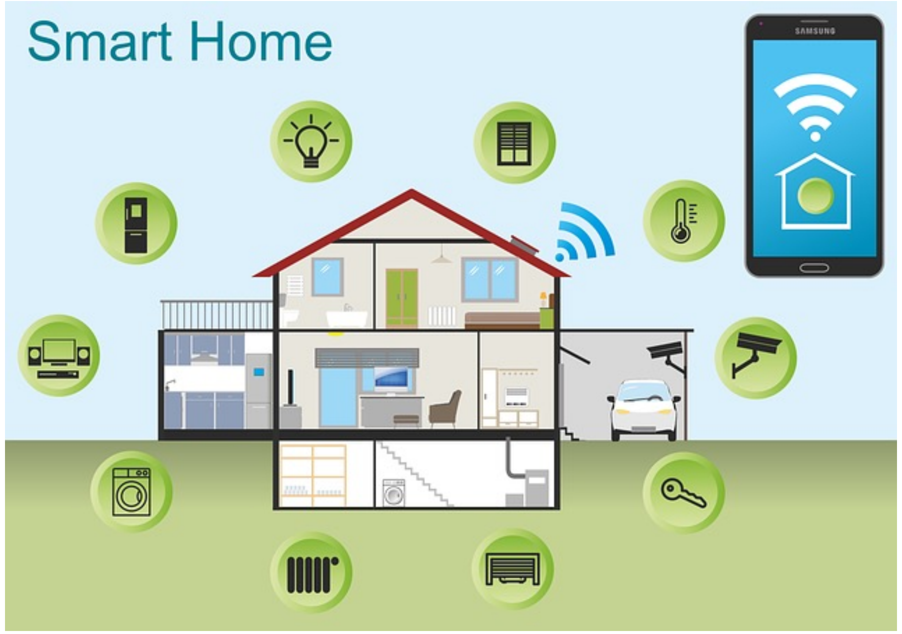
KDD2024
BARCELONA, SPAIN

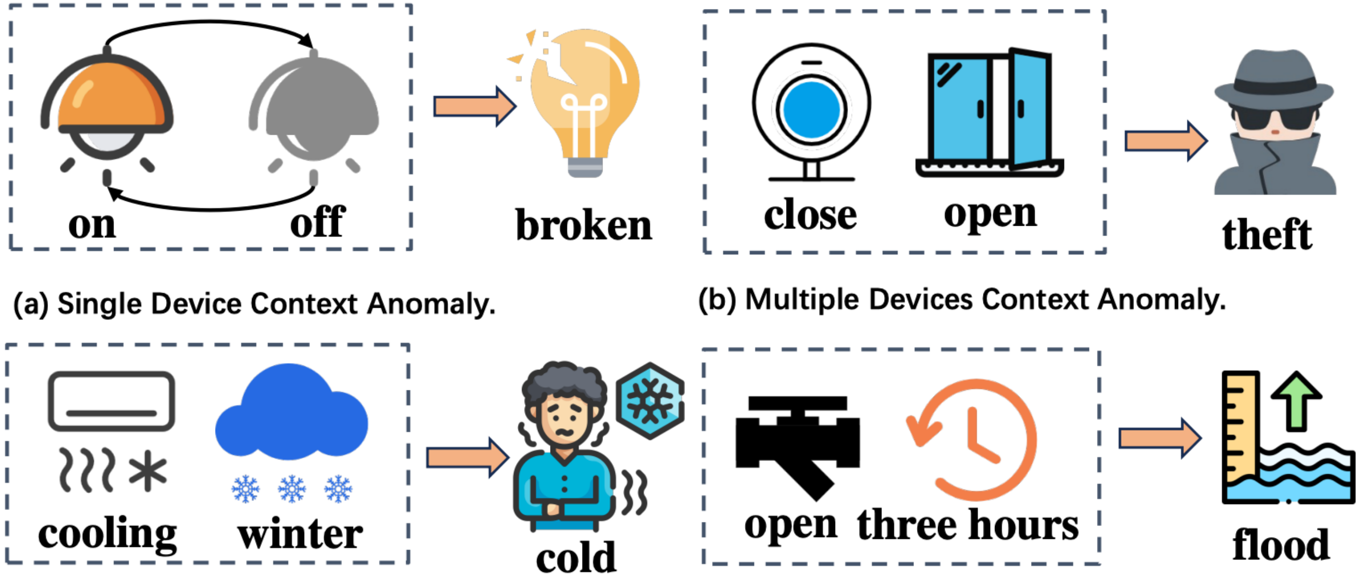
Make Your Home Safe: Time-aware Unsupervised User Behavior Anomaly Detection in Smart Homes via Loss-guided Mask

Jingyu Xiao*, Zhiyao Xu*, Qingsong Zou*, Qing Li ‡, Dan Zhao, Dong Fang
Ruoyu Li, Wenxin Tang, Kang Li, Xudong Zuo, Penghui Hu, Yong Jiang,
Zixuan Weng, Michael R. Lyu

1 Background

- Smart homes, powered by the Internet of Things, offer great convenience
- The abnormal behaviors pose substantial security risks within smart homes
  - Improper operations by users
  - Attacks from malicious attackers

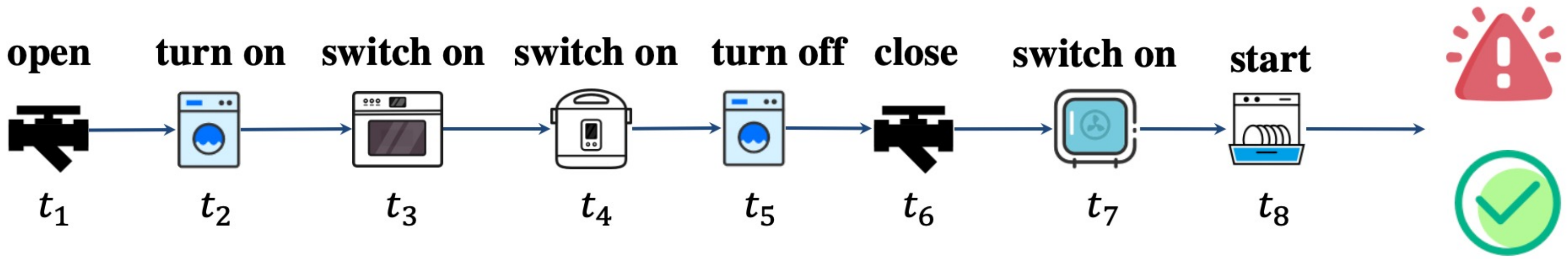




(a) Single Device Context Anomaly.
(b) Multiple Devices Context Anomaly.
(c) Device Control-Moment Context Anomaly.
(d) Device Control-Duration Context Anomaly.

