring-Monday, Main)	**4**	**95.95**	**197.27**	**105.6%**	**0.773**	**0.161**	**79.2%**
2013-01-18 H1 (Winter-Friday)	4	113.01	176.38	56.1%	0.687	0.361	47.5%
2013-03-30 H1 (Spring-Saturday)	1	7.39	89.76	1115.4%	0.225	0.014	93.6%
2013-03-30 H4 (Spring-Saturday)	2	64.36	135.08	109.9%	0.837	0.568	32.1%
2013-07-15 H1 (Summer-Monday)	4	72.62	150.57	107.3%	0.698	0.196	71.9%
2013-07-15 H2 (Summer-Monday)	3	58.85	164.06	178.8%	0.643	0.040	93.7%

Table 8: Generalization across the main and REDD test days.

Test Day	Active App.	RMSE Raw	RMSE Masked	RMSE Inc. ↑	F1 Raw	F1 Masked	F1 Red. ↓
2011-04-29 H2 (Spring-Friday, Main)	**5**	**44.48**	**115.47**	**165.9%**	**0.493**	**0.098**	**80.1%**
2011-04-30 H2 (Spring-Saturday)	5	61.53	123.75	101.1%	0.545	0.125	77.1%
2011-04-26 H1 (Spring-Tuesday)	3	80.36	170.30	111.9%	0.698	0.308	55.9%
2011-05-25 H3 (Spring-Wednesday)	3	111.89	173.51	55.1%	0.580	0.336	42.2%
2011-05-01 H4 (Spring-Sunday)	1	57.36	71.04	23.8%	0.934	0.360	61.4%

Deployment feasibility and adaptive-attacker robustness. Beyond frozen black-box attackers, we additionally evaluate battery feasibility under residential regulation and effectiveness against attackers that adapt to the defended distribution.

Table 9: Adaptive-attacker recoverability on UK-DALE, averaged over 16 configurations (4 architectures $\times$ 4 appliances).

Attacker setting	Mean F1	Rel. to unprotected	Effect
Unprotected signals	0.723	100%	No defense
Fixed attackers on defended signals	0.161	22.2%	Strong suppression
Scratch retraining on defended signals	0.284	39.3%	Defense remains effective
Fine-tuning on defended signals	0.265	36.6%	Defense remains effective

Table 10: Significance counts before and after multiple-comparison correction.

Dataset (n)	Metric	Uncorrected	After Holm	Deterministic[†]
UK-DALE (24)	RMSE Inc.	22/24	22/24	0
UK-DALE (24)	F1 Red.	22/24	22/24	2
REDD (30)	RMSE Inc.	28/30	26/30	0
REDD (30)	F1 Red.	21/30	15/30	4

Battery feasibility. Main residential battery regulations (US NFPA 855; EU 2023/1542; UK PAS 63100/BS 7671 Amdt. 4) impose no hard capacity cap below 20 kWh, well beyond the 8 kWh used here and consistent with commercial products (e.g., Tesla Powerwall, 13.5 kWh); typical UK usable capacity is 5–14 kWh. Table 6 shows the defense stays effective at 4 kWh, near this lower bound (60.2% RMSE increase, 42.2% F1 reduction), using control signals compatible with existing commercial battery APIs. SoC is bounded in $[0.1, 0.9]$ at every step (Eq. (13)), while matched battery use across methods is enforced by the throughput-budget constraint in Table 2.

Adaptive attackers. We evaluate attackers that retrain from scratch or fine-tune a clean-data checkpoint directly on defended signals, across four trainable architectures (BERT4NILM, S2P, DAE, ELECTRIcity) and four appliances (16 configurations per strategy; per-architecture breakdown in Appendix F.3). Table 9 reports the resulting mean event F1: the defense remains effective, with scratch retraining yielding 0.284 and fine-tuning 0.265, versus 0.723 unprotected—60.7% and 63.3% reductions, consistently across all four architectures. Even with full access to the defended distribution, adaptive attackers recover at most 39% of their unprotected F1, following evaluation practice for adaptive adversaries (Carlini et al., 2019; Tramèr et al., 2020).

Statistical robustness. Significance counts use a two-sided paired t-test per setting over five seeds, Holm–Bonferroni-corrected at $\alpha = 0.05$ (protocol in Appendix F.2); the large majority of attacker–appliance cases remain significant after correction, including all UK-DALE F1 cases. Table 10 reports the resulting counts by dataset and metric.

[†]Complete attacker failures (defended F1 $\equiv 0$), added to the Holm-corrected count for the total effective fraction in Table 1.

7 Conclusion

We studied black-box NILM privacy defense under attacker uncertainty, and proposed a proxy-guided HRL framework in which a fixed self-supervised aggregate-structure probe provides a surrogate training signal, while a curated signature library defines a physically realizable, task-aligned perturbation class. The learned defender transfers to six unseen NILM attackers on UK-DALE and REDD, remains effective across households, seasons, and weekdays/weekends without retraining, and remains robust under regulation-compliant battery capacities and adaptive attackers—supporting a general design principle: pairing a task-aligned surrogate signal with a structured perturbation class yields strong cross-attacker transfer without access to the deployed inference model.

Future directions. Extending the signature library through joint policy–library learning could broaden coverage to composite load patterns, the local surrogate-target alignment analysis (Appendix B) motivates studying broader transfer conditions across attacker families, and the adaptive-attacker results provide a basis for evaluating the framework in online residential deployments.

AI USE STATEMENT

Generative AI tools (LLM-based writing assistants) were used for editing: grammar, spelling, word choice, and LaTeX formatting during the preparation of this manuscript. No generative AI tool was used to generate research ideas, run experiments, analyze results, or write code implementing the proposed method. All AI-assisted text edits were reviewed by the authors, who take full responsibility for the final content of this work, including all text, claims, and reported results.

ETHICS STATEMENT

This work studies a privacy-*enhancing* technology: a household-side defense against appliance-level inference from smart meter data. All datasets (UK-DALE, REDD) are existing, publicly released, de-identified research benchmarks; no new human-subjects data was collected. Appendix D discusses two potential concerns: that the same mechanism could in principle be repurposed to obscure malicious load patterns from utility-side monitoring, and that wide-scale deployment could interact with grid demand forecasting; neither concern is specific to this method beyond other battery-based load-shaping defenses in the literature.

REPRODUCIBILITY STATEMENT

The privacy reward (Eq. (5)), battery feasibility constraints (Appendix A.2), signature-library construction procedure (Appendix E.2), privacy-probe architectures and training splits (Appendix E.3), defender hyperparameters and environment settings (Appendix E.4), and the statistical testing protocol (Appendix F.2) are fully specified to allow independent reproduction. Dataset preprocessing and train/test partitioning are described in Appendix E.1. Code and configuration files will be released upon acceptance.

REFERENCES

Nicholas Carlini, Anish Athalye, Nicolas Papernot, Wieland Brendel, Jonas Rauber, Dimitris Tsipras, Ian Goodfellow, Aleksander Madry, and Alexey Kurakin. On evaluating adversarial robustness. *arXiv preprint arXiv:1902.06705*, 2019.

Kunjin Chen, Qin Wang, Ziyu He, Kunlong Chen, Jun Hu, and Jinliang He. Convolutional sequence to sequence non-intrusive load monitoring. *the Journal of Engineering*, 2018(17):1860–1864, 2018.

Ambra Demontis, Marco Melis, Maura Pintor, Matthew Jagielski, Battista Biggio, Alina Oprea, Cristina Nita-Rotaru, and Fabio Roli. Why do adversarial attacks transfer? explaining transferability of evasion and poisoning attacks. In *28th USENIX security symposium (USENIX security 19)*, pp. 321–338, 2019.

