Potent but Stealthy: Rethink Profile Pollution Against Sequential Recommendation via Bi-Level Constrained Reinforcement Paradigm

Jiajie Su[1*], Zihan Nan[2*], Yunshan Ma[3], Xiaobo Xia[4,5], XiaoHua Feng[1], Weiming Liu[6], Xiang Chen[1], Xiaolin Zheng[1†], Chaochao Chen[1]

[1] Zhejiang University,
[2] Peking University
[3] Singapore Management University,
[4] National University of Singapore
[5] MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition,
University of Science and Technology of China
[6] Tiktok, Bytedance
{sujiajie, fengxiaohua, wasdnsxchen, xlzheng, zjuccc}@zju.edu.cn,
zhnan25@stu.pku.edu.cn, ysma@smu.edu.sg, xbx@nus.edu.sg, lwming95@gmail.com

Abstract

Sequential Recommenders, which exploit dynamic user intents through interaction sequences, are vulnerable to adversarial attacks. While existing attacks primarily rely on data poisoning, they require large-scale user access or fake profiles, thus lacking practicality. In this paper, we focus on the Profile Pollution Attack that subtly contaminates partial user interactions to induce targeted mispredictions. Previous PPA methods suffer from two limitations, *i.e.,* i) over-reliance on sequence horizon impact restricts fine-grained perturbations on item transitions, and ii) holistic modifications cause detectable distribution shifts. To address these challenges, we propose a constrained reinforcement driven attack CREAT that synergizes a bi-level optimization framework with multi-reward reinforcement learning to balance adversarial efficacy and stealthiness. We first develop a Pattern Balanced Rewarding Policy, which integrates pattern inversion rewards to invert critical patterns and distribution consistency rewards to minimize detectable shifts via unbalanced co-optimal transport. Then we employ a Constrained Group Relative Reinforcement Learning paradigm, enabling step-wise perturbations through dynamic barrier constraints and group-shared experience replay, achieving targeted pollution with minimal detectability. Extensive experiments demonstrate the effectiveness of CREAT.

Code&Datasets — https://github.com/SSndot/CREAT
Extended version — https://arxiv.org/abs/2511.09392

Introduction

Sequential Recommendation (SR) (Xie et al. 2022; Liu et al. 2023a; Su et al. 2023a) explores user evolving interests to make the next-item prediction. Although SRs are widely regarded as delivering trustworthy results, their sensitivity to sequential patterns renders them vulnerable to adversarial attacks (Nguyen et al. 2024; Du et al. 2024a). Recent research (Wang et al. 2023; Zhang et al. 2024) has mostly focus on *data poisoning* attacks, which manipulate SR by injecting a substantial amount of crafted sequences. But this attack relies on large-scale access to user accounts or the ability to create numerous fake profiles, which can be impractical in real-world scenarios. In this paper, we focus on a more targeted and stealthier attack strategy, *Profile Pollution Attack (PPA)*, which subtly contaminates partial individual user interaction histories to corrupt SR into a targeted misprediction on specific subtasks.

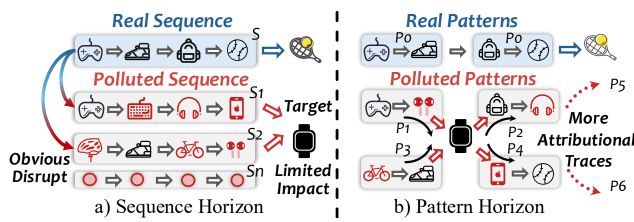


Figure 1: The motivation of CREAT.

Several previous studies attempted to conduct PPA against SR. One line of methods (Yue et al. 2021, 2022) crafts perturbations guided by gradient estimation on the attack loss then inserts influential items. Another line (Du et al. 2024b) leverages the influence function to measure the modification impact to training sequences on model parameters. As Fig 1 shows, these works largely follow a common paradigm, *i.e.,* assess the intensity of posterior attacks from the *sequence horizon*, which leverages the global structure of polluted sequences to modify recommender comprehension on user behaviors. Such a paradigm faces two limitations, **(i) Low attack intensity**. Relying on amplification of the whole polluted sequence effect confines the attack to

[*]These authors contributed equally.
[†]Xiaolin Zheng is the corresponding author.

user subsets with specific interests, failing to reshape model perception on fine-grained sequential transitions. **(ii) Subtle attack stealth**. Forcing the overall interests of sequences to align with attack targets induces a noticeable distributional shift. Since interference at the sequence horizon is a coarse-grained and user-level disruption, it often requires manipulating numerous sequences to achieve a significant effect, thus increasing the detection risk. These issues motivate us to a crucial question: ***How to exploit pivotal structures underlying the recommender collaborative modeling to maximize attack strength with minimal detectability?***

In light of this, we reformulate PPA into a bi-level optimization problem, wherein the *upper-level objective* seeks to maximize the utility of sequence perturbations, subject to a *lower-level constraint* enforcing a bounded degree of stealthiness in the crafted sequences. The upper-level formulation is grounded in a new theoretical concept, termed *pattern horizon*, which postulates that the model's prediction of the next item is inherently driven by an attributional trace over sequential pattern dependencies (Yin et al. 2024; Dang et al. 2025). As Fig 1 shows, inverting pivotal sequential patterns toward a target item enables a finer-grained exploration of how distinct perturbations propagate through the collaborative reasoning process of SR. Due to synergistic and cascading effects between patterns (Liu et al. 2023b), modifying a subset of patterns can be generalized to multiple similar patterns during the SR model's learning process, thereby amplifying the adversarial impact. At the lower level, in contrast to traditional sequence horizon perturbations that operate on holistic representations, the pattern-horizon-guided perturbations emphasize localized structural shifts. This allows the attacker to strategically calibrate pattern granularity and compositional balance, effectively regulating the distributional deviation between crafted and benign sequences. However, this bi-level formulation raises three key challenges, *i.e.,* ***Ch1: How to discern and reverse critical sequential patterns? Ch2: How to modulate the stealthiness of pattern-balanced perturbations? Ch3: How to synergistically optimize the coupled objectives across both levels?***

To tackle these challenges, we propose a Constrained REinforcement driven ATtack, termed `CREAT`, which leverages a group relative reinforcement learning constrained by stealth-aware conditions for targeted profile pollution. The key insight of `CREAT` lies in simulating the bi-level optimization problem with a multi-reward mechanism, *i.e.,* maximize the inversion effect of critical sequential patterns while minimize detection risk, thereby deriving an optimal pollution policy. For separately regulating the upper-level and lower-level objectives, we design a *Pattern Balanced Rewarding Policy (PBRP)*, integrating both inversion and consistency rewards to guide the perturbation. To uncover the most influential patterns (***Ch1***), we first develop the *pattern inversion reward* which identifies the optimal perturbation positions that simultaneously achieve maximal pattern-level semantic inversion and diversification. To regulate stealthiness of inverted patterns (***Ch2***), the *distribution consistency reward* adapts an unbalanced co-optimal transport to constrain the distributional shifts of polluted sequential representation from both sample and pattern aspects. Building upon the bi-level mechanism, we establish *Constrained Group Relative Reinforcement Learning (C-GRRL)*, which enables step-wise and self-reflective perturbations over polluted sequences. This paradigm consists of two stages, *i.e.,* critical pattern localization and constrained inversion optimization. In the *localization stage*, we train a sequence masker solely guided by the inversion reward, aiming to identify positions on target items that yield maximal adversarial impact. In the *constrained stage*, a dynamic barrier constraint that adaptively joins inversion and consistency rewards fine-tunes the masker, thereby aligning with the dual imperative of maximizing adversarial efficacy while preserving stealthiness (***Ch3***). Specifically, we employ a constrained group-relative policy within the bi-level optimization, which integrates a group-shared experience replay buffer and relative prioritized sampling, to accelerate the convergence toward optimal multi-step perturbations.