2 Problem Definition

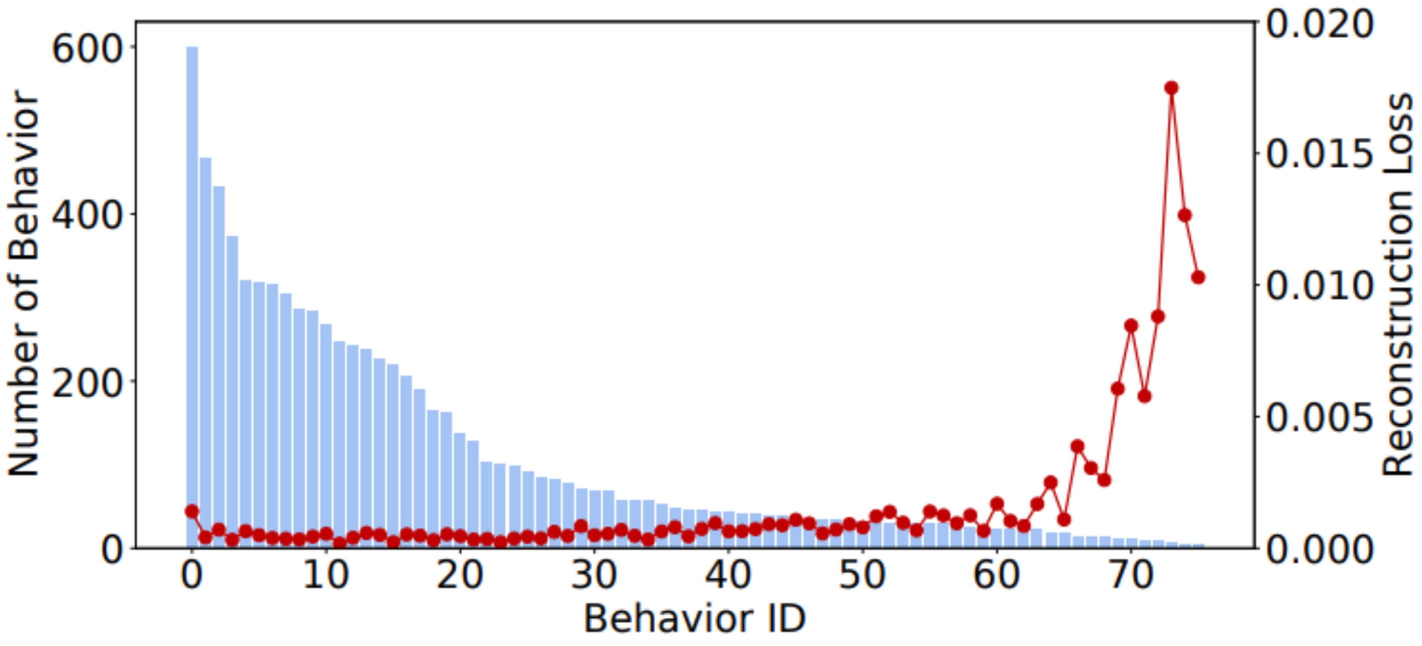
- **User Behaviors Sequence (UBS) anomaly detection in smart homes**
  - **Given a behavior sequence s, detecting whether there are anomalies in sequences.**
  - **Because there are fewer abnormal behaviors, unsupervised models are adopted.**
    - **6thSense [1] utilizes Naive Bayes to detect malicious behavior associated with sensors in smart homes.**
    - **Aegis[2] utilizes a Markov Chain-based machine learning technique to detect malicious behavior in smart homes.**
    - **ARGUS[3] designed an Autoencoder based on Gated Recurrent Units (GRU) to detect IoT infiltration attacks.**



[1] {6thSense}: A context-aware sensor-based attack detector for smart devices (USENIX Security 17)
[2] Aegis: A Context-Aware Security Framework for Smart Home Systems (ACSAC '19)
[3] ARGUS: Context-Based Detection of Stealthy IoT Infiltration Attacks (USENIX Security 2023)

3 Challenge #1

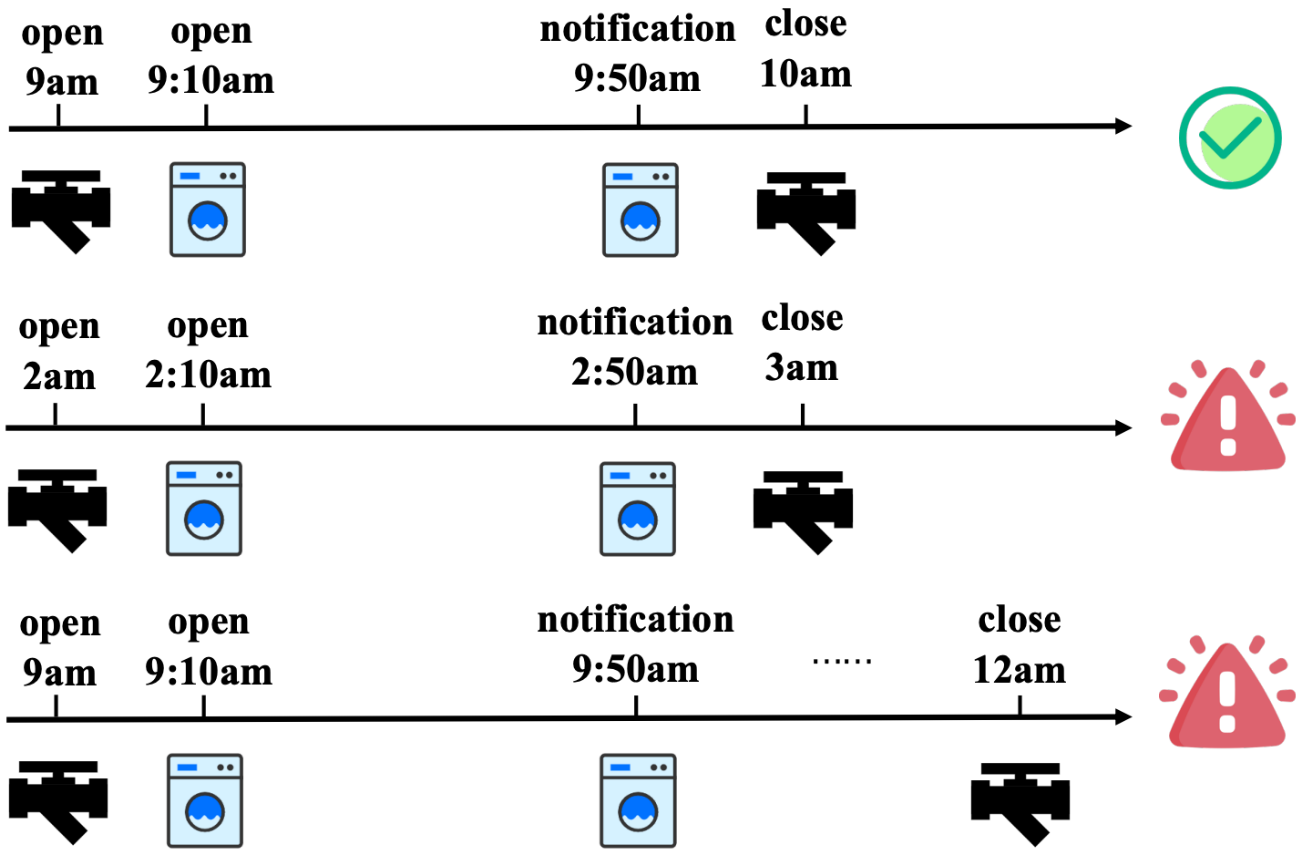
- **Behavior imbalance** leads to challenges in learning the semantics of these behaviors
  - Some behaviors, which occur frequently in similar contexts, can be easily inferred, while others that rarely appear or manifest in diverse contexts can be more challenging to infer.



