2021 26th International Computer Conference, Computer Society of Iran (CSICC) | 978-1-6654-1241-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/CSICC52343.2021.9420553

Injection Shilling Attack Tool for Recommender Systems

Fatemeh Rezaimehr
Computer Engineering Depatement
K.N.Toosi university of Tecnology
Tehran, Iran
f.rezaimehr@email.kntu.ac.ir

Chitra Dadkhah
Computer Engineering Depatement
K.N.Toosi university of Tecnology
Tehran, Iran
dadkhah@kntu.ac.ir

***Abstract*—Recommender systems help people in finding a particular item based on their preference from a wide range of products in online shopping rapidly. One of the most popular models of recommendation systems is the Collaborative Filtering Recommendation System (CFRS) that recommend the top-K items to active user based on peer grouping user ratings. The implementation of CFRS is easy and it can easily be attacked by fake users and affect the recommendation. Fake users create a fake profile to attack the RS and change the output of it. Different attack types with different features and attacking methods exist in which decrease the accuracy. It is important to detect fake users, remove their rating from rating matrix and recognize the items has been attacked. In the recent years, many algorithms have been proposed to detect the attackers but first, researchers have to inject the attack type into their dataset and then evaluate their proposed approach. The purpose of this article is to develop a tool to inject the different attack types to datasets. Proposed tool constructs a new dataset containing the fake users therefore researchers can use it for evaluating their proposed attack detection methods. Researchers could choose the attack type and the size of attack with a user interface of our proposed tool easily.**

Keywords—Recommender Systems, Attack Type, Collaborative Filter, Tool, Fake user, Shilling Attack

I. INTRODUCTION

Recommender systems use various data sources such as user-item rating matrix, user profile, and item attributes to extract correlations between users and items. Researchers categorize the recommender systems into five general categories including: Content-Based (CB), Collaborative Filtering (CF) knowledge-Based (KB), Demographic and Hybrid approaches [1]. CF systems analyze only historical interactions, while CB methods are based on item attributes, the KB recommendations are based on explicitly specified user requirements. In demographic approach, Recommender systems use demographic information of user such as gender and age. In Hybrid approach, the various aspects of different types of recommender systems are combined to achieve the best part of all of them.

CF method is the most successful techniques used in many applications such as movie, music, and book [2]. The similarity of users with the target user based on the rating matrix is calculated and the system recommend the top-K items to target user based on target user neighbors. In CFRS, the attacker introduces himself/herself as the most similar user to the target user by rating the items like real users to push/nuke an item. Attackers do not need to have much knowledge of the system for attacking and creating several profiles with a fake identity to influence the system output and change the Top-K recommendation. In this regard, many algorithms have been developed to detect and eliminate attackers. Since there is not any dataset that contains fake users for different attack type, so researchers have to inject the attacks to their datasets and then evaluate their proposed algorithms. The aim of this article is to design a tool to inject an attack to the dataset. Researchers could choose the attack type and size of attack with a user interface of our proposed tool easily.

The structure of the paper is as follows: section 1 describes the different types of attack. Section 2 explains the related works. In section 3 our proposed tool is described and finally section 4 outlines the conclusion.

II. ATTACK TYPES

Attacks are divided into three general categories: shilling attacks, relationship attacks and combined attacks [3]. In shilling attacks or rating attacks, the fake users increase/decrease the popularity of a particular item for push/nuke it. Relationship attacks affect the relationship of individuals and ultimately increases/decreases the popularity of a particular user. These attacks are more effective on social networks. The combined attacks, which are the most destructive attacks, combine both types of attacks. We have considered shilling attacks for CFRS in our proposed tool.

In shilling attacks, user enters biased profiles to influence system's behavior [2, 4-10]. Researchers have divided shilling attacks into push and nuke attacks. Intention of attackers in push attacks is to add a specific item to the Top-K recommendation. On the other hand, attackers in nuke attacks attempt to remove a particular item from recommendation list. Attackers create fake profiles and inject them into the system to influence the recommended list. Fig 1 shows the general form of the attack profile. There are many factors to categorize an attack as shown in Table 1.