Main contributions are: (1) We revisit the PPA against SR into a bi-level optimization problem, and propose a novel framework with group relative constrained reinforcement learning. (2) We devise the PBRP policy, developing pattern inversion reward to extract influential patterns and distribution consistency reward to control pattern stealth. (3) We establish the C-GRRL paradigm, realizing step-wise and self-reflected perturbation optimization. (4) Extensive experiments demonstrate the effectiveness of `CREAT`.

Related Work

Sequential Recommendation. SR characterizes dynamic user intents by modeling behavioral sequences. Early work models sequential patterns with Markov Chain assumption (Rendle, Freudenthaler, and Schmidt-Thieme 2010). Later, Recurrent Neural Networks (Wu et al. 2017), Convolutional Neural Networks (Tang and Wang 2018), Graph Neural Networks (Wu et al. 2019; Zheng et al. 2020; Su et al. 2023b), and Transformers (Kang and McAuley 2018) are developed to model interests over interactions. Besides, unsupervised learning based models (Xie et al. 2022) extract more informative user patterns by deriving self-supervision signals. Inspired by generative models, a series of diffusion-based SRs (Yang et al. 2023) merge, leveraging diffusion generative capabilities to produce personalized content. A series of methods (Li et al. 2023; Liu et al. 2024b) utilize large language models to enhance the accuracy of SR. But the vulnerability of SR to adversarial attacks based on malicious sequences remains a significant security problem.

Adversarial Attacks in Recommender Systems. Adversarial attacks (Zhang et al. 2021a, 2022; Wang et al. 2023, 2024) on recommender systems can be categorized into: (1) *data poisoning* and (2) *profile pollution*. Data poisoning attacks (Zhang et al. 2020; Song et al. 2020; Tang, Wen, and Wang 2020; Huang et al. 2021; Wu et al. 2023) compromise recommenders by injecting fabricated user profiles, skewing model outputs toward adversarial objectives. Conversely, profile pollution attacks (PPA) (Yue et al. 2021; Zhang et al. 2021b; Fan et al. 2021; Lin et al. 2022) directly tamper with user interaction records, subtly distorting individual recommendation streams without requiring large-scale data infiltration. In this paper, we focus on the PPA and aim to manipulate the recommendation with targeted goals. Existing studies on PPA is divided into four types, which are respectively based on injection (Xing et al. 2013; Meng et al. 2014; Zhang et al. 2019), replacement (Yue et al. 2022), repetition (Tang and Wang 2018), and expert knowledge (Yang, Gong, and Cai 2017). The exploration of PPA against SR remains at a nascent stage. One branch of methods (Yue et al. 2021, 2022) generates perturbations by estimating gradients of attack loss, injecting impactful items into the sequences. SimAlter (Yue et al. 2021) appends adversarial items by extending the targeted fast gradient sign method from the continuous to discrete item space. Replace (Yue et al. 2022) typically utilizes the loss gradient to guide the selection of injected items. However, these gradient-based attacks are constrained by insufficient optimization due to single-step gradient descent (Madry et al. 2017). Another branch of work, like INFAttack (Du et al. 2024b), employs influence function to quantify how modifications affect the model parameters. But the influence computation chain introduces substantial complexity while its accuracy deteriorates with deeper backbones. Although these works promote PPA to some extent, they assess pollution strength from *sequence horizon*, which overlooks exploring fine-grained patterns, thus constraining attack effectiveness and increasing the detection risk.

Methodology

Problem Formulation

Profile Pollution Attack against SR. Let Φ_θ denote a sequential recommender with parameters θ, where users $u \in \mathcal{U}$ and items $v \in \mathcal{V}$ are represented by chronological interaction sequences $\mathbf{s}_u = [v_1, \ldots, v_L]$. The recommender is trained on a dataset $\mathcal{D} = \{\mathbf{s}_u \mid u \in \mathcal{U}\}$ with the next-item prediction loss $\mathcal{L}(\cdot)$. The profile pollution attack (PPA) aims to perturb a certain *subset* of training sequences $\mathcal{S} \subseteq \mathcal{D}$ into $\mathcal{S}'$ by replacing limited interactions to maximize the recommendation exposure of a target item $v^* \in \mathcal{V}$. For each polluted sequence, the amount of perturbations M is bound by $M \leq K$, where K is a small constant. We assume the attacker knows the model architecture and loss function, or can obtain a surrogate model through prior extraction. This assumption is justified by recent advances in recommendation model extraction (Yue et al. 2021; Wang et al. 2025; Liu et al. 2025), which demonstrate that black-box recommenders can be reliably approximated with limited or even no user data, resulting in surrogate models with similar hidden representations and output behavior. Formally, the objective of PPA is to construct perturbed sequences as:

$$\hat{\theta} = \arg\min_{\theta} \sum_{\mathbf{s}_u \in (\mathcal{D} \setminus \mathcal{S}) \bigcup \mathcal{S}'} \mathcal{L}(\mathbf{s}_u; \theta),$$

$$\widetilde{\mathcal{S}} = \arg\max_{\mathcal{S}'} \mathbb{E}_{u \sim \mathcal{U}} \left[\mathrm{ER}(v^* \mid \Phi_{\hat{\theta}}(\mathbf{s}_u)) \right].$$

Here,$\hat{\theta}$ denotes the recommender parameters after pollution and $\widetilde{\mathcal{S}}$ indicates the optimal polluted sequences that invert the recommender training to maximize the exposure ratio $\mathrm{ER}(v^* \mid \Phi_{\hat{\theta}}(\mathbf{s}_u))$ of v^* in recommendation lists.