https://github.com/xjywhu/SmartGuard

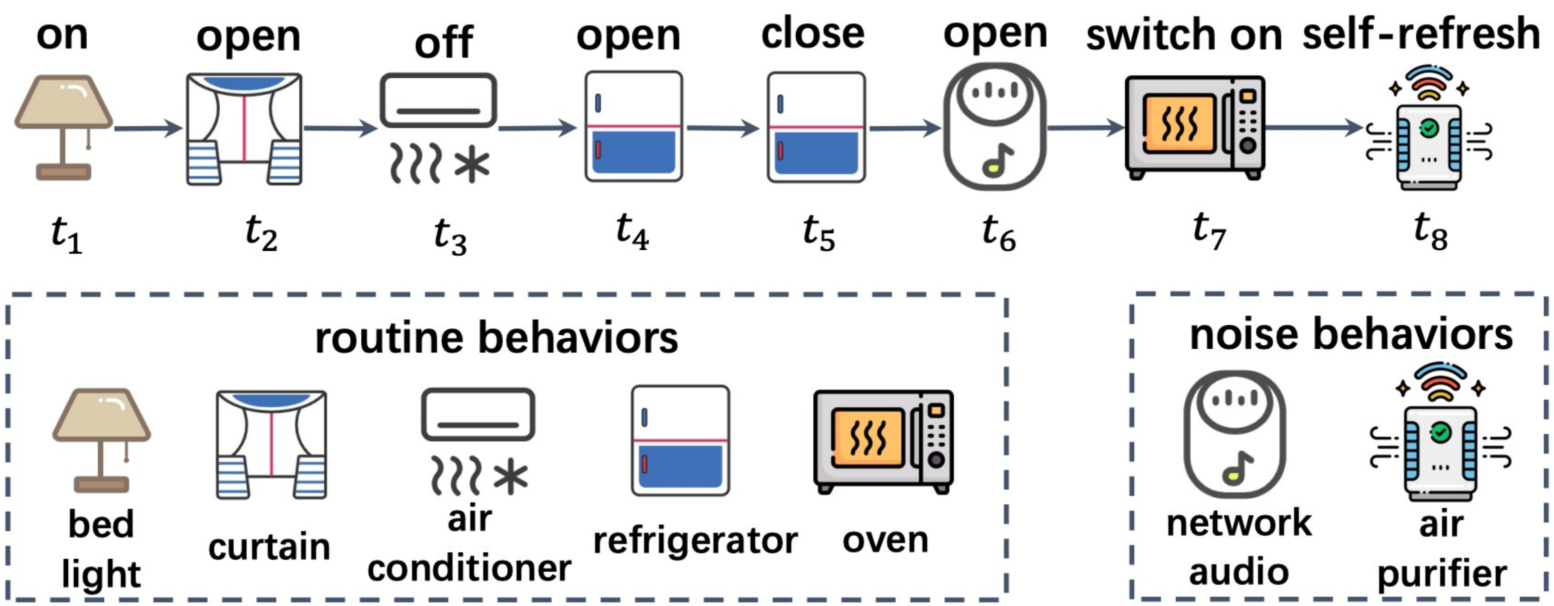
3 Challenge #2

- **Temporal context** plays a significant role in abnormal behavior detection
  - Timing and duration of user behaviors is overlooked by existing solutions



3 Challenge #3

- **Noise behaviors** in user behavior sequences interferes model's inference
  - 1) active behaviors, e.g., suddenly deciding to "turn on the network audio" to listen to music; 2) passive behavior from devices, e.g., the "self-refresh" of the air purifier