978-1-6654-1241-4/21/$31.00 ©2021 IEEE

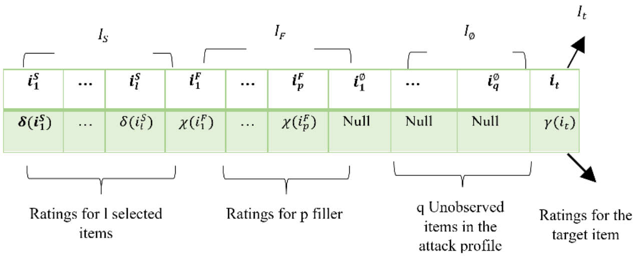


Fig. 1. The general form of an attack profile [2].

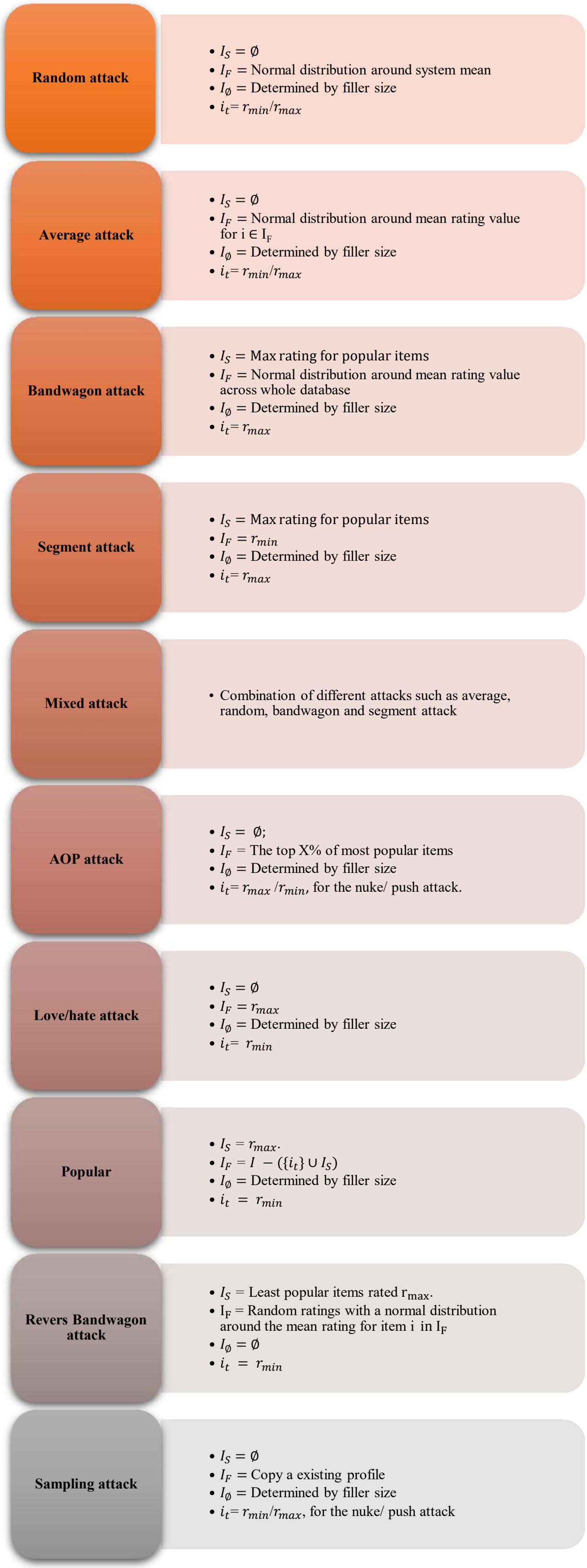
TABLE I. Description of attack feature.

	Feature	description
1	attack content	It indicates push or nuke items.
2	profile size	It is the number of ratings an attacker assigns to an attack profile.
3	Attack size	It is the number of profiles that an attacker injects to data set.
4	attack model	It describes items to be rated and attack strategy.

Researchers find the attackers behavior and their patterns so they can identify and remove them from the system. Rezeimehr and Dadkhah in [11] classified attacks into 10 categories as shown in Fig. 2.