Pattern Balanced Rewarding Policy

Perturbation Masker. Given a sequential recommender Φ_θ trained on interaction sequences $\mathcal{D}$, we design a perturbation masker $\mathcal{M}_\psi$ to identify optimal positions in a subset of training sequences $\mathcal{S} \subseteq \mathcal{D}$ for replacing items with the target item v^*, under a perturbation budget $M \leq K$. Formally, for a sequence $\mathbf{s} = [v_1, \ldots, v_L]$, the masker generates a binary mask $\mathbf{m} \in \{0, 1\}^L$ through a *step-wise reinforcement learning process*, where $\mathbf{m}_t = 1$ indicates replacing v_t with v^*. The perturbed sequence is constructed iteratively as:

$$\mathbf{s}'^{(i)} = \mathbf{s}'^{(i-1)} \odot (1 - \mathbf{m}^{(i)}) + v^* \cdot \mathbf{m}^{(i)},$$

where $\mathbf{m}^{(i)}$ is the mask vector at step i, and $\odot$ denotes element-wise multiplication. Unlike traditional perturbation, our masker follows a pattern balanced rewarding policy. At each step i, the masker selects the next position to perturb based on the current state $\mathbf{s}'^{(i-1)}$, and receives the reward based on the adversarial impact of the perturbation.

Pattern Inversion Reward. To amplify the attack effect, we propose a *Pattern Inversion Reward* that guides the masker to identify sub-pattern positions whose semantic distributions are notably different from that of the target item. By inserting the target item adjacent to these semantically divergent sub-patterns, the attacker can construct *spurious attributional paths* that link diverse user intent patterns to the target item. This misleads the recommender into falsely associating varied behavioral cues with the target, thereby increasing its exposure. This reward operates on two complementary dimensions, *i.e.,* the directionality and diversity of inversion pathways. **For the directionality of inversion**, we encourage the masker to maximize the semantic distance between the target item with both historical and future sequential contexts. Let $T^{(i)} = \{t_1, \ldots, t_i\}$ denote perturbed positions up to step i. For each $t_j \in T^{(i)}$, we compute embeddings of the predecessor $S^p_{t_j} = \mathbf{s}'^{(i)}_{[1:t_j-1]}$ and successor $S^f_{t_j} = \mathbf{s}'^{(i)}_{[t_j+1:L]}$ using the representation encoder φ_{rec} of Φ_θ. The directionality reward at step i is:

$$R^{(i)}_{\text{dir}} = \sum_{j=1}^{i} \left[D\left(\varphi_{\text{rec}}(S^p_{t_j}), \varphi_{\text{rec}}(v^*) \right) + D\left(\varphi_{\text{rec}}(v^*), \varphi_{\text{rec}}(S^f_{t_j}) \right) \right],$$

where $D(\cdot, \cdot)$ is the Euclidean distance. **For the diversity of inversion**, we refine the strategy by leveraging the synergistic effects among attack patterns. This inversion reward is designed to enhance the divergence among attack modes, ensuring heterogeneous attack paths and reduce detection risk. At step i, let $\mathcal{Y}^{(i)}$ be the set of subsequences in $\mathbf{s}'^{(i)}$ that exclude target items, we map each subsequence $y_p \in \mathcal{Y}^{(i)}$ to a unit-norm prototype as $\tilde{\varphi}(y_p) = \varphi_{\text{rec}}(y_p)/\|\varphi_{\text{rec}}(y_p)\|$. Then we form the diversity reward with the Gram matrix $G^{(i)}_y$:

$$G^{(i)}_y = \left[\tilde{\varphi}(y_k)^\top \tilde{\varphi}(y_l) \right]^{|\mathcal{Y}^{(i)}|}_{k,l=1}, \; R^{(i)}_{\text{div}} = \log\det\left(G^{(i)}_y \right).$$

Distribution Consistency Reward. To ensure the stealthiness of perturbed sequences, we introduce the *Distribution Consistency Reward*, which constrains the deviation between polluted sequences and their original counterparts in both instance-level and pattern-level semantics. We ground

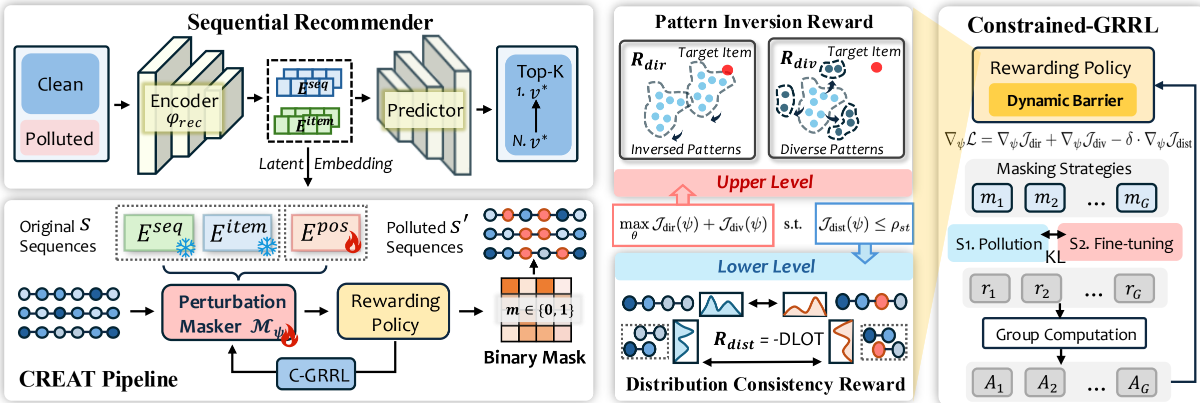


Figure 2: The proposed PPA framework of `CREAT`, which consists of three components, *i.e.,* the perturbation masker, pattern balanced rewarding policy with inversion and consistency reward, and constrained group relative reinforcement learning.

this reward in a dual-level co-optimal transport (`DLOT`) optimization (Tran et al. 2023), which provides a principled measure of distributional shifts by simultaneously aligning global sequence and local transitional patterns. At step i, let $\mathbf{s}$ denote the original sequence and $\mathbf{s}'^{(i)}$ the perturbed sequence. For the sequence-level, we obtain representations from the recommender encoder as $\mathbf{h}_{\text{orig}} = \varphi_{\text{rec}}(\mathbf{s})$ and $\mathbf{h}^{(i)}_{\text{pert}} = \varphi_{\text{rec}}(\mathbf{s}'^{(i)})$. For the pattern-level, the sets $\mathbf{p}_{\text{orig}} = \{\varphi_{\text{rec}}(\mathbf{s}_{[t:t+k]})\}_{t=1}^{L-k}$ and $\mathbf{p}_{\text{pert}} = \{\varphi_{\text{rec}}(\mathbf{s}'^{(i)}_{[t:t+k]})\}_{t=1}^{L-k}$ contain k-gram pattern embeddings derived from sliding windows over $\mathbf{s}$ and $\mathbf{s}'^{(i)}$. Then we construct the sequence-pattern spaces from the dual-level representations:

$$\mathbb{X}_{\text{orig}} = (\mathbf{h}_{\text{orig}}, \mathbf{p}_{\text{orig}}, \xi_{\text{orig}}), \quad \mathbb{X}^{(i)}_{\text{pert}} = (\mathbf{h}^{(i)}_{\text{pert}}, \mathbf{p}^{(i)}_{\text{pert}}, \xi_{\text{pert}}),$$

where ξ_{orig} and ξ_{pert} are scalar functions that define the sample-feature interactions. Unlike traditional balanced OT (Cuturi 2013; Flamary et al. 2021; Liu et al. 2022) and unbalanced OT (Pham et al. 2020; Séjourné, Vialard, and Peyré 2022), we incorporate two transport plans in `DLOT`, *i.e.,* π^s aligns entire sequences while π^f aligns intra-sequence patterns. The optimization of `DLOT` is established:

$$\texttt{DLOT} = \inf_{\pi^s,\pi^f} \iint |\xi_{\text{orig}}(\mathbf{h}_{\text{orig}}, \mathbf{p}_{\text{orig}}) - \xi_{\text{pert}}(\mathbf{h}^{(i)}_{\text{pert}}, \mathbf{p}^{(i)}_{\text{pert}})|^p d\pi^s d\pi^f$$
$$+ \sum_{j=1}^{2} \lambda_j \text{KL}\left(\pi^s_{\#j} \otimes \pi^f_{\#j} \| \mu^s_j \otimes \mu^f_j\right).$$