Ecenaz Erdemir, Pier Luigi Dragotti, and Deniz Gündüz. Privacy-cost trade-off in a smart meter system with a renewable energy source and a rechargeable battery. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2687–2691. IEEE, 2019.

Ecenaz Erdemir, Pier Luigi Dragotti, and Deniz Gündüz. Privacy-aware time-series data sharing with deep reinforcement learning. *IEEE Transactions on Information Forensics and Security*, 16: 389–401, 2020.

Farhad Farokhi. A fundamental bound on performance of non-intrusive load monitoring algorithms with application to smart-meter privacy. *IFAC-PapersOnLine*, 53(2):2280–2285, 2020.

Farhad Farokhi and Henrik Sandberg. Fisher information privacy with application to smart meter privacy using hvac units. In *Privacy in dynamical Systems*, pp. 3–17. Springer, 2019.

Jesus Gomez-Vilardebo and Deniz Gündüz. Smart meter privacy for multiple users in the presence of an alternative energy source. *IEEE Transactions on Information Forensics and Security*, 10(1): 132–141, 2014.

Jindong Gu, Xiaojun Jia, Pau de Jorge, Wenqain Yu, Xinwei Liu, Avery Ma, Yuan Xun, Anjun Hu, Ashkan Khakzar, Zhijiang Li, et al. A survey on transferability of adversarial examples across deep neural networks. *arXiv preprint arXiv:2310.17626*, 2023.

George W Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

Maria Kaselimi, Eftychios Protopapadakis, Athanasios Voulodimos, Nikolaos Doulamis, and Anastasios Doulamis. Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring. *Sensors*, 22(15):5872, 2022.

Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. In *2nd ACM int. conf. on embedded systems for energy-efficient built environments*, pp. 55–64, 2015a.

Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2(1):1–14, 2015b.

J Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Artificial intelligence and statistics*, pp. 1472–1482. PMLR, 2012.

J Zico Kolter and Matthew J Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, volume 25, pp. 59–62, 2011.

Donghe Li, Qingyu Yang, Feiye Zhang, Yiqun Wang, Yingchen Qian, and Dou An. Research on privacy issues in smart metering system: An improved tcn-based nilm attack method and practical drl-based rechargeable battery assisted privacy preserving method. *IEEE Transactions on Automation Science and Engineering*, 2023.

Yanpei Liu, Xinyun Chen, Chang Liu, and Dawn Song. Delving into transferable adversarial examples and black-box attacks. *arXiv preprint arXiv:1611.02770*, 2016.

Gautam Raj Mode and Khaza Anuarul Hoque. Adversarial examples in deep learning for multivariate time series regression. In *2020 ieee applied imagery pattern recognition workshop (aipr)*, pp. 1–10. IEEE, 2020.

Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, pp. 506–519, 2017.

Adrien Petralia, Philippe Charpentier, Youssef Kadhi, and Themis Palpanas. Nilmformer: Non-intrusive load monitoring that accounts for non-stationarity. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 4761–4772, 2025.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021. URL `http://jmlr.org/papers/v22/20-1364.html`.

Hasan Rafiq, Prajowal Manandhar, Edwin Rodriguez-Ubinas, Omer Ahmed Qureshi, and Themis Palpanas. A review of current methods and challenges of advanced deep learning-based non-intrusive load monitoring (nilm) in residential context. *Energy and Buildings*, 305:113890, 2024.

Mohammadhadi Shateri. Privacy-cost management in smart meters: Classical vs deep q-learning with mutual information. In *2023 IEEE 11th international conference on smart energy grid engineering (SEGE)*, pp. 109–113. IEEE, 2023.

Mohammadhadi Shateri, Francisco Messina, Pablo Piantanida, and Fabrice Labeau. Privacy-cost management in smart meters using deep reinforcement learning. In *IEEE PES ISGT-Europe*, pp. 929–933, 2020.

Mohammadhadi Shateri, Francisco Messina, Pablo Piantanida, and Fabrice Labeau. Privacy-cost management in smart meters with mutual-information-based reinforcement learning. *IEEE Internet of Things Journal*, 9(22):22389–22398, 2021.

Stavros Sykiotis, Maria Kaselimi, Anastasios Doulamis, and Nikolaos Doulamis. Electricity: An efficient transformer for non-intrusive load monitoring. *Sensors*, 22(8):2926, 2022.

Gamze Kirman Tokgoz, Onat Gungor, Tajana Rosing, and Baris Aksanli. Intarg: Informed real-time adversarial attack generation for time-series regression. *arXiv preprint arXiv:2604.11928*, 2026.

Florian Tramèr, Nicholas Carlini, Wieland Brendel, and Aleksander Madry. On adaptive attacks to adversarial example defenses. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 1633–1645, 2020.

Zhenrui Yue, Camilo Requena Witzig, Daniel Jorde, and Hans-Arno Jacobsen. Bert4nilm: A bidirectional transformer model for non-intrusive load monitoring. In *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, pp. 89–93, 2020.

Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for non-intrusive load monitoring. In *AAAI conf. on artificial intelligence*, volume 32, 2018.

Jiahao Zhang, Chenbei Lu, Hongyu Yi, and Chenye Wu. User-perceptional privacy protection in nilm: A differential privacy approach. *Applied Energy*, 382:125233, 2025.

Ruichang Zhang, Youcheng Sun, and Mustafa A Mustafa. Proactive load-shaping strategies with privacy-cost trade-offs in residential households based on deep reinforcement learning. In *2024 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 160–166. IEEE, 2024a.

Ruichang Zhang, Kaiyue Wu, Youcheng Sun, and Mustafa A Mustafa. Privacy-preserving load-shaping strategies for smart meters using deep reinforcement learning. In *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pp. 1–5. IEEE, 2024b.

Appendix

A Cost Modeling and Battery Constraints

This appendix provides the battery feasibility constraints and operational cost model used in this work. The feasibility constraints define the executable action space of the low-level battery executor, while the cost model quantifies the economic and degradation implications of battery operation. Unless otherwise specified, the main privacy experiments use $\lambda = 1$, corresponding to privacy-focused optimization, and report electricity and battery degradation costs separately.

A.1 Cost Modelling: Battery Degradation and Operational Cost

Battery operation incurs both time-of-use electricity cost and battery degradation cost. These quantities correspond to C_t^{energy} and C_t^{deg} in Eq. (6). In the main privacy experiments, $\lambda = 1$, so Eq. (6) reduces to the privacy reward in Eq. (5); the following costs are evaluated and reported separately.

We adopt a time-of-use electricity tariff $c_{\text{e}}(t)$. Under the sign convention $P_{\text{batt}}(t) > 0$ for charging and $P_{\text{batt}}(t) < 0$ for discharging, the ToU energy cost is defined as

$$C_t^{\text{energy}} = c_{\text{e}}(t)\,\Delta t \begin{cases} \dfrac{P_{\text{batt}}(t)}{\eta_{\text{c}}}, & P_{\text{batt}}(t) > 0 \quad \text{(charging)}, \\ \eta_{\text{d}}\, P_{\text{batt}}(t), & P_{\text{batt}}(t) < 0 \quad \text{(discharging)}, \\ 0, & P_{\text{batt}}(t) = 0. \end{cases} \tag{7}$$

Here, C_t^{energy} may take negative values, reflecting cost savings when the battery discharges to offset grid consumption during high-price periods. This corresponds to temporal energy shifting rather than feed-in tariffs, as no electricity is sold back to the grid.

We set $\eta_{\text{c}} = \eta_{\text{d}} = 1$ in all experiments. Under this setting, Eq. (7) simplifies to

$$C_t^{\text{energy}} = c_{\text{e}}(t)\,\Delta t\, P_{\text{batt}}(t). \tag{8}$$

Battery cycling also induces long-term degradation. Following throughput-based cycle aging models, the degradation cost is modeled as proportional to the processed battery energy:

$$C_t^{\text{deg}} = c_{\text{deg}}\,|P_{\text{batt}}(t)|\,\Delta t, \tag{9}$$

where c_{deg} denotes the amortized degradation cost per processed kilowatt-hour.