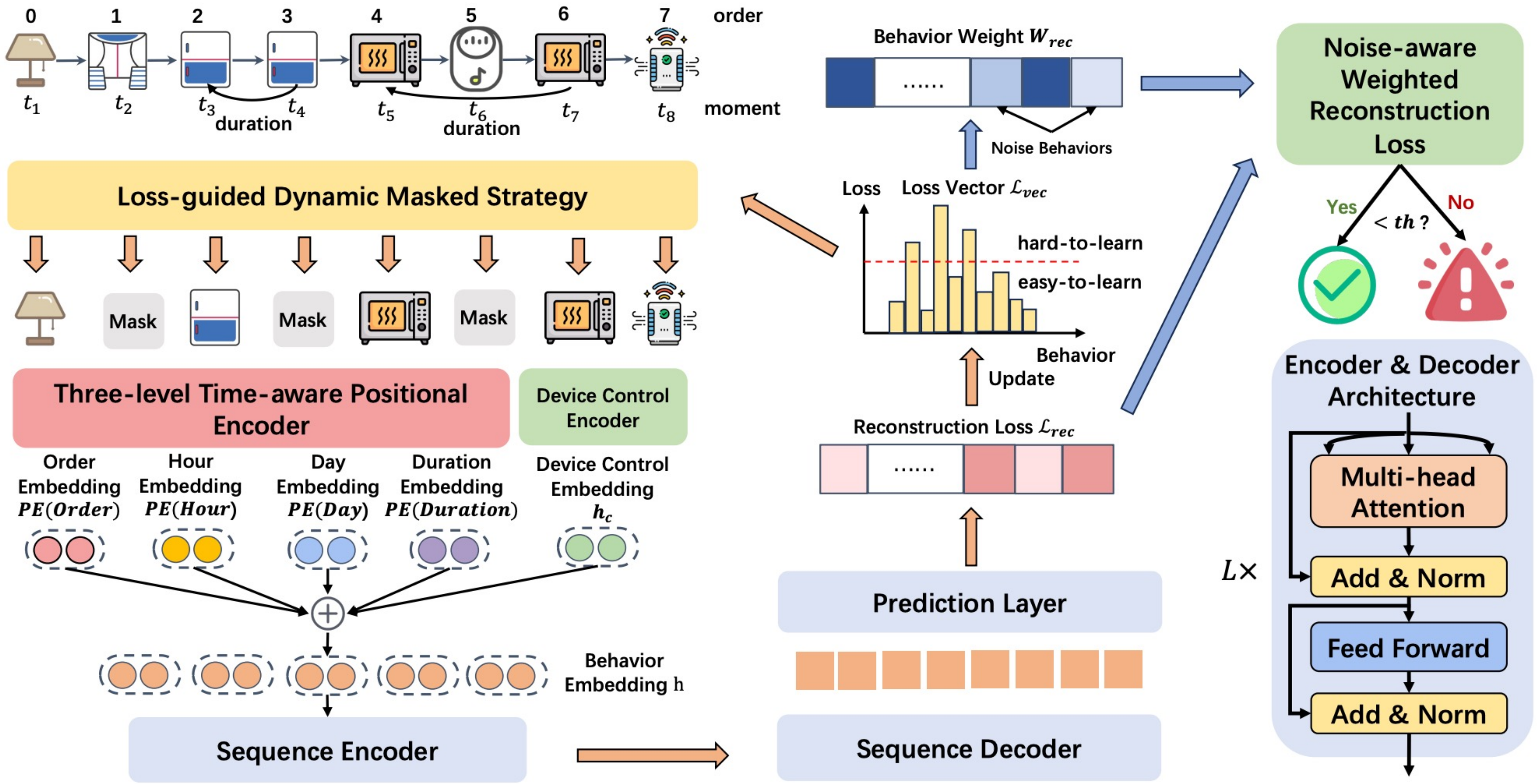


4 Overview

KDD2024 BARCELONA, SPAIN

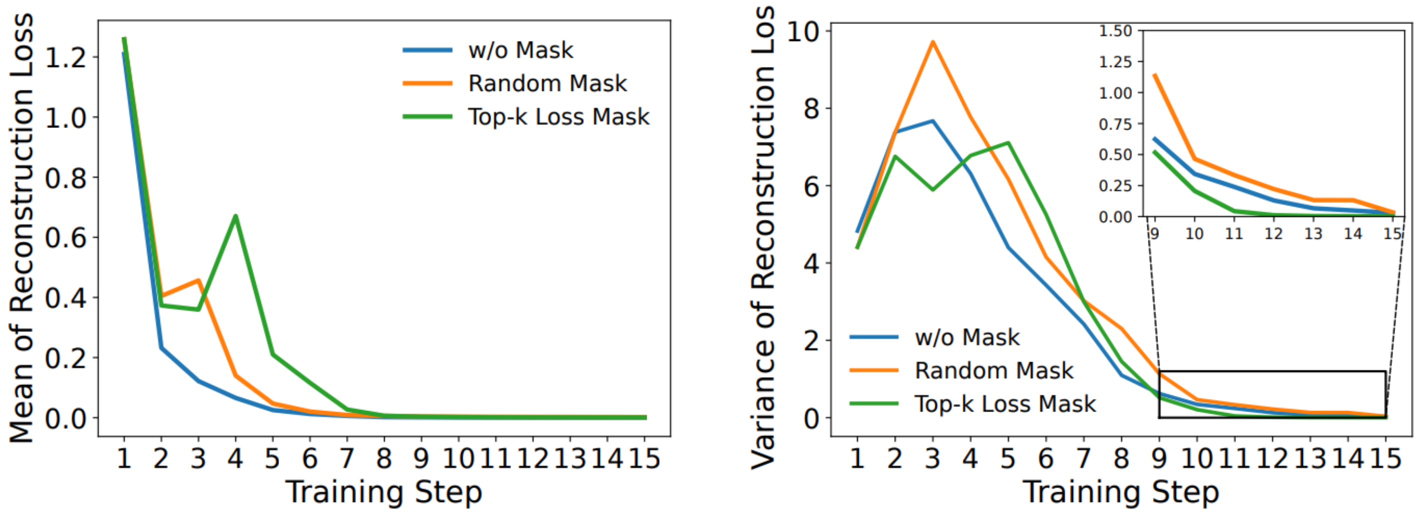
- We propose SmartGuard:
  - A novel approach for accurate user behavior anomaly detection in smart homes
- Idea #1: Loss-guided Dynamic Mask Strategy (LDMS)
  - To mask the behaviors with high reconstruction loss
  - To promote the model's learning of infrequent hard-to-learn behaviors
- Idea #2: Three-level Time-aware Position Embedding (TTPE)
  - To integrate temporal information into positional embedding for detecting temporal context anomalies.
  - Considering order-level, moment-level and duration-level information.
- Idea #3: Noise-aware Weighted Reconstruction Loss (NWRL)
  - To assign distinct weights to routine behaviors and noise behaviors, thereby mitigating the impact of noise behaviors.

4 Solution: SmartGuard



5 Loss-guided Dynamic Mask Strategy

- **Preliminary experiment： 1) without mask; 2) random mask; 3) top-k loss mask**



(a) Mean of reconstruction loss. (b) Variance of reconstruction loss.

- **Conclusion: 1) the model without mask shows the fastest convergence trend, whereas the loss of the model with mask fluctuates. 2) the model with top-*k* loss mask strategy shows lowest variance towards the end of training**

5 Loss-guided Dynamic Mask Strategy

KDD2024 BARCELONA, SPAIN

- **Method:**
  - **First, we encourage the model to learn the relatively easy task to accelerate convergence, i.e., behavior sequence reconstruction without mask.**
  - **Then, top-k loss mask strategy are adopt to encourage the model to learn the hard-to-learn behaviors with high reconstruction loss.**

$$\mathcal{L}_{\text{vec}}^{ep} = \left\{\ell_1, \ell_2, \ldots, \ell_c, \ldots, \ell_{|C|}\right\}, c \in C, \quad mask(i) = \begin{cases} 1, & \text{if } i \in sorted\_index[: \lfloor n \cdot r \rfloor] \\ 0, & \text{if } i \notin sorted\_index[: \lfloor n \cdot r \rfloor] \end{cases}, i \in [1, n],$$

$$\ell_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \ell_c^i, \quad sorted\_index = \operatorname{argsort}\left(\left\{\mathcal{L}_{\text{vec}}^{ep}(b_1), \mathcal{L}_{\text{vec}}^{ep}(b_2), \ldots.\mathcal{L}_{\text{vec}}^{ep}(b_n)\right\}\right),$$