III. Related Work

Srikanth et al. introduced two new measures, RDMB and CIDA, for identifying fake profiles and target items [12]. If the rate given by the suspected user for target item is higher/lower than the average ratings, the attack is push/nuke and the user is removed. Cai et al. selected the optimal list of neighbors for each active user using the VNS (Value-based Neighbor Selection) method. Their proposed system provided recommendations that could maximize expected profits and the level of attackers (number of attackers on the neighbors list) [13]. They used the RDMA measures to find the suspicious rating of the items and identify the attackers based on the difference between the user rate and the average rate of the items. To find other attackers, users were clustered hierarchically, and then GRDMA is calculated for each cluster. They believed that attackers have behavior similar to the other users in the system to effect recommendations. They also believed that the distance between the attackers and the real users is very short. Therefore, users in the cluster with the highest GRDMA are more likely to be attackers. Yang et al. proposed a Bayesian-based method for detection attacks [14]. They used matrix factorization to reduce the dimensions of the user-item matrix and extract user attributes using feature extraction. They updated these features with the Bayesian model and created a user-user graph that represents user relationships and tagged users based on the extracted features.



Random attack

- $I_S = \emptyset$
- $I_F =$ Normal distribution around system mean
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{min}/r_{max}$

Average attack

- $I_S = \emptyset$
- $I_F =$ Normal distribution around mean rating value for i ∈ I_F
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{min}/r_{max}$

Bandwagon attack

- $I_S =$ Max rating for popular items
- $I_F =$ Normal distribution around mean rating value across whole database
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{max}$

Segment attack

- $I_S =$ Max rating for popular items
- $I_F = r_{min}$
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{max}$

Mixed attack

- Combination of different attacks such as average, random, bandwagon and segment attack

AOP attack

- $I_S = \emptyset$;
- $I_F =$ The top X% of most popular items
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{max}/r_{min}$, for the nuke/ push attack.

Love/hate attack

- $I_S = \emptyset$
- $I_F = r_{max}$
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{min}$

Popular

- $I_S = r_{max}$.
- $I_F = I - (\{i_t\} \cup I_S)$
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{min}$

Revers Bandwagon attack

- $I_S =$ Least popular items rated r_{max}.
- $I_F =$ Random ratings with a normal distribution around the mean rating for item i in I_F
- $I_\emptyset = \emptyset$
- $i_t = r_{min}$

Sampling attack

- $I_S = \emptyset$
- $I_F =$ Copy a existing profile
- $I_\emptyset =$ Determined by filler size
- $i_t = r_{min}/r_{max}$, for the nuke/ push attack

Fig. 2. The Detail of different attacks [11].

Rezeimehr and Dadkhah surveyed the different attack detection method in CFRS from 2009-2019 [11].

IV. Our Proposed Tool

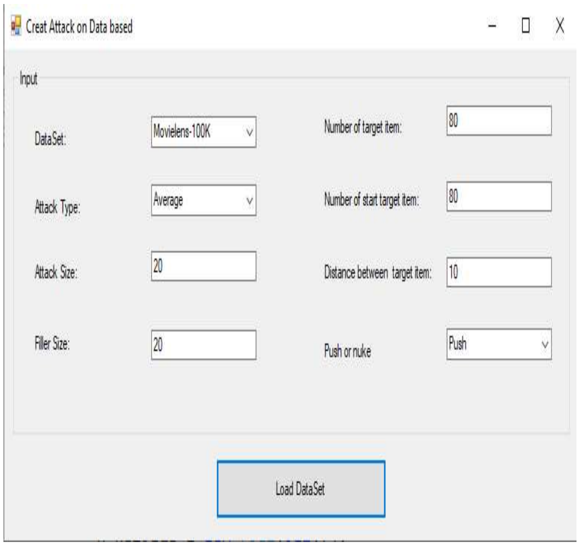
As mentioned, in the field of attack detection in collaborative filtering recommender systems, an integrated dataset containing the attackers is needed. Hence, we created a tool to add a different attack types on datasets. This tool has created with a C# programming language. All of the features and attack types could be inserted by the user with user interface of the tool. We have considered the same featured describe in table 1 in this part.

Fig 3 shows the user interface of the tool. The values of each box must be specified by the user. If the user does not enter the information, the default values will be considered.

First, the dataset can be selected by the researcher to inject the attack. The datasets considered in this tool are most commonly used in recommender systems such as Movielens (100K and 1M), Netflix, Amazon and Jester. Then attack type must be selected between Average, Random, Bandwagon and Segment strategy and the target item must be indicated as nuke or push. In the end, the researcher must identify the features of the attack described in table II.