The overall operational cost at time t is therefore

$$C_t^{\text{op}} = C_t^{\text{energy}} + C_t^{\text{deg}}. \tag{10}$$

When operational costs are considered during policy optimization, C_t^{op} is combined with the privacy reward through the trade-off coefficient λ in Eq. (6).

A.2 Battery Feasibility Constraints

The masked load follows the sign convention: $P_{\text{batt}}(t) > 0$ denotes battery charging and increases the reported aggregate load, while $P_{\text{batt}}(t) < 0$ denotes battery discharging and decreases the reported aggregate load.

All battery actions generated by the low-level executor are subject to hard physical constraints. First, the battery power is bounded by the maximum charging and discharging rates:

$$-P_{\max}^{\text{dis}} \le P_{\text{batt}}(t) \le P_{\max}^{\text{ch}}, \tag{11}$$

where $P_{\max}^{\text{ch}}$ and $P_{\max}^{\text{dis}}$ denote the maximum charging and discharging power, respectively.

The battery state of charge evolves according to

$$\text{SoC}_{t+1} = \text{SoC}_t + \frac{\eta_{\text{c}}[P_{\text{batt}}(t)]_+\Delta t}{E_{\max}} - \frac{[-P_{\text{batt}}(t)]_+\Delta t}{\eta_{\text{d}} E_{\max}}, \tag{12}$$

where $E_{\max}$ is the battery capacity, η_{c} and η_{d} are the charging and discharging efficiencies, Δt is the sampling interval, and $[x]_+ = \max(x, 0)$.

To avoid unrealistic operation near full depletion or full charge, the battery is constrained to operate within a safe state-of-charge range:

$$0.1 \leq \mathrm{SoC}_t \leq 0.9. \tag{13}$$

When a daily throughput budget is used, the cumulative processed battery energy within one day is additionally constrained by

$$\sum_{t \in \mathcal{D}} |P_{\text{batt}}(t)|\, \Delta t \leq B_{\mathrm{day}}, \tag{14}$$

where $\mathcal{D}$ denotes the set of timesteps within the day and B_{day} is the daily throughput budget.

During signature replay, a high-level action first defines a nominal battery trajectory:

$$\tilde{P}_{\text{batt}}(t+\tau) = \sigma(a_t^{\mathrm{high}}) p_\tau^{(k)}, \qquad \tau = 1, \ldots, L_k, \tag{15}$$

where $\sigma(a_t^{\mathrm{high}}) = +1$ denotes charging and $\sigma(a_t^{\mathrm{high}}) = -1$ denotes discharging. The actually executed battery action $P_{\text{batt}}(t+\tau)$ is obtained by enforcing the constraints in Eqs. (11)–(14). If the nominal replay violates the charge/discharge power limits, the state-of-charge range, or the remaining throughput budget, the executor clips the action to the feasible range or terminates the replay.

B Theoretical Analysis for Proxy-Guided Signature Manipulation

This appendix provides a sufficient-condition analysis for the proxy-guided objective used in Section 4, formalizing a local condition under which optimizing a surrogate privacy probe improves the expected inference risk over a heterogeneous attacker family. This condition can be viewed as a policy-constrained analogue of surrogate-target alignment in transfer-based black-box adversarial attacks.

B.1 Notation and Objective

Let $\mathcal{D}$ denote the data distribution over aggregate load windows and appliance-level targets. For a defense policy π_θ, let $\mathbf{x}^{\pi_\theta}$ denote the masked aggregate window induced by feasible battery actions. For a NILM attacker f_ϕ, the inference risk is

$$R(f_\phi, \pi_\theta) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\mathcal{L}\left(f_\phi(\mathbf{x}^{\pi_\theta}), \mathbf{y} \right) \right], \tag{16}$$

where $\mathcal{L}$ measures appliance-level reconstruction error. The ideal robust objective under attacker uncertainty is

$$\mathcal{R}_{\mathrm{avg}}(\pi_\theta) = \mathbb{E}_{f_\phi \sim \mu} \left[R(f_\phi, \pi_\theta) \right], \tag{17}$$

where μ denotes the latent distribution over possible NILM attackers. Since μ is unknown during training, the defender instead optimizes the proxy risk induced by the fixed aggregate-structure privacy probe:

$$R_p(\pi_\theta) = R(f_\eta, \pi_\theta). \tag{18}$$

Let

$$z_\theta = \psi(\mathbf{x}^{\pi_\theta}) \tag{19}$$

denote the NILM-relevant representation induced by the current policy. The representation map $\psi(\cdot)$ is used only as an analytical abstraction of task-level commonality among NILM attackers, rather than as an explicitly learned feature extractor shared by all models.

For compactness, we write the representation-level proxy risk and expected attacker-family risk as

$$r_p(z_\theta) = R_p(\pi_\theta), \qquad r_{\mathrm{avg}}(z_\theta) = \mathcal{R}_{\mathrm{avg}}(\pi_\theta), \tag{20}$$

where the dependence on $\mathcal{D}$ is implicit. The corresponding representation-level risk gradients are

$$u_p = \nabla_z r_p(z_\theta), \qquad u_{\mathrm{avg}} = \nabla_z r_{\mathrm{avg}}(z_\theta). \tag{21}$$

B.2 Policy-Reachable Representation Directions

The defense policy induces representation changes through feasible battery actions, subject to capacity, power, state-of-charge, throughput, and cost constraints. Therefore, the relevant transfer condition is defined over directions locally reachable by the policy.

Let

$$J_\theta = \frac{\partial z_\theta}{\partial \theta} \tag{22}$$

denote the local policy-to-representation Jacobian. The local policy-reachable representation subspace is

$$\mathcal{T}_\theta = \mathrm{Range}(J_\theta). \tag{23}$$

Let P_θ denote the orthogonal projection onto $\mathcal{T}_\theta$. For any representation-level gradient u, define its reachable component as

$$u^{\mathrm{reach}} = P_\theta u. \tag{24}$$

Since directions orthogonal to $\mathcal{T}_\theta$ cannot be induced by local policy changes, they do not affect the policy-gradient update:

$$J_\theta^\top u = J_\theta^\top u^{\mathrm{reach}}. \tag{25}$$

B.3 Local Surrogate-Target Alignment

We now state a local sufficient condition connecting the proxy risk and the expected attacker-family risk. The condition is defined in the metric induced by the policy-to-representation mapping.

Let

$$M_\theta = J_\theta J_\theta^\top \tag{26}$$

denote the policy-induced metric on the reachable representation subspace. For any two reachable representation directions $a, b \in \mathcal{T}_\theta$, their corresponding policy-gradient inner product is

$$\left\langle J_\theta^\top a, J_\theta^\top b \right\rangle = a^\top M_\theta b. \tag{27}$$

Condition A1: Policy-induced structure-to-risk alignment. At the current policy parameter θ, suppose the reachable proxy-risk direction and the reachable expected-risk direction satisfy

$$\left(u_p^{\mathrm{reach}}\right)^\top M_\theta u_{\mathrm{avg}}^{\mathrm{reach}} \geq \gamma \left(u_p^{\mathrm{reach}}\right)^\top M_\theta u_p^{\mathrm{reach}} - \varepsilon_{\mathrm{tr}}, \qquad \gamma > 0, \tag{28}$$

where $\varepsilon_{\mathrm{tr}} \geq 0$ captures residual transfer error.

This condition states that, after accounting for how feasible signature-mimicry actions change aggregate structures, the probe-risk direction is locally aligned with the expected NILM-risk direction over the attacker family. It does not require the probe to predict appliance-level outputs or replicate any deployed NILM attacker. Instead, it assumes that, within the signature-reachable action space, aggregate structures that strongly degrade probe reconstruction are also structures that affect downstream NILM disaggregation.