6 Three-level Time-aware Positional Encoder

- **Three-level Time-aware Positional Encoder**
  - **Order-level:**

$$order \in [0, n-1]$$

  - **Moment-level: hour of day, day of week**
  - **Duration-level:**

$$duration_b = t(b) - t(b_{next})$$

  - **Positional Encoder:**

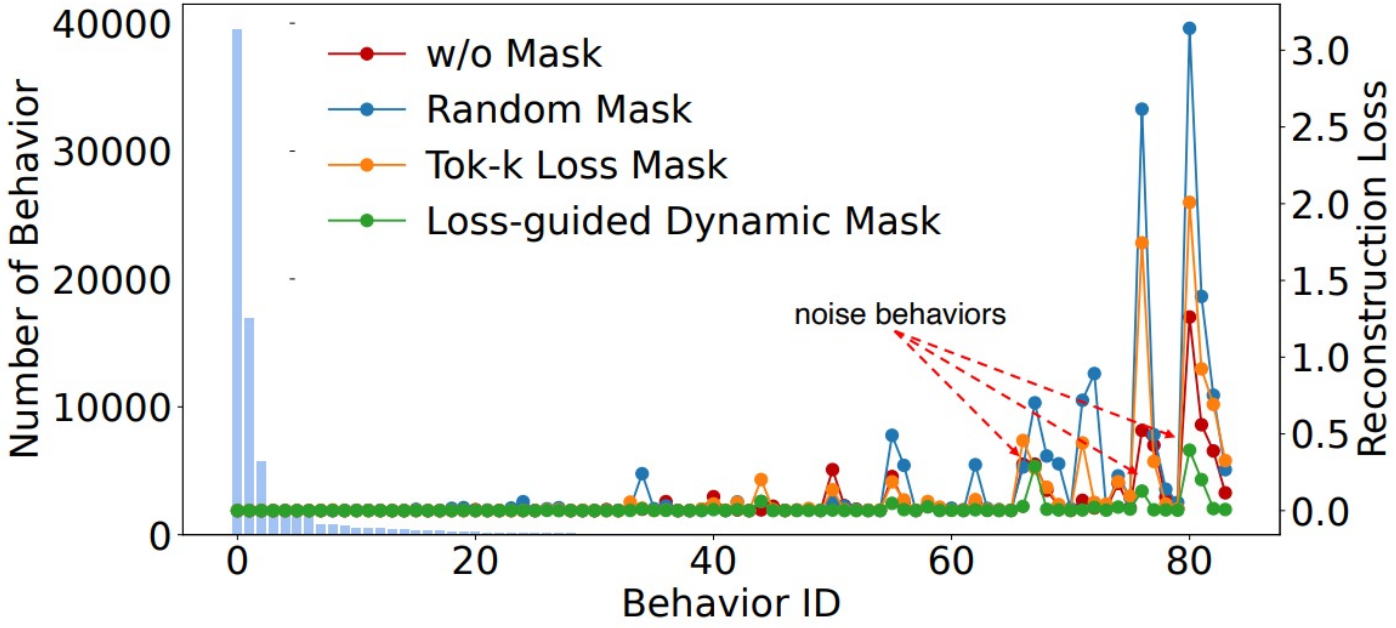
$$\overline{PE} = w_{order} \cdot PE(pos) + w_{hour} \cdot PE(hour) + w_{day} \cdot PE(day) + w_{dur} \cdot PE(duration),$$

$$PE_{(\cdot,2i)} = \sin\left(\cdot/10000^{2i/d}\right)$$

$$PE_{(\cdot,2i+1)} = \cos\left(\cdot/10000^{2i/d}\right)$$

7 Noise-aware Weighted Reconstruction Loss

- **Noise-aware Weighted Reconstruction Loss**



$$\mathcal{W}_{vec} = \text{sigmoid}\left(-\frac{\text{relu}(\mathcal{L}_{vec} - \mathbb{E}(\mathcal{L}_{vec}))}{\sqrt{\text{Var}(\mathcal{L}_{vec})} \cdot \mu}\right) \quad p_i = \frac{\mathcal{W}_{vec}(b_i)}{\sum_{j=1}^{n} \mathcal{W}_{vec}(b_j)} \quad score(s) = -\frac{1}{|s|}\sum_{i=1}^{|s|} p_i y_i \log \hat{y}_i.$$

8 Datasets

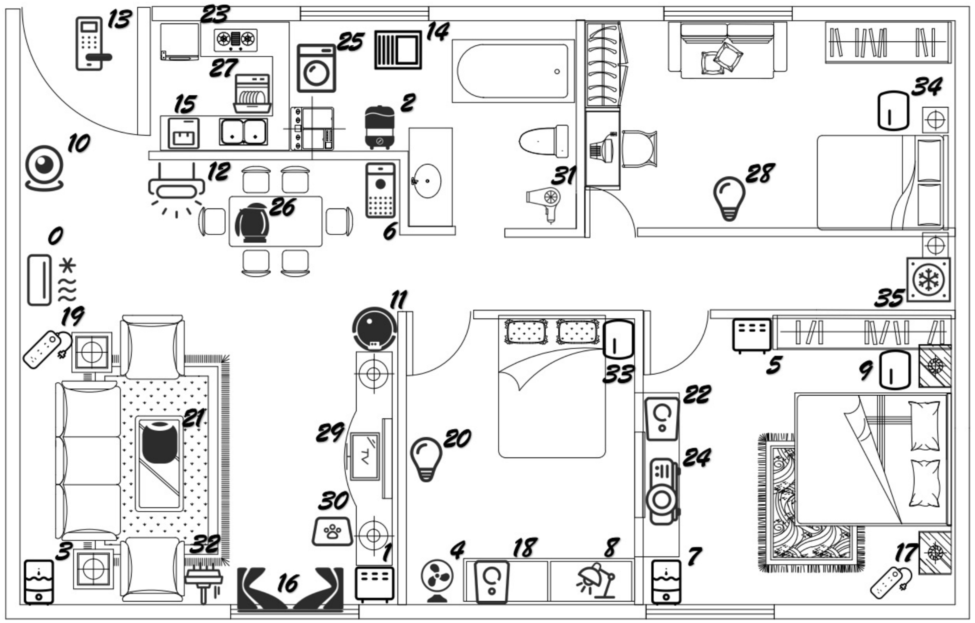
- **We use three real-world datasets to evaluate SmartGuard**
  - **SP/FR from public dataset, AN collected by ourselves.**
  - **Datasets are split into training/validation/testing with a ratio of 7:1:2.**
  - **10 types of anomaly behaviors**

Name	Time period (Y-M-D)	Sizes	# Devices	# Device controls
AN	2022-07-31~2022-08-31	1,765	36	141
FR	2022-02-27~2022-03-25	4,423	33	222
SP	2022-02-28~2022-03-30	15,665	34	234

Anomaly	Type	Anomaly	Type
Light flickering	SD	Open the airconditioner's cool mode in winter	DM
Camera flickering	SD	Open the window at midnight	DM
TV flickering	SD	Open the watervalve at midnight	DM
Open the window while smartlock lock	MD	Shower for long time	DD
Close the camera while smartlock lock	MD	Microwave runs for long time	DD

9 Datasets

- **Testbed**
  - **Three volunteers were recruited to simulate the typical daily activities of a standard family, assuming the roles of an adult male, an adult female, and a child. The experimental platform comprises a comprehensive selection of 36 popular market-available devices**



No.	Device	No.	Device	No.	Device
0	AC	12	LED	24	projector
1	heater	13	locker	25	washing_machine
2	dehumidifier	14	bathheater	26	kettle
3	humidifier_1	15	water_cooler	27	dishwasher
4	fan	16	curtains	28	bulb_1
5	standheater	17	outlet	29	TV
6	aircleaner	18	audio	30	pet_feeder
7	humidifier_2	19	plug	31	hair_dryer
8	desklight	20	bulb_2	32	window_cleaner
9	bedight_1	21	soundbox_1	33	bedlight_2
10	camera	22	soundbox_2	34	bedlight_3
11	sweeper	23	refrigerator	35	cooler