To tolerate partial mass mismatch and enhance robustness, KL-divergence terms penalize deviations of the marginal distributions of π^s and π^f from their empirical counterparts μ^s and μ^f. A mass constraint $m(\pi^s) = m(\pi^f)$ is imposed to ensure transport consistency across levels. We present the derivation details in the extended version. The consistency reward is set as `DLOT` distance between as:

$$R^{(i)}_{\text{dist}} = -\texttt{DLOT}_{\lambda_1,\lambda_2}\left(\mathbb{X}_{\text{orig}}, \mathbb{X}'^{(i)}_{\text{pert}}\right),$$

Constrained Group Relative Learning

To jointly optimize the multi-objective rewards under stealth-aware constraints, we introduce a two-stage optimization paradigm, which enables progressive learning of perturbation strategies by first distilling attack-effective behaviors and then aligning with distributional constraints.

Constrained Reinforcement with Dynamic Barrier. We model the perturbation decision process as a bi-level reinforcement learning problem, where the policy aims to maximize the inversion rewards while satisfying stealth constraints. We formulate the bi-level objective for each mask step, where each reward or constraint term is defined as the expected discounted return:

$$\max_{\theta} \mathcal{J}_{\text{dir}}(\psi) + \mathcal{J}_{\text{div}}(\psi) \quad \text{s.t.} \quad \mathcal{J}_{\text{dist}}(\psi) \leq \rho_{st}$$

$$\mathcal{J}_r(\psi) = \mathbb{E}_{\tau\sim\pi_\psi}\left[\sum_{t=0}^{T} \gamma^t R_r(s_t, a_t, s_{t+1}).\right], r \in \{\text{dir}, \text{div}, \text{dist}\}$$

Here, s_t denotes the state at step t representing the partially perturbed sequence $\mathbf{s}'^{(i)}$, a_t represents the action of masking, and γ^t is the discount factor. The constraint $\mathcal{J}_{\text{dist}}(\psi) \leq \rho_{st}$ enforces stealthiness by bounding the expected distribution consistency reward. It is worth noting that ρ_{st} is a threshold dynamically derived from the distribution of trajectories in group relative policy, which ensures the stealth bound adapts to the evolving policy and group-wise sequence characteristics. To solve this problem, we rewrite the problem as a min-max Lagrangian formulation:

$$\mathcal{L}(\psi, \delta) = \mathcal{J}_{\text{dir}}(\psi) + \mathcal{J}_{\text{div}}(\psi) - \delta \cdot (\mathcal{J}_{\text{dist}}(\psi) - \rho_{st}), \quad \delta \geq 0$$

We compute the policy gradients for each term via the policy gradient theorem, and further derive the policy gradient $\nabla_\psi \mathcal{L}$ for perturbation masker $\mathcal{M}_\psi$:

$$\nabla_\psi \mathcal{L} = \nabla_\psi \mathcal{J}_{\text{dir}} + \nabla_\psi \mathcal{J}_{\text{div}} - \delta \cdot \nabla_\psi \mathcal{J}_{\text{dist}}.$$

Different from static constraints, we tend to dynamically adjust the penalty term, based on real-time constraint violation and gradient alignment to ensure a balanced optimization. When gradients of $\mathcal{J}_{\text{dir/div}}$ and $\mathcal{J}_{\text{dist}}$ conflict, the numerator δ reduces to prioritize attack efficacy. Severe stealth violations increase δ to suppress detectable perturbations.

Based on the dynamic barrier design (Gong and Liu 2021), we can give out the closed-form expression as

$$\delta = \left[\frac{\mathcal{J}_{\text{dist}} - \rho_{st} - \nabla_\psi \mathcal{J}_{\text{dist}}^\top \nabla_\psi (\mathcal{J}_{\text{dir}} + \mathcal{J}_{\text{div}})}{\|\nabla_\psi \mathcal{J}_{\text{dist}}\|^2 + \kappa} \right]_+ .$$

$\kappa > 0$ is for numerical stability. Then policy gradients are

$$\nabla_\psi (\mathcal{J}_{\text{dir}} + \mathcal{J}_{\text{div}}) = \mathbb{E}_\tau \left[\sum_t \nabla_\psi \log \pi_\psi(a_t|s_t) \cdot (\hat{A}_{R_{dir}} + \hat{A}_{R_{div}}) \right],$$
$$\nabla_\psi \mathcal{J}_{\text{dist}} = \mathbb{E}_\tau \left[\sum_t \nabla_\psi \log \pi_\psi(a_t|s_t) \cdot \hat{A}_{R_{dist}}(s_t, a_t) \right].$$

This ensures the policy prioritizes stealthiness only when constraints are violated, balancing efficacy and detectability without sacrificing convergence stability. The perturbation policy is updated using gradient ascent on the Lagrangian as $\psi_{t+1} = \psi_t + \eta \left(\nabla_\psi \mathcal{J}_{\text{dir}} + \nabla_\psi \mathcal{J}_{\text{div}} - \delta \cdot \nabla_\psi \mathcal{J}_{\text{dist}} \right)$, where η denotes the learning rate.

Group Relative Optimization. To stabilize optimization and accelerate convergence, inspired by the GRPO paradigm (Liu et al. 2024a), we further reform the constrained reinforcement learning with a group relative strategy. Formally, we divide the training into two stages, i.e., *localization stage* and *constrained fine-tuning stage*. In the localization stage, we train the perturbation masker using only the pattern-level rewards R_{dir} and R_{div}. This stage is analogous to supervised fine-tuning, allowing the policy to explore effective inversion behaviors in an unconstrained space. We obtain the masker policy with the pure attacking goal as π_{att}. In the constrained stage, we introduce the distribution consistency reward R_{dist} as a constraint and perform constrained reinforcement learning with with dynamic barrier, guided by a GRPO-based surrogate objective. In each masking step, we sample a group of G trajectories $\{o_{i,t}, r_{i,t}\}_{i=1}^G$ under the current policy π_ψ, where $o_{i,t}$ is the trajectory and $r_{i,t}$ is the reward aggregated from multi-objective signals. We obtain the group-wise baseline $\mu_{G,t} = \frac{1}{G}\sum_{i=1}^G r_{i,t}$ and reward standard deviation $\sigma_{G,t} = \sqrt{\frac{1}{G-1}\sum_{i=1}^G (r_{i,t} - \mu_{G,t})^2}$. Besides, we can also set the dynamic stealthy threshold for the constraint reward as $\rho_{st} = \mu_G^{dist} + \lambda_{st} \cdot \sigma_G^{dist}$. We construct the normalized advantage estimate $\hat{A}_{i,t} = \frac{r_{i,t} - \mu_{G,t}}{\sigma_{G,t} + \epsilon}$, where ϵ is a positive constant for numerical stability. Finally, the policy is optimized with a clipped surrogate objective:

$$\mathcal{J}(\psi) = \mathbb{E}_\tau \left[\frac{1}{G} \sum_{i=1}^G \frac{1}{\|o_i\|} \sum_{t=1}^{\|o_i\|} \min \left(\frac{\pi_\psi(o_{i,t})}{\pi_{\text{att}}(o_{i,t})} \hat{A}_{i,t}, \right.\right.$$
$$\left.\left. \text{clip}\left(\frac{\pi_\psi(o_{i,t})}{\pi_{\text{att}}(o_{i,t})}, 1-\epsilon, 1+\epsilon \right) \hat{A}_{i,t} \right) \right]$$

This group relative normalization enhances training stability by reducing sensitivity to outliers and scales advantage estimates adaptively across trajectories. Combining this with bi-level optimization, C-GRRL supports efficient and stable learning of stealthy, high-impact perturbation strategies.

Empirical Study

Experimental Setups

Datasets & Backbone Recommender Systems. Following (Yue et al. 2022), we evaluate on three datasets, i.e. ML-1M, ML-20M, and Amazon Beauty, and preprocess them according to (Sun et al. 2019). We choose two representative SR as backbones, i.e., NARM and BERT4Rec. The dataset and backbone details are shown in the extended version.

Comparison Methods. For fair comparison, we set all attack SOTAs under the same setting, *i.e.,* PPA against white-box SR: 1) **Popular Attack** injects sequences with target items and filler popular items. 2) **Random Attack** randomly selects filler items. 3) **SimAlter Attack** (Yue et al. 2021) constructs adversarial sequences with semantic associations. 4) **Replace Attack** (Yue et al. 2022) identifies fragile items by gradients and replaces them with adversarial candidates. 5) **SSL Attack** (Wang et al. 2023) realizes data poisoning against SR with generative adversarial networks, and we adapt it into our white-box setting for fair comparison.

Evaluation Protocols. We evaluate with the exposure ratio of the target item, which is measured with Hit Ratio (HR), Normalized Discounted Cumulative Gain (NDCG), and Mean Reciprocal Rank (MRR).

Implementation Details. We present the implementation details and parameter settings in the extended version.

Experimental Results and Analysis

Overall Attack Performance (RQ1). We evaluate the attack performance of `CREAT` and baselines on NARM and Bert4Rec for three datasets. From Table 1, we find: 1) `CREAT` significantly enhances target item exposure across diverse scenarios, delivering nearly ten times higher exposure compared to *Pure*. Specifically, it outperforms the best baselines by over 20% across all metrics. 2) Dense datasets favor appending strategies (*e.g.,* Random, Popular), whereas short sequences exhibit heightened susceptibility to substitution strategies (*e.g.,* SimAlter, Replace). The divergence emerges as longer sequences intensify positional bias in attention mechanisms, while shorter sequences' dependence on sparse high-impact features elevates substitution risks. 3) On the most challenging dataset, *i.e., Beauty*, all baselines exhibit poor attack performance, *e.g.,* the well-performed SOTA *SSL Attack* only achieve 0.0109 on HR@10. In contrast, `CREAT` achieved 0.2601 on *HR@10* and 0.1038 on *NDCG@10*, highlighting the superior efficacy of deep pattern extraction over gradient-based strategies.

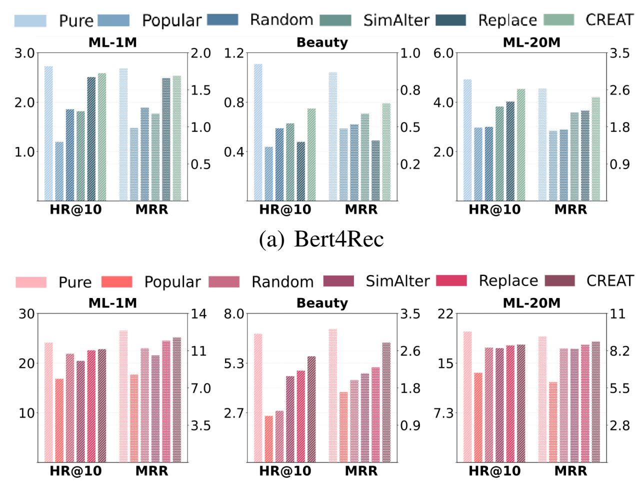
Ablation Studies (RQ2). We design variants as: (a) *w/o dir* removes directionality reward. (b) *w/o div* removes diversity reward. (c) *w/o pv* removes the whole pattern inversion reward. (d) *w/o dist* removes distribution consistency reward. From Table 2: 1) *w/o dir* and *w/o div* both outperform *w/o pv*, thus either the directionality or diversity reward contributes to the promotion. 2) `CREAT` performs better than *w/o dir* and *w/o div*, so relying solely on either pattern reward has its limitations. Utilizing only R_{dir} leads to the clustering of inverted patterns in the representation space, while solely on R_{div} results in insufficient disturbance of the sub-patterns.