Because the attacker distribution μ is unknown, this condition cannot be directly verified during training. Its practical relevance is evaluated empirically by testing whether policies trained against f_η degrade unseen NILM attackers.

B.4 Policy-Level Alignment

The following proposition shows that Condition A1 directly induces alignment between the proxy-gradient direction and the expected-risk gradient in policy-parameter space.

Proposition 1: Policy-level proxy alignment. Under Condition A1,

$$\langle \nabla_\theta R_p(\pi_\theta), \nabla_\theta \mathcal{R}_{\mathrm{avg}}(\pi_\theta) \rangle \geq \gamma \left\| \nabla_\theta R_p(\pi_\theta) \right\|^2 - \varepsilon_{\mathrm{tr}}. \tag{29}$$

Proof. By the chain rule and Eq. (25),

$$\nabla_\theta R_p(\pi_\theta) = J_\theta^\top u_p^{\text{reach}}, \qquad \nabla_\theta \mathcal{R}_{\text{avg}}(\pi_\theta) = J_\theta^\top u_{\text{avg}}^{\text{reach}}. \tag{30}$$

Therefore,

$$\begin{aligned}\langle \nabla_\theta R_p(\pi_\theta), \nabla_\theta \mathcal{R}_{\text{avg}}(\pi_\theta)\rangle &= \left\langle J_\theta^\top u_p^{\text{reach}}, J_\theta^\top u_{\text{avg}}^{\text{reach}}\right\rangle \\ &= \left(u_p^{\text{reach}}\right)^\top J_\theta J_\theta^\top u_{\text{avg}}^{\text{reach}} \\ &= \left(u_p^{\text{reach}}\right)^\top M_\theta u_{\text{avg}}^{\text{reach}}.\end{aligned} \tag{31}$$

Applying Condition A1 gives

$$\begin{aligned}\langle \nabla_\theta R_p(\pi_\theta), \nabla_\theta \mathcal{R}_{\text{avg}}(\pi_\theta)\rangle &\geq \gamma \left(u_p^{\text{reach}}\right)^\top M_\theta u_p^{\text{reach}} - \varepsilon_{\text{tr}} \\ &= \gamma \left\| J_\theta^\top u_p^{\text{reach}}\right\|^2 - \varepsilon_{\text{tr}} \\ &= \gamma \left\|\nabla_\theta R_p(\pi_\theta)\right\|^2 - \varepsilon_{\text{tr}}.\end{aligned} \tag{32}$$

This proves Eq. (29). □

B.5 Local Improvement of the Robust Objective

The previous proposition establishes policy-level alignment. We next state a local improvement result for a small proxy-gradient ascent step.

Condition A2: Local smoothness. Assume that $\mathcal{R}_{\text{avg}}(\pi_\theta)$ is locally L-smooth as a function of θ. That is, for a sufficiently small update $\Delta\theta$,

$$\mathcal{R}_{\text{avg}}(\pi_{\theta+\Delta\theta}) \geq \mathcal{R}_{\text{avg}}(\pi_\theta) + \langle \nabla_\theta \mathcal{R}_{\text{avg}}(\pi_\theta), \Delta\theta\rangle - \frac{L}{2}\|\Delta\theta\|^2. \tag{33}$$

Proposition 2: Proxy-guided local improvement. Let the policy be updated by a small proxy-gradient ascent step:

$$\theta^+ = \theta + \eta \nabla_\theta R_p(\pi_\theta), \tag{34}$$

where $\eta > 0$ is the step size. Under Conditions A1 and A2,

$$\mathcal{R}_{\text{avg}}(\pi_{\theta^+}) - \mathcal{R}_{\text{avg}}(\pi_\theta) \geq \eta \left[\gamma \left\|\nabla_\theta R_p(\pi_\theta)\right\|^2 - \varepsilon_{\text{tr}}\right] - \frac{L\eta^2}{2}\left\|\nabla_\theta R_p(\pi_\theta)\right\|^2. \tag{35}$$

Thus, for sufficiently small η, the proxy-guided update improves the expected attacker-family risk whenever the surrogate-target alignment term dominates the residual transfer error.

Proof. Set

$$\Delta\theta = \eta \nabla_\theta R_p(\pi_\theta). \tag{36}$$

Substituting this update into Eq. (33) gives

$$\mathcal{R}_{\text{avg}}(\pi_{\theta^+}) - \mathcal{R}_{\text{avg}}(\pi_\theta) \geq \eta \left\langle \nabla_\theta \mathcal{R}_{\text{avg}}(\pi_\theta), \nabla_\theta R_p(\pi_\theta)\right\rangle - \frac{L\eta^2}{2}\left\|\nabla_\theta R_p(\pi_\theta)\right\|^2. \tag{37}$$

Applying Proposition 1 yields Eq. (35). □

B.6 Empirical evidence of Condition A1.

Condition A1 has a directly testable empirical implication: a policy trained against f_η should degrade unseen NILM attackers in $\mathcal{F}$. The cross-attacker transfer results in Section 6 (Table 1) provide consistent evidence: across 24 attacker–appliance pairs on UK-DALE and 30 on REDD, the proxy-guided policy increases attacker reconstruction error by over 100% on average and reduces F1 by 80%. This is the empirical signature of $\gamma > 0$ holding broadly across the attacker family.

Algorithm 1 Proxy-Guided HRL Training for Signature Mimicry

Input: Fixed privacy probe f_η, signature library $\mathcal{S}$, battery environment $\mathcal{E}$
Output: Trained manager policy π_θ

Initialize manager policy π_θ
for each training episode **do**
  Reset environment and battery state
  while episode not finished **do**
    Observe state s_t
    Sample high-level action $a_t^{\text{high}} \sim \pi_\theta(\cdot|s_t)$
    if $a_t^{\text{high}} = 0$ **then**
      Apply no mimic action and advance one step
    else
      Select the signature and replay direction from a_t^{high}
      Replay the selected signature using the deterministic executor subject to hard battery constraints
    end if
    Construct the resulting masked window $\mathbf{x}_t^{\pi_\theta}$
    Construct the corresponding unmodified window $\mathbf{x}_t^0$
    Compute privacy reward using Eq. (5)
    Store transition in the rollout buffer
  end while
  Update manager policy π_θ using PPO
end for
Return trained policy π_θ

B.7 Signature Mimic as a Structured Perturbation Class

The main text motivates signature mimic as a task-aligned perturbation mechanism. Here we formalize this design choice at the level of feasible perturbation classes.

Let Δ_{batt} denote the set of perturbations that can be physically realized by the battery under capacity, power, state-of-charge, throughput, and cost constraints. A generic battery-induced perturbation satisfies

$$\delta_{\pi_\theta} \in \Delta_{\text{batt}}. \tag{38}$$

Signature mimic restricts this feasible set to perturbations generated by replaying trajectories from a real appliance-signature library $\mathcal{S}$:

$$\Delta_{\text{sig}} = \{\delta \in \Delta_{\text{batt}} : \delta \text{ is induced by replaying a selected signature } s \in \mathcal{S}\}. \tag{39}$$

Thus,

$$\Delta_{\text{sig}} \subseteq \Delta_{\text{batt}}. \tag{40}$$

This restriction does not make the perturbation class more general. Instead, it makes the perturbations more task-aligned. The goal is not to maximize arbitrary distortion of the aggregate signal, but to introduce plausible appliance-like decoys that interfere with NILM disaggregation. Compared with unstructured random perturbations sampled from Δ_{batt}, perturbations in Δ_{sig} are constructed from real appliance trajectories and therefore retain appliance-like temporal structure.