10 Baselines and Evaluation Metrics

- Baselines: we compare SmartGuard with 8 competitors
  - Local Outiler Factor (LOF)
  - Isolation Forest (IF)
  - 6thSense utilizes Naive Bayes to detect malicious behavior
  - Aegis utilizes a Markov Chain-based technique to detect malicious behavior.
  - OCSVM
  - Autoencoder
  - ARGUS designed an AE based on Gated Recurrent Units to detect IoT infiltration attacks.
  - TransformerAutoencoder (TransAE) uses self-attention mechanism in the encoder and decoder to achieve context-aware anomaly detection.
- Evaluation Metrics:
  - Precision, Recall, F1-Score
  - False Positive Rate, False Negative Rate

11 Questions

- **RQ1 (Performance).** Compared with other methods, does SmartGuard achieve better anomaly detection performance?
- **RQ2 (Ablation study).** How will model performance change if we remove key modules of SmartGuard?
- **RQ3 (Parameter study).** How do key parameters affect the performance of SmartGuard?
- **RQ4 (Interpretability study).** Can SmartGuard give reasonable explanations for the detection results?
- **RQ5 (Embedding space analysis).** Does SmartGuard successfully learn useful embeddings of behaviors and correct correlations between device controls and time?

Experimental Results

- **RQ1:** Compared with other methods, does SmartGuard achieve better performance?
- **A1:** SmartGuard can outperform competitors in many situations.

Dataset	Type	Metric	LOF	IF	6thSense	Aegis	OCSVM	DBSCAN	Glow	HomeGuardian	Tang	Anomaly Transformer	SSMCTB	Autoencoder	ARGUS	TransAE	SmartGuard(Ours)
		Recall	0.0275	0.4105	0.468	0.2902	0.5399	0.9307	0.4915	0.4092	0.6502	0.5699	0.719	0.9832	0.9858	0.9882	0.9986
	SD	Precision	0.4773	0.6305	0.584	0.5	0.6413	0.7518	0.8319	0.6285	0.7027	0.9689	0.9839	0.9999	0.9998	0.9934	0.9948
		F1 Score	0.0519	0.4972	0.5196	0.3672	0.5862	0.8318	0.6179	0.4956	0.6755	0.7177	0.8308	0.9915	0.9928	0.9908	0.9967
		Recall	0.0745	0.4039	0.5941	0.4431	0.6039	0.9451	0.2431	0.351	0.1892	0.2196	0.1745	0.5156	0.5666	0.6216	0.9745
	MD	Precision	0.76	0.5988	0.6516	0.5045	0.7163	0.6667	0.62	0.5793	0.7089	0.8889	0.9082	0.9531	0.9632	0.9635	0.9921
		F1 Score	0.1357	0.4824	0.6215	0.4718	0.6553	0.7818	0.3493	0.4371	0.2987	0.3522	0.2928	0.6692	0.7135	0.7557	0.9832
		Recall	0.0784	0.4373	0.3745	0.5647	0.351	0.9451	0.2686	0.4157	0.2838	0.2	0.1255	0.5196	0.5313	0.6078	0.9961
AN	DM	Precision	0.7407	0.6335	0.6749	0.5647	0.5408	0.6667	0.6432	0.6235	0.5526	0.8793	0.8767	0.9529	0.9611	0.9628	0.9922
		F1 Score	0.1418	0.5174	0.4817	0.5647	0.4257	0.7818	0.379	0.4988	0.375	0.3259	0.2196	0.6725	0.6843	0.7452	0.9941
		Recall	0.0961	0.3451	0.198	0.7804	0.4961	0.9431	0.2431	0.2941	0.1628	0.2353	0.5078	0.5137	0.5117	0.5294	0.998
	DD	Precision	0.7903	0.5641	0.7214	0.6419	0.7485	0.6662	0.62	0.5415	0.5213	0.8955	0.9664	0.9527	0.9597	0.9574	0.9923
		F1 Score	0.1713	0.4282	0.3108	0.7044	0.5967	0.7808	0.3493	0.3812	0.2481	0.3727	0.6658	0.6675	0.6675	0.6818	0.9951
		Recall	0.0509	0.4614	0.5769	0.4941	0.4466	0.9996	0.3316	0.4623	0.4342	0.2738	0.2535	0.6134	0.6317	0.6756	0.9925
	ALL	Precision	0.0731	0.6758	0.6503	0.5608	0.7759	0.6931	0.7146	0.6867	0.5009	0.9133	0.9372	0.9680	0.9738	0.9717	0.9931
		F1 Score	0.06	0.5484	0.6114	0.5254	0.5669	0.8186	0.453	0.5526	0.4652	0.4213	0.3991	0.7509	0.7663	0.797	0.9928
		Recall	0.3541	0.2444	0.2907	0.3916	0.5918	0.8404	0.6152	0.246	0.3036	0.8513	0.9999	0.9816	0.9796	0.9864	0.9979
	SD	Precision	0.7467	0.7242	0.7355	0.5406	0.7492	0.8216	0.8983	0.7178	0.9688	0.9808	0.9836	0.9999	0.9998	0.9978	0.9885
		F1 Score	0.4804	0.3655	0.4167	0.4542	0.6612	0.8308	0.7302	0.3664	0.4623	0.9115	0.9917	0.9907	0.9897	0.9921	0.9932
		Recall	0.4275	0.298	0.6069	0.6568	0.4384	0.8051	0.8947	0.298	0.1915	0.8175	0.5554	0.9726	0.9875	0.9782	0.9984
	MD	Precision	0.661	0.729	0.6599	0.5682	0.7503	0.8771	0.8954	0.7403	0.8418	0.9704	0.957	0.9999	0.9692	0.9967	0.9831
		F1 Score	0.5192	0.423	0.6323	0.6093	0.5534	0.8396	0.8950	0.4249	0.312	0.8874	0.7029	0.9861	0.9783	0.9874	0.9907
		Recall	0.3825	0.3191	0.5461	0.762	0.392	0.8145	0.9477	0.3191	0.1774	0.4111	0.5179	0.4952	0.6676	0.6529	0.9985
FR	DM	Precision	0.6553	0.7596	0.6971	0.6177	0.6675	0.8964	0.9007	0.7544	0.8023	0.9428	0.954	0.9489	0.9575	0.9621	0.9841
		F1 Score	0.483	0.4494	0.6124	0.6823	0.494	0.8535	0.9236	0.4485	0.2905	0.5725	0.6714	0.6508	0.7867	0.7779	0.9912
		Recall	0.3572	0.185	0.5358	0.9743	0.6267	0.8369	0.5605	0.1775	0.4526	0.1559	0.0636	0.4397	0.7398	0.6098	0.9981
	DD	Precision	0.5642	0.5805	0.6515	0.5874	0.6584	0.8271	0.7873	0.5764	0.9766	0.8118	0.6832	0.9507	0.9667	0.9668	0.9862
		F1 Score	0.4375	0.2806	0.588	0.7329	0.6422	0.832	0.6548	0.2715	0.6185	0.2616	0.1164	0.6013	0.8382	0.7479	0.9921
		Recall	0.3309	0.3595	0.4375	0.644	0.763	0.9996	0.7522	0.3518	0.5079	0.4317	0.2673	0.73	0.7526	0.8119	0.9982
	ALL	Precision	0.8183	0.8019	0.6657	0.5803	0.8062	0.6822	0.8829	0.8133	0.8383	0.9384	0.9077	0.9674	0.9683	0.9706	0.9858
		F1 Score	0.4712	0.4964	0.528	0.6105	0.784	0.811	0.8123	0.4912	0.6326	0.5914	0.413	0.8321	0.8469	0.8842	0.9919
		Recall	0.2197	0.2643	0.6979	0.1618	0.5332	0.8208	0.7425	0.2447	0.2809	0.7887	0.7359	0.9824	0.9795	0.9172	0.9862
	SD	Precision	0.7043	0.7138	0.7538	0.3271	0.7276	0.828	0.9042	0.7124	0.9672	0.9687	0.9718	0.9999	0.9999	0.9828	0.9801
		F1 Score	0.335	0.3857	0.7248	0.2165	0.6155	0.8244	0.8154	0.3643	0.4354	0.8695	0.8376	0.9911	0.9896	0.9489	0.9831
		Recall	0.2786	0.3399	0.6317	0.7445	0.384	0.7331	0.935	0.2818	0.1852	0.1949	0.5419	0.5645	0.9696	0.9936	0.9961
	MD	Precision	0.6587	0.727	0.6568	0.5986	0.7272	0.7326	0.8879	0.7126	0.8096	0.8361	0.9442	0.9547	0.9998	0.9796	0.9703
		F1 Score	0.3916	0.4632	0.644	0.6636	0.5026	0.7328	0.9109	0.4039	0.3015	0.3161	0.6886	0.7095	0.9845	0.9866	0.983
		Recall	0.278	0.3465	0.608	0.8122	0.5351	0.8576	0.8751	0.3443	0.2908	0.7731	0.3658	0.3074	0.5297	0.5451	0.9198
SP	DM	Precision	0.7895	0.8469	0.807	0.7427	0.7781	0.8571	0.9175	0.8512	0.9277	0.9681	0.9449	0.9533	0.9679	0.9631	0.9818
		F1 Score	0.4112	0.4918	0.6935	0.7759	0.6341	0.8573	0.8958	0.4903	0.4428	0.8597	0.5274	0.4649	0.6847	0.6962	0.9498
		Recall	0.2109	0.1763	0.5449	0.8001	0.8293	0.8623	0.376	0.1698	0.9883	0.1285	0.8453	0.6455	0.6455	0.6456	0.9961
	DD	Precision	0.5522	0.5151	0.7589	0.5538	0.6536	0.8429	0.6143	0.5208	0.7955	0.6272	0.9398	0.9494	0.9412	0.9394	0.9621
		F1 Score	0.3052	0.2627	0.6343	0.6545	0.7311	0.8525	0.4665	0.2561	0.8815	0.2134	0.8901	0.7685	0.7658	0.7653	0.9788
		Recall	0.258	0.3324	0.6649	0.6037	0.7318	0.7945	0.7888	0.3108	0.3042	0.3382	0.5287	0.5049	0.7275	0.7276	0.9667
	ALL	Precision	0.8123	0.7992	0.7366	0.6041	0.8546	0.7988	0.8826	0.8007	0.8817	0.8986	0.9526	0.9678	0.9813	0.9698	0.9762
		F1 Score	0.3916	0.4695	0.6989	0.6039	0.7884	0.7966	0.833	0.4478	0.4523	0.4915	0.68	0.6635	0.8356	0.8314	0.9714