Attack	NARM						Bert4Rec					
	HR@1	HR@5	HR@10	NDCG@5	NDCG@10	MRR	HR@1	HR@5	HR@10	NDCG@5	NGCG@10	MRR
ML-1M												
Pure	0.0038	0.0109	0.0174	0.0075	0.0096	0.0098	0.0000	0.0076	0.0128	0.0034	0.0051	0.0062
Popular	0.0128	0.0428	0.0695	0.0280	0.0365	0.0406	0.0105	0.0632	0.1065	0.0372	0.0515	0.0452
Random	0.0104	0.0326	0.0560	0.0217	0.0291	0.0333	0.0043	0.0305	0.0560	0.0173	0.0255	0.0254
SimAlter	0.0143	0.0459	0.0658	0.0306	0.0369	0.0360	0.0102	0.0385	0.0563	0.0271	0.0328	0.0324
Replace	0.0164	0.0575	0.1027	0.0369	0.0515	0.0514	0.0155	0.0569	0.0938	0.0365	0.0493	0.0512
SSLAttack	0.0148	0.0565	0.1069	0.0350	0.0510	0.0564	0.0145	0.0578	0.0970	0.0356	0.0483	0.0453
CREAT	**0.0305**	**0.0855**	**0.1428**	**0.0580**	**0.0765**	**0.0749**	**0.0187**	**0.0811**	**0.1492**	**0.0448**	**0.0666**	**0.0593**
Beauty												
Pure	0.0003	0.0015	0.0033	0.0008	0.0014	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000	0.0022
Popular	0.0005	0.0018	0.0065	0.0009	0.0023	0.0032	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026
Random	0.0010	0.0051	0.0103	0.0029	0.0042	0.0049	0.0000	0.0000	0.0001	0.0000	0.0000	0.0031
SimAlter	0.0019	0.0093	0.0178	0.0055	0.0082	0.0116	0.0000	0.0013	0.0048	0.0006	0.0017	0.0068
Replace	0.0027	0.0151	0.0608	0.0070	0.0216	0.0228	0.0000	0.0000	0.0000	0.0000	0.0000	0.0034
SSLAttack	0.0095	0.0354	0.0751	0.0226	0.0296	0.0270	0.0022	0.0074	0.0109	0.0049	0.0060	0.0097
CREAT	**0.0424**	**0.1080**	**0.1504**	**0.0759**	**0.0895**	**0.0801**	**0.0076**	**0.1191**	**0.2601**	**0.0582**	**0.1038**	**0.0773**
ML-20M												
Pure	0.0020	0.0108	0.0215	0.0063	0.0097	0.0136	0.0001	0.0015	0.0046	0.0007	0.0018	0.0070
Popular	0.0516	0.1516	0.2207	0.1024	0.1247	0.1074	0.0201	0.1542	0.2431	0.0905	0.1171	0.0945
Random	0.0411	0.1287	0.2080	0.0851	0.1096	0.1019	0.0193	0.1062	0.1634	0.0746	0.0930	0.0892
SimAlter	0.0318	0.1204	0.2048	0.0756	0.1036	0.0955	0.0235	0.0815	0.1398	0.0523	0.0710	0.0662
Replace	0.0356	0.1243	0.1906	0.0799	0.1012	0.0921	0.0084	0.0455	0.0872	0.0269	0.0402	0.0420
SSLAttack	0.0437	0.1202	0.1743	0.0825	0.0999	0.0893	0.0217	0.0723	0.1135	0.0473	0.0605	0.0583
CREAT	**0.0585**	**0.1905**	**0.3041**	**0.1243**	**0.1608**	**0.1392**	**0.0241**	**0.1621**	**0.2628**	**0.0932**	**0.1255**	**0.1025**

Table 1: The overall performance on three datasets. The best results are boldfaced, and the second-best results are underlined. All improvements are significant with p-value < 0.05 based on t-tests.

Attack	NARM				Bert4Rec			
	ML-1M		Beauty		ML-1M		Beauty	
	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
w/o dir	0.0117	0.0024	0.5105	0.3663	0.0140	0.0056	0.6307	0.4812
w/o div	0.0143	0.0055	0.5547	0.4318	0.0151	0.0063	0.6603	0.5941
w/o pv	0.0025	0.0013	0.5051	0.3538	0.0053	0.0019	0.5044	0.2859
w/o dist	0.0271	0.0116	0.6291	0.5221	0.0212	0.0083	0.7461	0.7034
CREAT	**0.0157**	**0.0063**	**0.5633**	**0.4439**	**0.0186**	**0.0068**	**0.6885**	**0.6285**

Table 2: Ablation studies on each reward.



(a) Bert4Rec

(b) NARM

Figure 3: Effects on recommendation accuracy (%).

3) The constrained reward R_{dist} decreases attack effectiveness on both datasets, but its impact varies. This is due to behavioral characteristics, *i.e.,* high randomness in user behavior makes patterns more fragile, while stable preferences allow effective feature injection even under covert conditions.

Stealth Verification (RQ3). We evaluate the stealthiness of CREAT from three aspects. *First*, we present the performance of each attack under the defense SOTA methods in the extended version, where we prove CREAT still outperforms other attacks even under the strongest defenses. *Second*, we evaluate the side effects that each PPA brings to the overall SR performance. We present the recommendation accuracy of all users on both backbones before and after PPA in Figure 3. From it, we observe all attacks damage the performance of SR. Compared to SOTAs, CREAT results in slighter accuracy degradation across all datasets, indicating that: 1) CREAT efficiently attacks by uncovering finer-grained sequential pattern correlations with minimal disturbance. 2) Since CREAT causes less degradation in the overall accuracy, it is less likely to trigger detection mechanisms that rely on fluctuations in SR performance, thereby making the attack more stealthy. *Third*, we employ t-SNE (Van der Maaten and Hinton 2008) to visualize the latent distribution of original sequences and polluted sequences. From Figure 4: 1) the adversarial samples from *Popular* and *Random* are distinguishable from original samples, while *SimAlter* with feature alignment improves stealth but still has detectable anomalies. 2) Compared to the best SOTA *Replace*,

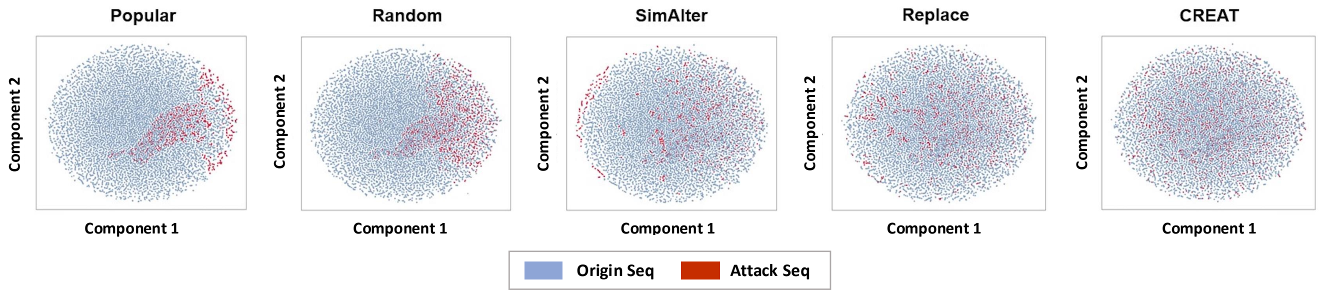


Figure 4: The t-SNE visualization of original sequences and polluted sequences.

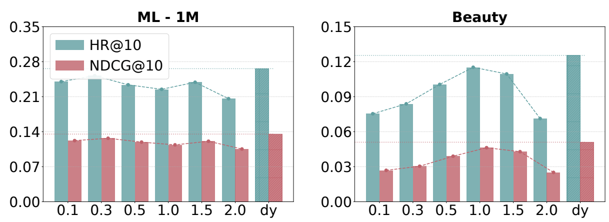


Figure 5: The effects of the dynamic barrier.

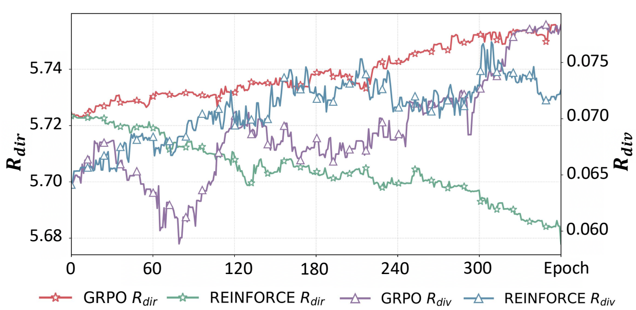


Figure 6: The convergence of CREAT.