This provides a formal interpretation of signature mimic as a structured adversarial perturbation class tailored to NILM inference. The effectiveness of this design choice is evaluated empirically through comparisons with random or unstructured perturbation baselines and through ablations over the signature library.

C Algorithm

The full training procedure is summarized in Algorithm 1.

D BROADER IMPACT

This work develops a privacy defense for smart meter users against appliance-level inference attacks, aiming to restore user control over fine-grained behavioral information that smart meter deployments would otherwise expose. The intended societal benefit is improved residential energy privacy without requiring changes to the metering infrastructure or assumptions about the deployed inference model.

We acknowledge two potential concerns. *Misuse risk.* The same proxy-guided framework that learns to inject misleading appliance signatures could in principle be repurposed to obscure malicious load patterns, such as energy theft, from utility-side monitoring. We note that the framework operates under hard battery feasibility constraints and does not alter the user's actual energy consumption—only the reported aggregate trajectory. Distinguishing legitimate privacy protection from concealment of misuse requires utility-side detection mechanisms and is a policy-level question beyond the scope of this work. *Grid-level effects.* Wide-scale deployment of battery-based load shaping interacts with grid demand forecasting and could affect distribution-network operations if adopted at scale. Coordinated deployment protocols and forecasting-aware extensions are an important direction for practical adoption.

E EXPERIMENT SETTINGS

E.1 DATASET DESCRIPTION

We evaluate our method on the UK-DALE dataset Kelly & Knottenbelt (2015b), a widely used public dataset for NILM research. It contains appliance-level and aggregate power readings from multiple UK households. We use data from House 1, downsampled to 1-minute resolution, following prior works Zhang et al. (2024a;b). We extract 10 days of user load data (from March 19 to March 29, 2013) as the training set, and reserve a separate day (March 18, 2013) as the test set. During attacker training, the 10-day data is further split into a training and a cross-validation subset (80:20). The test day is only used to evaluate the performance of the learned defender policy.

To assess cross-dataset applicability, we additionally conduct experiments on the REDD dataset Kolter & Johnson (2011). REDD contains aggregate and appliance-level power measurements from multiple residential buildings in the United States. Similar to the UK-DALE setup, we use continuous data from Building 2, downsampled to 1-minute resolution. We select 11 consecutive days, using the first 10 days (from April 19 to April 29, 2011) for training and the 11th day (April 30, 2011) for testing. The REDD dataset is used to evaluate how well the proposed defense framework performs when facing a range of mainstream NILM attackers under a different data distribution.

E.2 APPLIANCE SIGNATURE SELECTION

We construct a trajectory-level appliance signature library $\mathcal{S}$ from the UK-DALE House 1 dataset, which contains power traces of 37 labeled household appliances. All candidate and selected signatures are extracted exclusively from the UK-DALE House 1 training period, with no temporal overlap with any evaluation day. Once constructed, the library is frozen and reused without target-specific adaptation. The purpose of this library is to provide realistic appliance-level usage patterns that can be replayed by the battery controller during mimicry. Since NILM attackers often rely on distinctive appliance activation patterns in the aggregate load, we focus on intermittent high-power appliances whose signatures are visually and statistically salient, such as kettles, toasters, microwaves, ovens, washer–dryers, and dishwashers.

Candidate event extraction. We first identify the top-6 energy-consuming intermittent high-power appliances and extract candidate usage trajectories from a 10-day period. For each appliance, we apply a threshold-based event segmentation scheme. A device-specific activation threshold is used to determine whether the appliance is operating. The appliance-level power sequence is scanned sequentially: when the power first exceeds the threshold, a new event is initiated and its start time is recorded. Consecutive samples above the threshold are grouped into the same event, forming a continuous usage trajectory. The event is terminated once the power falls below the threshold. To remove noise and spurious activations, events shorter than a predefined minimum duration are discarded. This process yields 188 candidate appliance usage trajectories.

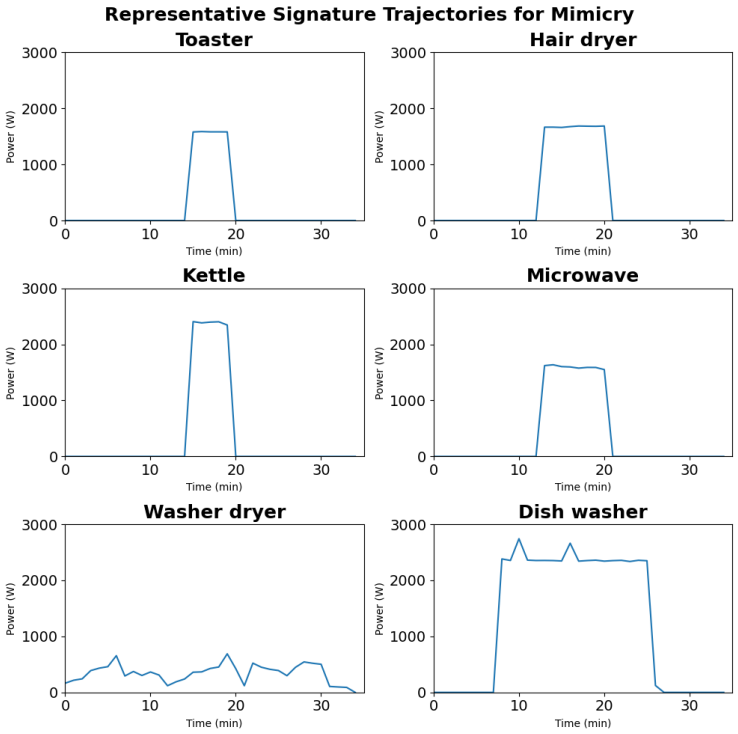

Figure 5: Representative signature trajectories in the constructed signature library. All trajectories are padded to a unified length of 35 minutes for visualization. The signatures exhibit diverse temporal lengths, power magnitudes, and usage patterns, illustrating the diversity and realism of the signature library used for mimicry.

Filtering and ranking. To keep the mimic actions physically meaningful and to simplify the reinforcement learning action space, we further filter the candidate trajectories by duration and power magnitude. Specifically, we retain events whose duration is between 5 and 35 minutes and whose median power lies between 300 and 2500 W. These constraints remove extremely short, low-power, or unusually long events that are less suitable for battery-based replay.

The remaining candidates are ranked according to two criteria. First, we prioritize trajectories with larger standard deviation, as they exhibit stronger temporal variation and therefore provide more distinctive usage patterns. Second, we rank by median power, since higher-power events are more visible in the aggregate load and are more likely to affect NILM inference. To avoid the final library being dominated by a single appliance type, we impose a balancing constraint and retain at most two signatures per appliance category. Signatures are then selected greedily from the ordered candidate list until the final library contains 10 representative trajectories.

Final signature library. The resulting library $\mathcal{S}$ contains 10 appliance signatures covering short high-power bursts and medium-duration usage patterns. Typical examples include kettle, toaster, microwave, washer-dryer, and dishwasher signatures. All selected signatures have durations between 5 and 35 minutes, with most concentrated in the 5–10 minute range. In terms of power magnitude, most signatures fall between 1500 and 2500 W, corresponding to common high-power intermittent household appliances.

Figure 5 visualizes representative trajectories from the final mimic library. The selected signatures exhibit diverse temporal lengths, power magnitudes, and usage patterns, supporting realistic appliance-level mimicry under battery control.

E.3 PRIVACY PROBE TRAINING

The architecture of the Transformer-based Seq2Seq privacy probe model consists of a linear embedding layer projecting input to a 64-dimensional latent space, followed by sinusoidal positional encoding. The encoder comprises 3 Transformer layers with 4 attention heads and a dropout rate of 0.1. A final linear projection maps each hidden state to a 1D power prediction.