Experimental Results

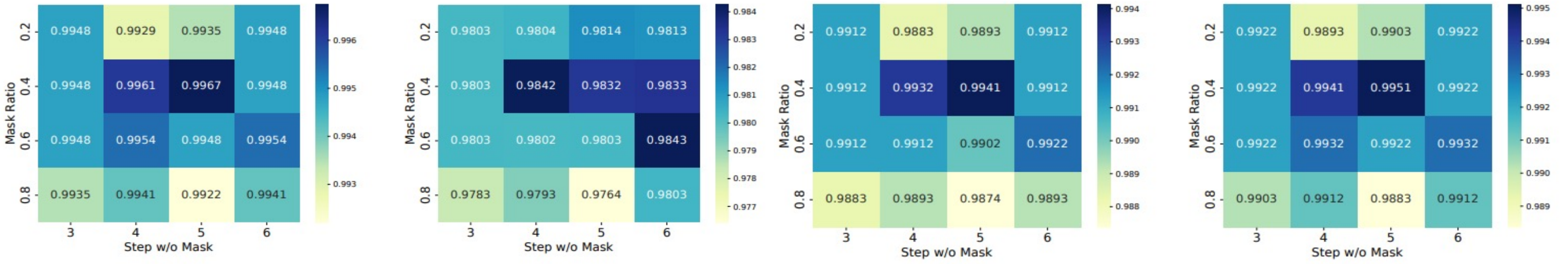
- **RQ2:** How will model performance change if we remove key modules of SmartGuard?
- **A2:** Each component of SmartGuard has a positive impact on results. The combination of all components brings the best results, which is much better than using any subset of the three components.

Table 4: The F1-Score of 5 variants (C_0-C_4) on AN dataset.

LDMS	TTPE	NWRL		SD	MD	DM	DD
X	X	X	C_0	0.9908	0.7557	0.7452	0.6818
Y	Y	X	C_1	0.9877	0.9708	0.9767	0.9817
Y	X	Y	C_2	0.9883	0.6716	0.6783	0.6799
X	Y	Y	C_3	0.9902	0.9766	0.9835	0.9855
Y	Y	Y	C_4	**0.9967**	**0.9832**	**0.9941**	**0.9951**

Experimental Results

- **RQ3:** How do key parameters affect the performance of SmartGuard?
- **A3:** SmartGuard achieves the optimal performance when $r = 0.4$ and $N = 5$.

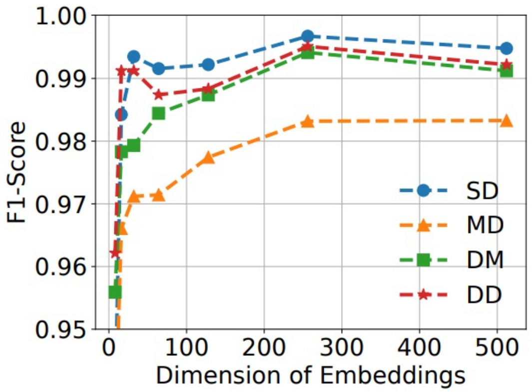


(a) SD Anomaly. (b) MD Anomaly. (c) DM Anomaly. (d) DD Anomaly.