Attack	NARM			Bert4Rec		
	Head	Medium	Tail	Head	Medium	Tail
ML-1M						
Pure	0.0513	0.0174	0.0000	0.2149	0.0128	0.0000
SimAlter	0.3079	0.0658	0.0000	0.3081	0.0563	0.0000
Replace	0.3535	0.1027	0.0036	0.4829	0.0938	0.0000
CREAT	**0.5664**	**0.1428**	**0.0451**	**0.5228**	**0.1492**	**0.0707**
Beauty						
Pure	0.0472	0.0033	0.0000	0.0254	0.0000	0.0000
SimAlter	0.1504	0.0178	0.0004	0.5277	0.0048	0.0000
Replace	0.3350	0.0608	0.0000	0.4947	0.0000	0.0000
CREAT	**0.4714**	**0.1504**	**0.0102**	**0.6587**	**0.2601**	**0.1999**
ML-20M						
Pure	0.0974	0.0215	0.0000	0.2139	0.0046	0.0000
SimAlter	0.4031	0.2048	0.0000	0.3401	0.1398	0.0000
Replace	0.3865	0.1906	0.0000	0.3816	0.0872	0.0000
CREAT	**0.4317**	**0.3041**	**0.0953**	**0.4590**	**0.2628**	**0.1502**

Table 3: Attack across various target popularity (HR@10).

CREAT shows more stealth, achieving indistinguishability between polluted and clean sequences, indicating high integration in spatial density and local clustering patterns.

Adaptability of Dynamic Barrier (RQ4). To justify necessity of dynamically adjusting constraint penalty, we replace the constrained reinforcement with *dynamic barrier* by fixed penalty coefficients. From Figure 5, we observe that setting the penalty to static values, *i.e.,* ranging from 0.1 to 2.0, leads to suboptimal or unstable performance. Contrastingly, CREAT (dy) achieves the best performance, indicating the ability of dynamic barrier to adaptively adjust the penalty in response to constraint violations and optimization dynamics.

GRPO Convergence (RQ5). We prove GRPO's convergence through first-360-epoch reward trajectories, compared with traditional REINFORCE. Figure 6 shows REINFORCE preserves consistency reward but suffers persistent pattern inversion reward decline due to its lack of group-wise baseline normalization for reward functions, whereas GRPO achieves sustained pattern inversion reward growth stabilizing post-epoch-300. Despite initial volatility, GRPO's consistency reward stabilizes post-epoch-120 and converges post-epoch-330, confirming its capability.

Effect of Target Item Popularity (RQ6). To investigate the effect of target item popularity on attacks, we conduct more experiments. From Table 3, we find: 1) *SimAlter* and *Replace* exhibit significantly weaker tail-item attack capabilities than CREAT, demonstrating near-zero efficacy across all datasets. 2) CREAT exhibits superior exposures across all range of popularity. Our attack demonstrates a greater relative improvement on lower-popularity targets, with tail items exhibit higher enhancement than medium and head targets. Since attackers in real scenarios typically focus on increasing the illicit exposure of long-tail items rather than already popular ones, this advantage has practical significance.

Conclusion and Future Work

We rethink the PPA against SR into a bi-level optimization problem, where we design a pattern balanced rewarding policy to reverse key patterns and implement a constrained group relative reinforcement learning. Experiments demonstrate superiority of CREAT. But there are currently no effective defenses against CREAT, which pose potential security risks to recommenders. In future, we will extend adaptive defenses against such attacks for recommender safety.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No.72192823 and No.624B2131) and the Fundamental Research Funds for the Central Universities. Xiaobo Xia is partially supported by the MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition at the University of Science and Technology of China (Grant No. 2421002).

References

Cuturi, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

Dang, Y.; Liu, Y.; Yang, E.; Huang, M.; Guo, G.; Zhao, J.; and Wang, X. 2025. Data augmentation as free lunch: Exploring the test-time augmentation for sequential recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1466–1475.

Du, L.; Yuan, Q.; Chen, M.; Sun, M.; Cheng, P.; Chen, J.; and Zhang, Z. 2024a. PARL: Poisoning Attacks Against Reinforcement Learning-based Recommender Systems. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*, 1331–1344.

Du, X.; Chen, Y.; Zhang, Y.; and Tang, J. 2024b. Precision Profile Pollution Attack on Sequential Recommenders via Influence Function. *arXiv preprint arXiv:2412.01127*.

Fan, W.; Derr, T.; Zhao, X.; Ma, Y.; Liu, H.; Wang, J.; Tang, J.; and Li, Q. 2021. Attacking black-box recommendations via copying cross-domain user profiles. In *2021 IEEE 37th international conference on data engineering (ICDE)*, 1583–1594. IEEE.

Flamary, R.; Courty, N.; Gramfort, A.; Alaya, M. Z.; Boisbunon, A.; Chambon, S.; Chapel, L.; Corenflos, A.; Fatras, K.; Fournier, N.; et al. 2021. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78): 1–8.

Gong, C.; and Liu, X. 2021. Bi-objective trade-off with dynamic barrier gradient descent. *NeurIPS 2021*.

Huang, H.; Mu, J.; Gong, N. Z.; Li, Q.; Liu, B.; and Xu, M. 2021. Data poisoning attacks to deep learning based recommender systems. *arXiv preprint arXiv:2101.02644*.

Kang, W.-C.; and McAuley, J. 2018. Self-attentive sequential recommendation. In *ICDM*, 197–206. IEEE.

Li, J.; Zhang, W.; Wang, T.; Xiong, G.; Lu, A.; and Medioni, G. 2023. GPT4Rec: A generative framework for personalized recommendation and user interests interpretation. *arXiv preprint arXiv:2304.03879*.

Lin, C.; Chen, S.; Zeng, M.; Zhang, S.; Gao, M.; and Li, H. 2022. Shilling black-box recommender systems by learning to generate fake user profiles. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1): 1305–1319.

Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; et al. 2024a. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Liu, F.; Zhang, H.; Lan, Y.; and Li, M. 2025. FewMEA: Few-shot Model Extraction Attack against Sequential Recommenders. In *Proceedings of the 2025 International Conference on Multimedia Retrieval*, ICMR '25, 917–925. New York, NY, USA: Association for Computing Machinery. ISBN 9798400718779.

Liu, Q.; Wu, X.; Wang, Y.; Zhang, Z.; Tian, F.; Zheng, Y.; and Zhao, X. 2024b. Llm-esr: Large language models enhancement for long-tailed sequential recommendation. In *NeurIPS*, volume 37, 26701–26727.

Liu, W.; Zheng, X.; Chen, C.; Su, J.; Liao, X.; Hu, M.; and Tan, Y. 2023a. Joint internal multi-interest exploration and external domain alignment for cross domain sequential recommendation. In *Proceedings of the ACM Web Conference 2023*, 383–394.