The LSTM probe uses a two-layer LSTM encoder with a hidden size of 64. A single-channel input sequence is first processed by the stacked recurrent layers, which output a 64-dimensional hidden representation at each time step. A final linear projection maps each hidden state to a scalar power estimate.

The 1D-CNN probe is built with three temporal convolutional layers. After reshaping the input to channel-first format, the model applies two hidden Conv1D blocks (64 channels, kernel size 3), each followed by ReLU activation, and a final Conv1D layer that maps features to a single output channel. Symmetric padding is used to preserve sequence length throughout the network, and the output is reshaped back to time-first format.

We first generate candidate 60-minute segments via repeated sliding-window passes (stride = 30 minutes) over the 14,400-minute (10-day) period, with each pass beginning at a randomly sampled starting offset uniformly drawn from [0, 1439] and proceeding to the end of the sequence. Within each pass, the first 80% of generated segments (by temporal order) are tentatively tagged as the training set and the last 20% as the validation set. To eliminate leakage arising from the randomized offsets, we discard any training-tagged segment whose end time extends into the final two validation days(i.e., exceeds 11,520 minutes), and any validation-tagged segment whose start time extends into the first eight training days (i.e., precedes 11,520 minutes). This guarantees strict non-overlap in absolute time between the training and validation sets. The procedure repeats across independently sampled offsets until 5,000 segments are collected for each set. The remaining one-day data is reserved exclusively for evaluation. The model is trained using mean squared error (MSE) loss as loss function.

E.4 DEFENDER TRAINING PARAMETERS AND SM ENVIRONMENT SETTINGS

The PPO defender training proceeds for 300 PPO episodes, with 1,440 time steps per episode (corresponding to one full day). The PPO optimizer uses a learning rate of 2×10^{-5}, a discount factor $\gamma = 0.99$, GAE parameter $\lambda = 0.95$. We adopt a rollout horizon of 4,096 steps, 4 optimization epochs per update, a minibatch size of 2,048, and a clipping range of 0.1. Entropy regularization is enabled with a coefficient of 0.01 to encourage early exploration, and a target KL divergence of 0.01 is used to stabilize policy updates.

The defender operates within a simulated battery and SM environment with the following constraints: a maximum capacity of 8 kWh, peak charge/discharge power of 4 kW, and a SoC constraint between 10% and 90%. The SoC is initialized to 50% at the beginning of each episode. All actions are applied at 1-minute resolution.

E.5 COMPUTATIONAL RESOURCES.

All experiments were conducted on a single NVIDIA RTX 3050 Ti GPU. Defender training takes approximately 1.2 hours per run; full reproduction of the main results (both datasets, five seeds, six attackers) requires roughly 30 GPU-hours.

E.6 ATTACKER ADAPTATION DETAILS

Each attacker preserves its original formulation and architecture, with dataset-specific adaptations to the 1-minute sampling rate; these adaptations preserve the corresponding temporal context and inference formulation, and all are fixed using only unmasked training/validation data.

Table 11: Per-attacker adaptation to 1-minute resolution.

Attacker	Adaptation
BERT4NILM / ELECTRIcity	Original architecture and 60-sample windows unchanged; normalization statistics recomputed from each dataset's clean training split (60 minutes of context at 1-minute resolution).
DAE	60-sample window unchanged; balanced sampling, weighted MSE, and early stopping address sparse appliance activations after minute-level aggregation.
S2P	Original ∼6s-resolution window of 599 samples spans ∼1 hour; at 1-minute resolution we use 61 samples to preserve the same ∼1-hour receptive field, not to shrink it.
CO / FHMM	Number of states unchanged (three); candidate-appliance count set by dataset coverage (CO: 10/8, FHMM: 12/8 on UK-DALE/REDD), which expands rather than restricts recovery targets.

Table 12: Detailed Raw, Masked, Delta, and p-value results on UK-DALE over five random seeds.

Attacker	Appliance	Raw Data				Masked Data				Delta (Masked - Raw)				p-value			
		RMSE	MAE	SAE	F1	RMSE	MAE	SAE	F1	RMSE	MAE	SAE	F1	RMSE	MAE	SAE	F1
CO	Aggregate	176.97	31.53	0.07	–	393.16±26.34	194.98±7.13	0.84±0.07	–	216.18±26.34	163.44±7.13	0.77±0.07	–	0.00	0.00	0.00	–
	Fridge freezer	79.24	41.01	0.81	0.70	82.41±1.33	40.33±0.86	0.84±0.01	0.16±0.03	3.17±1.33	-0.68±0.86	0.02±0.01	-0.54±0.03	0.01	0.19	0.02	0.00
	Kettle	86.62	4.85	0.05	0.95	274.23±11.35	33.14±2.83	0.59±0.29	0.25±0.14	187.61±11.35	28.29±2.83	0.54±0.29	-0.70±0.14	0.00	0.00	0.02	0.00
	Microwave	41.76	2.37	0.02	0.96	167.45±24.36	18.83±5.02	0.49±0.29	0.08±0.11	125.70±24.36	16.46±5.02	0.47±0.29	-0.87±0.11	0.00	0.00	0.03	0.00
	Toaster	94.89	6.40	0.21	0.89	188.11±9.56	20.57±2.22	0.90±0.12	0.00±0.00	93.23±9.56	14.17±2.22	0.69±0.12	-0.89±0.00	0.00	0.00	0.00	0.00
FHMM	Aggregate	150.54	20.62	0.02	–	389.00±26.01	183.97±7.81	0.74±0.07	–	238.46±26.01	163.35±7.81	0.72±0.07	–	0.00	0.00	0.00	–
	Fridge freezer	77.11	39.07	0.78	0.70	82.87±2.22	40.64±1.26	0.83±0.02	0.16±0.03	5.76±2.22	1.57±1.26	0.05±0.02	-0.54±0.03	0.01	0.07	0.01	0.00
	Kettle	86.62	4.85	0.05	0.95	274.23±11.35	33.14±2.83	0.59±0.29	0.25±0.14	187.61±11.35	28.29±2.83	0.54±0.29	-0.70±0.14	0.00	0.00	0.02	0.00
	Microwave	41.76	2.37	0.02	0.96	167.45±24.36	18.83±5.02	0.49±0.29	0.08±0.11	125.70±24.36	16.46±5.02	0.47±0.29	-0.87±0.11	0.00	0.00	0.03	0.00
	Toaster	94.89	6.40	0.21	0.89	188.11±9.56	20.57±2.22	0.90±0.12	0.00±0.00	93.23±9.56	14.17±2.22	0.69±0.12	-0.89±0.00	0.00	0.00	0.00	0.00
DAE	Aggregate	46.09	22.38	0.01	–	471.59±36.74	224.61±11.86	0.46±0.08	–	425.50±36.74	202.23±11.86	0.44±0.08	–	0.00	0.00	0.00	–
	Fridge freezer	62.07	28.91	0.10	0.74	78.79±1.08	40.09±1.17	0.63±0.03	0.24±0.02	16.72±1.08	11.18±1.17	0.53±0.03	-0.50±0.02	0.00	0.00	0.00	0.00
	Kettle	172.13	35.48	0.25	0.72	323.42±16.39	70.15±8.34	0.57±0.31	0.24±0.09	151.29±16.39	34.67±8.34	0.32±0.31	-0.48±0.09	0.00	0.00	0.11	0.00
	Microwave	121.10	33.13	1.29	0.31	198.14±15.75	50.26±5.29	2.22±0.21	0.16±0.05	77.04±15.75	17.13±5.29	0.93±0.21	-0.15±0.05	0.00	0.00	0.00	0.00
	Toaster	151.07	36.44	0.74	0.57	216.61±17.21	49.77±5.81	0.67±0.44	0.15±0.13	65.54±17.21	13.33±5.81	-0.06±0.44	-0.42±0.13	0.00	0.01	0.79	0.00
S2P	Aggregate	26.62	12.72	0.00	–	489.24±34.84	233.44±12.27	0.47±0.08	–	462.62±34.84	220.71±12.27	0.46±0.08	–	0.00	0.00	0.00	–
	Fridge freezer	38.28	19.45	0.10	0.87	80.18±2.13	40.86±0.62	0.67±0.02	0.17±0.03	41.90±2.13	21.41±0.62	0.57±0.02	-0.70±0.03	0.00	0.00	0.00	0.00
	Kettle	127.35	15.49	0.36	0.91	278.11±18.24	44.24±4.88	0.28±0.18	0.21±0.17	150.75±18.24	28.75±4.88	-0.08±0.18	-0.70±0.17	0.00	0.00	0.42	0.00
	Microwave	150.39	24.99	1.14	0.37	223.62±21.83	43.87±7.30	2.12±0.54	0.18±0.05	73.23±21.83	18.88±7.30	0.97±0.54	-0.19±0.05	0.00	0.01	0.02	0.00
	Toaster	80.27	10.80	0.03	0.94	224.70±25.56	41.86±8.95	0.67±0.53	0.11±0.07	144.44±25.56	31.06±8.95	0.64±0.53	-0.83±0.07	0.00	0.00	0.07	0.00
ELECTRICITY	Aggregate	16.20	14.11	0.05	–	495.12±35.71	250.55±12.46	0.53±0.08	–	478.92±35.71	236.43±12.46	0.48±0.08	–	0.00	0.00	0.00	–
	Fridge freezer	33.13	17.19	0.12	0.84	83.88±1.65	42.80±0.63	1.00±0.01	0.07±0.03	50.75±1.65	25.61±0.63	0.88±0.01	-0.77±0.03	0.00	0.00	0.00	0.00
	Kettle	88.04	8.37	0.21	0.94	323.55±44.29	53.82±11.85	0.24±0.14	0.24±0.09	235.51±44.29	45.45±11.85	0.03±0.14	-0.70±0.09	0.00	0.00	0.68	0.00
	Microwave	172.84	25.86	1.20	0.50	188.11±25.85	35.08±8.12	0.96±0.69	0.19±0.06	15.27±25.85	9.22±8.12	-0.24±0.69	-0.31±0.06	0.30	0.09	0.53	0.00
	Toaster	129.56	21.01	0.14	0.76	249.34±28.75	51.65±10.71	0.84±0.53	0.20±0.11	119.79±28.75	30.64±10.71	0.70±0.53	-0.56±0.11	0.00	0.00	0.06	0.00
BERT4NILM	Aggregate	11.86	11.32	0.04	–	491.00±34.54	247.27±12.20	0.53±0.08	–	479.14±34.54	235.95±12.20	0.48±0.08	–	0.00	0.00	0.00	–
	Fridge freezer	31.46	16.22	0.09	0.84	83.43±0.46	42.73±0.41	1.02±0.01	0.05±0.02	51.97±0.46	26.50±0.41	0.92±0.01	-0.79±0.02	0.00	0.00	0.00	0.00
	Kettle	55.75	6.79	0.17	0.97	336.92±43.63	57.87±12.34	0.31±0.07	0.25±0.12	281.17±43.63	51.08±12.34	0.14±0.07	-0.72±0.12	0.00	0.00	0.02	0.00
	Microwave	164.56	26.23	0.94	0.52	193.75±22.96	44.02±8.14	0.67±0.41	0.18±0.04	29.19±22.96	17.79±8.14	-0.26±0.41	-0.35±0.04	0.06	0.01	0.27	0.00
	Toaster	121.86	16.81	0.34	0.76	262.04±32.90	51.07±11.57	0.94±0.84	0.20±0.11	140.18±32.90	34.26±11.57	0.60±0.84	-0.57±0.11	0.00	0.00	0.23	0.00