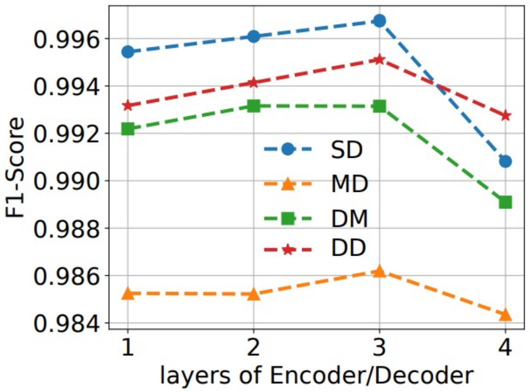
Figure 7: Performance under different mask ratio and step w/o mask on AN dataset.

Experimental Results

- **RQ3: How do key parameters affect the performance of SmartGuard?**
- **A3: SmartGuard achieves the optimal performance when embedding size = 256 and Layer = 3**



(a) Embedding dimension.



(b) Layers of encoder/decoder.

Figure 9: The influence of embedding dimension and encoder/decoder layer number on AN dataset.

Experimental Results

- **RQ4:** Can SmartGuard give reasonable explanations for the detection results?
- **A4:** SmartGuard delivers highly interpretable results

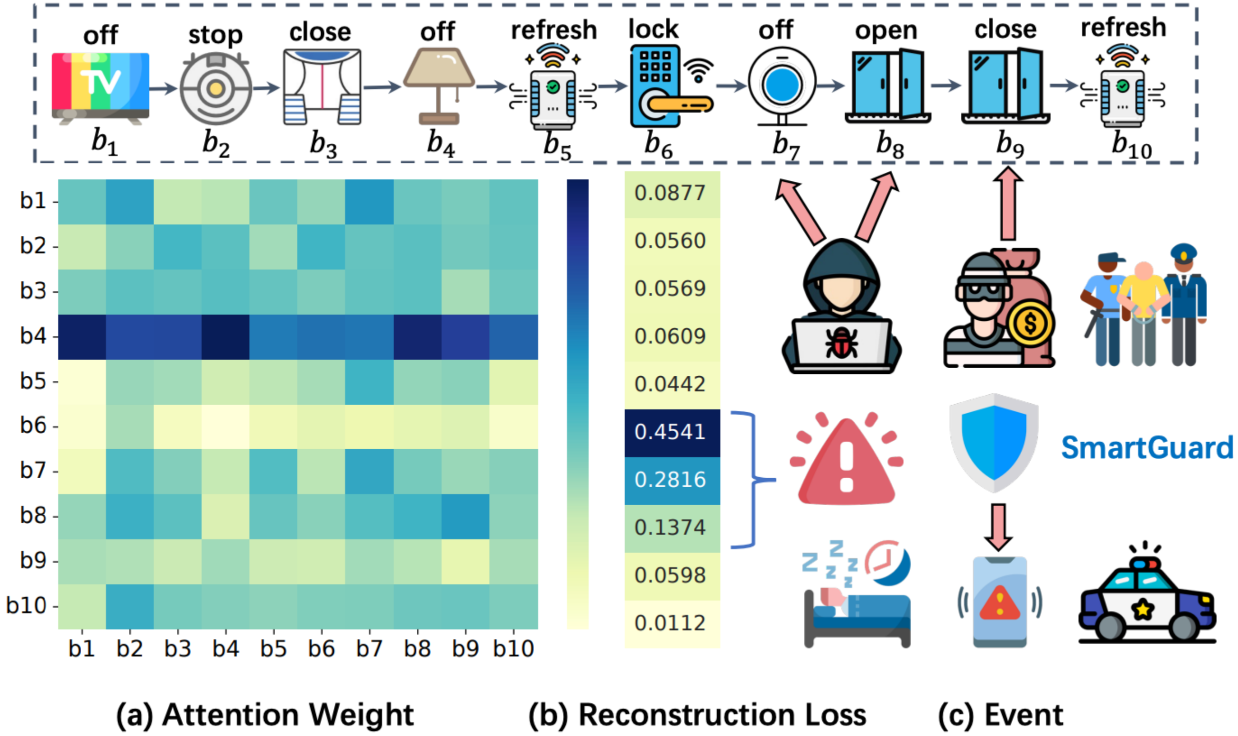
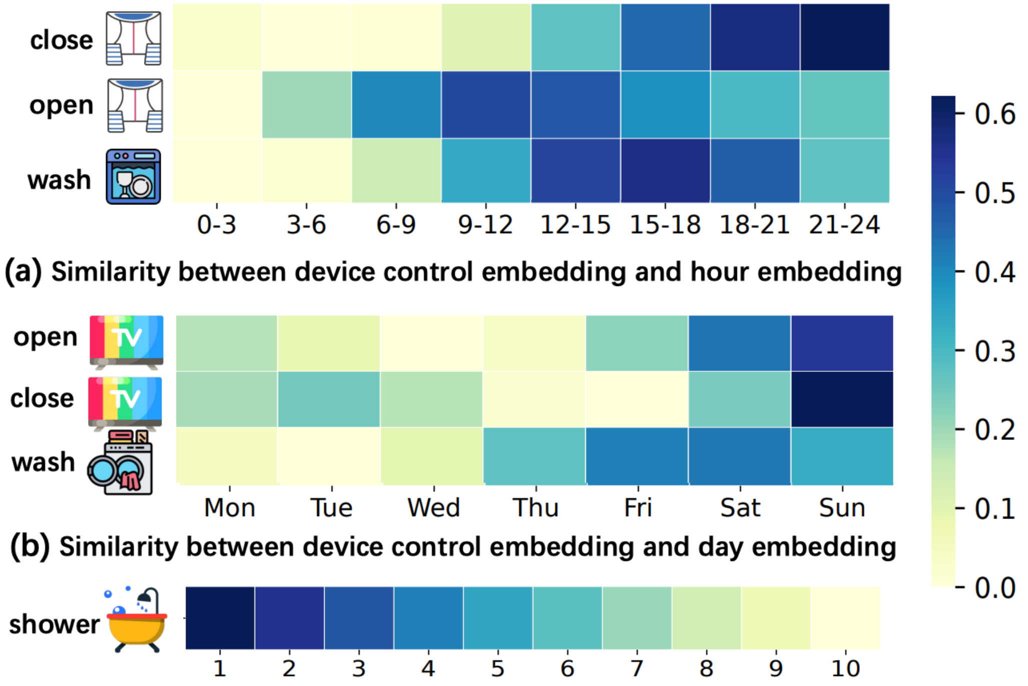


Figure 10: (a) Attention weights, (b) reconstrution loss and (c) the corresponding events.

Experimental Results

- **RQ5: Does SmartGuard successfully learn useful embeddings of behaviors and correct correlations between device controls and time?**
- **A5: SmartGuard can effectively mine the contextual relationship between behavior and time.**



(a) Similarity between device control embedding and hour embedding

(b) Similarity between device control embedding and day embedding

(c) Similarity between device control embedding and duration embedding

Thank you!

- **Speaker: Jingyu Xiao**
- **Codes:** https://github.com/xjywhu/SmartGuard
- **Homepage**: https://whalexiao.github.io/
- **Email**: jy-xiao21@mails.tsinghua.edu.cn