Liu, W.; Zheng, X.; Su, J.; Hu, M.; Tan, Y.; and Chen, C. 2022. Exploiting Variational Domain-Invariant User Embedding for Partially Overlapped Cross Domain Recommendation. In *The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, 312–321. ACM.

Liu, X.; Li, Z.; Gao, Y.; Yang, J.; Cao, T.; Wang, Z.; Yin, B.; and Song, Y. 2023b. Enhancing user intent capture in session-based recommendation with attribute patterns. *Advances in Neural Information Processing Systems*, 36: 30821–30839.

Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; and Vladu, A. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*.

Meng, W.; Xing, X.; Sheth, A.; Weinsberg, U.; and Lee, W. 2014. Your online interests: Pwned! a pollution attack against targeted advertising. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 129–140.

Nguyen, T. T.; Quoc Viet hung, N.; Nguyen, T. T.; Huynh, T. T.; Nguyen, T. T.; Weidlich, M.; and Yin, H. 2024. Manipulating Recommender Systems: A Survey of Poisoning Attacks and Countermeasures. *ACM Computing Surveys*, 57(1).

Pham, K.; Le, K.; Ho, N.; Pham, T.; and Bui, H. 2020. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, 7673–7682. PMLR.

Rendle, S.; Freudenthaler, C.; and Schmidt-Thieme, L. 2010. Factorizing personalized markov chains for next-basket recommendation. In *WWW*, 811–820.

Séjourné, T.; Vialard, F.-X.; and Peyré, G. 2022. Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, 4995–5021. PMLR.

Song, J.; Li, Z.; Hu, Z.; Wu, Y.; Li, Z.; Li, J.; and Gao, J. 2020. Poisonrec: an adaptive data poisoning framework for attacking black-box recommender systems. In *2020 IEEE 36th international conference on data engineering (ICDE)*, 157–168. IEEE.

Su, J.; Chen, C.; Lin, Z.; Li, X.; Liu, W.; and Zheng, X. 2023a. Personalized Behavior-Aware Transformer for

Multi-Behavior Sequential Recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, 6321–6331. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701085.

Su, J.; Chen, C.; Liu, W.; Wu, F.; Zheng, X.; and Lyu, H. 2023b. Enhancing hierarchy-aware graph networks with deep dual clustering for session-based recommendation. In *Proceedings of the ACM web conference 2023*, 165–176.

Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; and Jiang, P. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. arXiv:1904.06690.

Tang, J.; and Wang, K. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *WSDM*, 565–573.

Tang, J.; Wen, H.; and Wang, K. 2020. Revisiting adversarially learned injection attacks against recommender systems. In *Proceedings of the 14th ACM Conference on Recommender Systems*, 318–327.

Tran, Q. H.; Janati, H.; Courty, N.; Flamary, R.; Redko, I.; Demetci, P.; and Singh, R. 2023. Unbalanced co-optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10006–10016.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Wang, Y.; Su, J.; Chen, C.; Han, M.; Zhang, C.; and Wang, J. 2025. Sim4Rec: Data-Free Model Extraction Attack on Sequential Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(12): 12766–12774.

Wang, Z.; Gao, M.; Yu, J.; Ma, H.; Yin, H.; and Sadiq, S. 2024. Poisoning attacks against recommender systems: A survey. *arXiv preprint arXiv:2401.01527*.

Wang, Z.; Yu, J.; Gao, M.; Yin, H.; Cui, B.; and Sadiq, S. 2023. Poisoning Attacks Against Contrastive Recommender Systems. *arXiv preprint arXiv:2311.18244*.

Wu, C.; Lian, D.; Ge, Y.; Zhu, Z.; and Chen, E. 2023. Influence-driven data poisoning for robust recommender systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10): 11915–11931.

Wu, C.-Y.; Ahmed, A.; Beutel, A.; Smola, A. J.; and Jing, H. 2017. Recurrent recommender networks. In *WSDM*, 495–503.

Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; and Tan, T. 2019. Session-based recommendation with graph neural networks. In *AAAI*, volume 33, 346–353.

Xie, X.; Sun, F.; Liu, Z.; Wu, S.; Gao, J.; Zhang, J.; Ding, B.; and Cui, B. 2022. Contrastive learning for sequential recommendation. In *ICDE*, 1259–1273. IEEE.

Xing, X.; Meng, W.; Doozan, D.; Snoeren, A. C.; Feamster, N.; and Lee, W. 2013. Take this personally: Pollution attacks on personalized services. In *22nd USENIX Security Symposium (USENIX Security 13)*, 671–686.

Yang, G.; Gong, N. Z.; and Cai, Y. 2017. Fake co-visitation injection attacks to recommender systems. In *NDSS*.

Yang, Z.; Wu, J.; Wang, Z.; Wang, X.; Yuan, Y.; and He, X. 2023. Generate what you prefer: Reshaping sequential recommendation via guided diffusion. In *NeurIPS*, 24247–24261.

Yin, M.; Wang, H.; Guo, W.; Liu, Y.; Zhang, S.; Zhao, S.; Lian, D.; and Chen, E. 2024. Dataset regeneration for sequential recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3954–3965.

Yue, Z.; He, Z.; Zeng, H.; and McAuley, J. 2021. Black-box attacks on sequential recommenders via data-free model extraction. In *RecSys*, 44–54.

Yue, Z.; Zeng, H.; Kou, Z.; Shang, L.; and Wang, D. 2022. Defending substitution-based profile pollution attacks on sequential recommenders. In *RecSys*, 59–70.

Zhang, H.; Li, Y.; Ding, B.; and Gao, J. 2020. Practical data poisoning attack against next-item recommendation. In *WWW*, 2458–2464.

Zhang, H.; Tian, C.; Li, Y.; Su, L.; Yang, N.; Zhao, W. X.; and Gao, J. 2021a. Data poisoning attack against recommender system using incomplete and perturbed data. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2154–2164.

Zhang, K.; Cao, Q.; Wu, Y.; Sun, F.; Shen, H.; and Cheng, X. 2024. Lorec: Large language model for robust sequential recommendation against poisoning attacks. *arXiv preprint arXiv:2401.17723*.

Zhang, S.; Yin, H.; Chen, T.; Huang, Z.; Nguyen, Q. V. H.; and Cui, L. 2022. Pipattack: Poisoning federated recommender systems for manipulating item promotion. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 1415–1423.

Zhang, X.; Chen, J.; Zhang, R.; Wang, C.; and Liu, L. 2021b. Attacking recommender systems with plausible profile. *IEEE Transactions on Information Forensics and Security*, 16: 4788–4800.

Zhang, Y.; Xiao, J.; Hao, S.; Wang, H.; Zhu, S.; and Jajodia, S. 2019. Understanding the manipulation on recommender systems through web injection. *IEEE Transactions on Information Forensics and Security*, 15: 3807–3818.

Zheng, Y.; Liu, S.; Li, Z.; and Wu, S. 2020. Dgtn: Dual-channel graph transition network for session-based recommendation. In *ICDMW*, 236–242. IEEE.