F Original Experiment results

F.1 Detailed Main Results

Tables 12 and 13 report the detailed Raw, Masked, Delta, and p-value results used to generate Fig. 3.

F.2 Statistical Testing Protocol

For each (attacker × appliance) setting, we run the full pipeline under 5 independent random seeds (42, 123, 2024, 3407, 9876), producing paired no-defense and defended results per seed. We apply a two-sided paired t-test on the per-seed differences (defended − no-defense) for each metric (RMSE, MAE, SAE, F1), with significance declared at $p < 0.05$; the pairing is by seed, so each test has 4

Table 13: Detailed Raw, Masked, Delta, and p-value results on REDD over five random seeds.

Attacker	Appliance	Raw Data				Masked Data				Delta (Masked - Raw)				p-value			
		RMSE	MAE	SAE	F1	RMSE	MAE	SAE	F1	RMSE	MAE	SAE	F1	RMSE	MAE	SAE	F1
CO	Aggregate	69.69	13.72	0.03	–	444.12±44.37	281.58±13.68	0.56±0.09	–	374.44±44.37	267.86±13.68	0.53±0.09	–	0.00	0.00	0.00	–
	Fridge	77.27	56.16	0.33	0.87	118.83±1.01	83.55±0.83	0.92±0.02	0.19±0.02	41.56±1.01	27.39±0.83	0.59±0.02	-0.68±0.02	0.00	0.00	0.00	0.00
	Light	48.79	22.11	0.26	0.22	46.99±1.33	22.19±0.88	0.73±0.04	0.06±0.07	-1.80±1.33	0.08±0.88	0.48±0.04	-0.16±0.07	0.04	0.84	0.00	0.01
	Microwave	138.65	33.33	2.89	0.06	326.20±21.71	81.38±9.53	8.62±1.24	0.00±0.01	187.55±21.71	48.05±9.53	5.73±1.24	-0.05±0.01	0.00	0.00	0.00	0.00
	Sockets 1	33.86	8.92	0.44	0.20	33.29±0.13	4.71±0.42	0.88±0.04	0.13±0.18	-0.57±0.13	-4.21±0.42	0.44±0.04	-0.06±0.18	0.00	0.00	0.00	0.49
	Sockets 2	148.77	98.30	7.95	0.07	75.18±3.35	13.35±2.69	0.68±0.26	0.06±0.08	-73.59±3.35	-84.95±2.69	-7.27±0.26	-0.01±0.08	0.00	0.00	0.00	0.85
FHMM	Aggregate	73.52	41.97	0.08	–	446.71±81.86	262.56±25.94	0.34±0.13	–	373.19±81.86	220.59±25.94	0.26±0.13	–	0.00	0.00	0.01	–
	Fridge	66.89	40.61	0.32	0.76	102.32±1.56	79.73±1.13	0.45±0.02	0.54±0.01	35.43±1.56	39.12±1.13	0.13±0.02	-0.21±0.01	0.00	0.00	0.00	0.00
	Light	96.23	71.92	2.84	0.30	45.07±2.60	20.90±1.71	0.81±0.11	0.07±0.13	-51.16±2.60	-51.02±1.71	-2.04±0.11	-0.23±0.13	0.00	0.00	0.00	0.02
	Microwave	141.72	34.66	3.06	0.06	349.26±51.61	89.13±18.17	9.62±2.37	0.00±0.01	207.53±51.61	54.47±18.17	6.56±2.37	-0.05±0.01	0.00	0.00	0.00	0.00
	Sockets 1	33.50	5.43	0.81	0.01	33.38±0.06	4.91±0.32	0.91±0.05	0.03±0.07	-0.12±0.06	-0.52±0.32	0.10±0.05	0.03±0.07	0.01	0.02	0.01	0.43
	Sockets 2	80.39	17.65	0.61	0.40	73.40±1.10	11.34±0.50	0.70±0.03	0.05±0.06	-6.99±1.10	-6.31±0.50	0.09±0.03	-0.35±0.06	0.00	0.00	0.00	0.00
DAE	Aggregate	28.92	15.25	0.01	–	480.67±62.38	285.13±16.30	0.37±0.11	–	451.75±62.38	269.88±16.30	0.35±0.11	–	0.00	0.00	0.00	–
	Fridge	29.06	14.88	0.01	0.95	114.22±3.72	82.69±2.77	0.78±0.04	0.26±0.06	85.16±3.72	67.81±2.77	0.77±0.04	-0.69±0.06	0.00	0.00	0.00	0.00
	Light	33.81	15.63	0.03	0.30	45.57±12.39	14.67±3.43	0.47±0.19	0.10±0.15	11.76±12.39	-0.96±3.43	0.44±0.19	-0.19±0.15	0.10	0.57	0.01	0.04
	Microwave	26.07	6.01	0.52	0.75	118.84±23.04	26.41±7.32	2.50±0.92	0.00±0.00	92.77±23.04	20.40±7.32	1.99±0.92	-0.75±0.00	0.00	0.00	0.01	–
	Sockets 1	33.06	8.01	0.53	0.24	34.32±0.11	13.06±0.46	1.98±0.09	0.30±0.06	1.26±0.11	5.05±0.46	1.45±0.09	0.05±0.06	0.00	0.00	0.00	0.10
	Sockets 2	37.49	12.42	0.21	0.28	226.49±53.25	73.36±10.66	4.30±1.05	0.09±0.09	189.00±53.25	60.95±10.66	4.09±1.05	-0.20±0.09	0.00	0.00	0.00	0.01
S2P	Aggregate	39.09	12.63	0.01	–	468.86±32.56	274.56±14.17	0.44±0.09	–	429.78±32.56	261.93±14.17	0.43±0.09	–	0.00	0.00	0.00	–
	Fridge	34.43	16.12	0.08	0.95	117.25±2.33	84.07±2.18	0.87±0.03	0.15±0.05	82.82±2.33	67.95±2.18	0.79±0.03	-0.81±0.05	0.00	0.00	0.00	0.00
	Light	17.71	7.92	0.06	0.46	40.57±0.84	18.38±0.52	0.80±0.11	0.14±0.17	22.86±0.84	10.46±0.52	0.74±0.11	-0.33±0.17	0.00	0.00	0.00	0.01
	Microwave	21.71	5.99	0.54	0.67	88.40±14.50	14.01±2.84	0.30±0.24	0.00±0.00	66.69±14.50	8.02±2.84	-0.24±0.24	-0.67±0.00	0.00	0.00	0.09	–
