SNS COLLEGE OF ENGINEERING
Kurumbapalayam (Po), Coimbatore – 641 107
An Autonomous Institution
Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

COURSE NAME : 19AD701 RECOMMENDER SYSTEMS

IV YEAR /VII SEMESTER

Unit 3- COLLABORATIVE FILTERING

Topic 5 : Attacks on Collaborative recommender systems

Collaborative recommender systems, like other types of recommendation systems, are susceptible to various attacks and vulnerabilities. These attacks can compromise the integrity, privacy, and effectiveness of the recommendation process. Here are some common attacks on collaborative recommender systems:

1. **Data Poisoning Attacks**:

- **Profile Injection**: Attackers can create fake user profiles or inject fraudulent user-item interactions into the system to influence recommendations. This can lead to personalized recommendations that are manipulated to serve the attacker's interests.
- **Shilling Attacks**: Shilling attacks involve creating a group of fake user accounts that rate or review specific items positively or negatively to manipulate recommendations for those items. These attacks can distort item rankings and reduce recommendation quality.

2. Data Inference Attacks:

- **Inference of Sensitive Information**: By analyzing the recommendations provided to a target user, an attacker may infer sensitive information about the user, such as their health condition, political beliefs, or personal interests.
- **Membership Inference**: Attackers may try to infer whether a specific item or user is part of the recommendation system's dataset, compromising user privacy.

3. **Collusion Attacks**:

- **Collusion Among Users**: A group of users may collude to artificially boost or lower the ratings or interactions for specific items, manipulating the recommendations for those items.

- **Collusion with Malicious Items**: Attackers may introduce malicious items into the system and manipulate interactions with those items to promote them or harm the reputation of competitors.

4. **Model Poisoning Attacks**:

- **Adversarial Inputs**: Attackers may manipulate the input data to the recommendation algorithm, introducing adversarial items or users designed to degrade the quality of recommendations.
- **Adversarial Machine Learning**: Attackers can try to exploit vulnerabilities in machine learning models used for recommendation to manipulate their behavior, causing biased or harmful recommendations.

5. **Profile Inversion Attacks**:

- **Profile Reverse Engineering**: Attackers may attempt to reverse-engineer user profiles by analyzing the recommendations provided to determine the users' preferences, interests, or demographics.

6. **Sybil Attacks**:

- **Sybil Accounts**: Attackers create multiple fake accounts (Sybil accounts) to manipulate recommendations, increase the visibility of their items, or inflate their own ratings or reviews.

7. **Data Crawling Attacks**:

- **Web Scraping**: Attackers may use web scraping techniques to collect user-generated content, such as ratings, reviews, or user profiles, to build alternative recommendation models or for other malicious purposes.

8. **Denial of Service (DoS) Attacks**:

- **Overloading the System**: Attackers can flood the recommendation system with excessive requests or interactions to degrade its performance or cause it to become unresponsive.

9. Fairness Attacks:

- **Bias and Discrimination**: Attackers may exploit biases in recommendation systems to amplify existing biases or create new ones, leading to unfair or discriminatory recommendations.

To defend against these attacks, collaborative recommender systems should implement various security and privacy measures, including data validation, user authentication, access control, and anomaly detection. Additionally, techniques like differential privacy and federated learning can help protect user privacy and mitigate some of the vulnerabilities associated with collaborative filtering systems. It's crucial for developers and system operators to stay vigilant and continuously monitor and update their systems to address emerging threats.