	Sockets 1	35.96	9.22	0.65	0.26	40.08±8.52	7.18±3.00	0.58±0.16	0.06±0.09	4.12±8.52	-2.04±3.00	-0.07±0.16	-0.20±0.09	0.34	0.20	0.38	0.01
	Sockets 2	22.87	4.56	0.21	0.78	180.42±60.98	43.52±16.65	2.60±1.66	0.08±0.09	157.54±60.98	38.96±16.65	2.39±1.66	-0.70±0.09	0.00	0.01	0.03	0.00
ELECTRICITY	Aggregate	7.72	6.00	0.02	–	510.80±50.87	301.77±14.85	0.40±0.10	–	503.08±50.87	295.77±14.85	0.38±0.10	–	0.00	0.00	0.00	–
	Fridge	22.57	8.84	0.01	0.97	121.09±1.54	85.05±2.07	0.96±0.03	0.03±0.06	98.52±1.54	76.21±2.07	0.95±0.03	-0.94±0.06	0.00	0.00	0.00	0.00
	Light	19.34	7.93	0.08	0.40	38.21±0.17	11.57±0.07	0.60±0.01	0.01±0.02	18.86±0.17	3.64±0.07	0.52±0.01	-0.40±0.02	0.00	0.00	0.00	0.00
	Microwave	11.81	3.91	0.14	1.00	333.52±55.63	89.82±22.03	8.74±3.02	0.01±0.01	321.71±55.63	85.92±22.03	8.60±3.02	-0.99±0.01	0.00	0.00	0.00	0.00
	Sockets 1	7.60	3.85	0.05	0.04	37.14±0.80	18.88±1.98	2.70±0.26	0.27±0.20	29.55±0.80	15.03±1.98	2.64±0.26	0.23±0.20	0.00	0.00	0.00	0.07
	Sockets 2	12.32	5.17	0.17	0.71	92.20±15.54	19.80±3.62	0.44±0.28	0.05±0.05	79.88±15.54	14.63±3.62	0.27±0.28	-0.66±0.05	0.00	0.00	0.10	0.00
BERT4NILM	Aggregate	4.22	3.24	0.01	–	498.50±41.98	300.22±12.80	0.43±0.09	–	494.29±41.98	296.98±12.80	0.42±0.09	–	0.00	0.00	0.00	–
	Fridge	19.68	8.34	0.01	0.98	120.19±1.14	85.62±1.58	0.95±0.03	0.19±0.03	100.51±1.14	77.29±1.58	0.94±0.03	-0.79±0.03	0.00	0.00	0.00	0.00
	Light	19.84	7.96	0.03	0.43	38.14±0.02	11.36±0.03	0.60±0.00	0.01±0.02	18.30±0.02	3.41±0.03	0.57±0.00	-0.42±0.02	0.00	0.00	0.00	0.00
	Microwave	11.82	3.24	0.12	0.89	418.96±29.23	116.65±13.85	12.38±1.89	0.00±0.01	407.14±29.23	113.41±13.85	12.25±1.89	-0.88±0.01	0.00	0.00	0.00	0.00
	Sockets 1	11.10	3.99	0.19	0.04	34.00±0.23	8.08±0.41	1.67±0.08	0.00±0.00	22.90±0.23	4.09±0.41	1.48±0.08	-0.04±0.00	0.00	0.00	0.00	–
	Sockets 2	28.64	9.27	0.62	0.52	74.28±0.06	20.34±0.07	1.91±0.01	0.00±0.00	45.64±0.06	11.08±0.07	1.29±0.01	-0.52±0.00	0.00	0.00	0.00	–

degrees of freedom. This yields 24 settings on UK-DALE (6 attackers $\times$ 4 appliances) and 30 on REDD (6 attackers $\times$ 5 appliances); aggregate-channel results are reported separately and excluded from these counts.

We apply the Holm–Bonferroni step-down procedure to control the family-wise error rate at $\alpha = 0.05$, treating each dataset's settings as one family. We use Holm rather than Benjamini–Hochberg because attackers evaluated on the same defended signal are dependent, and Holm makes no independence assumption. Table 10 in the main text (Section 6) reports the resulting counts before and after correction.

Settings where the defended F1 is exactly 0.000 in all 5 seeds (zero variance) leave the t-statistic undefined; we verified these are complete attacker failures rather than a data-processing artefact (e.g., on the UK-DALE toaster, the undefended attacker predicts events above the 500 W threshold while the defended-signal output never crosses it, giving F1$= 0$ with no seed-to-seed variation). We report these separately as deterministic reductions in Table 10 and add them to the Holm-corrected count to obtain the total effective significant fraction reported in Table 1 (26/30 RMSE, 19/30 F1 on REDD).

F.3 Adaptive-Attacker Full Configuration

Table 14 reports the per-architecture breakdown underlying Table 9 in the main text. Scratch retraining and fine-tuning both improve every trainable attacker relative to its fixed counterpart, but none approaches its unprotected performance; scratch retraining is slightly stronger overall and outperforms fine-tuning on three of the four architectures.

Table 14: Adaptive-attacker mean F1 by architecture on UK-DALE (4 appliances per architecture).

Architecture	**Fixed**	**Scratch retraining**	**Fine-tuning**
BERT4NILM	0.149	0.225	0.232
S2P	0.179	0.325	0.283
DAE	0.177	0.306	0.283
ELECTRIcity	0.139	0.280	0.260