TABLE II. Attack features in tool

Features		Description
	Attack size	The number of profiles that an attacker adds to the system
	Filler size	The number of items to be rated in each profile.
	Number of target item	The number of item to be attacked.
	Number started target item	Which item is the first item to be attacked?
	Distance between target item	What is the distance between the target items?
	Push or Nuke	Identify the type of attack for target item



DataSet: Movielens-100K

Attack Type: Average

Attack Size: 20

Filler Size: 20

Number of target item: 80

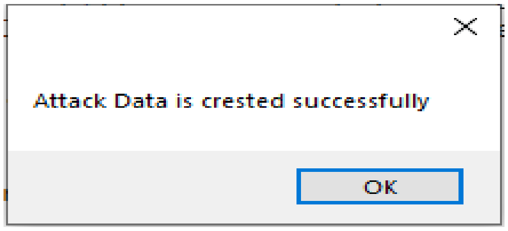
Number of start target item: 80

Distance between target item: 10

Push or nuke Push

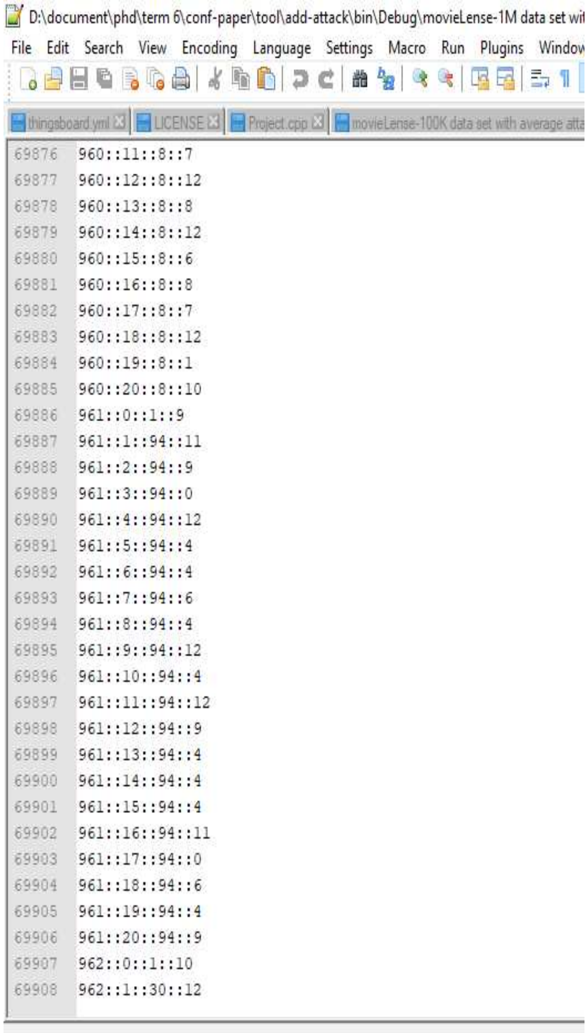
Fig. 3. User interface of the proposed tool.

After specifying all the required information, by clicking on the button *Load Data set*, injection of the attacks will start. Finally, after the message shown in Fig 4 is displayed, the new dataset containing the attacks will be generated and saved to the file. Fig 5 shows an executable instance of the MovieLens dataset based on the information in Fig. 3.



Attack Data is crested successfully

Fig. 4. End message.


D:\document\phd\term 6\conf-paper\tool\add-attack\bin\Debug\movieLense-1M data set wi
File Edit Search View Encoding Language Settings Macro Run Plugins Window
thingsboard.yml
LICENSE
Project.cpp
movieLense-100K data set with average atta

```
960::11::8::7
960::12::8::12
960::13::8::8
960::14::8::12
960::15::8::6
960::16::8::8
960::17::8::7
960::18::8::12
960::19::8::1
960::20::8::10
961::0::1::9
961::1::94::11
961::2::94::9
961::3::94::0
961::4::94::12
961::5::94::4
961::6::94::4
961::7::94::6
961::8::94::4
961::9::94::12
961::10::94::4
961::11::94::12
961::12::94::9
961::13::94::4
961::14::94::4
961::15::94::4
961::16::94::11
961::17::94::0
961::18::94::6
961::19::94::4
961::20::94::9
962::0::1::10
962::1::30::12
```

Fig. 5. New dataset contains attacks.

V. Conclusion

Recommendation systems are useful web applications which can also be used in industrial and commercial services. Nowadays, online shopping is very popular so the need for recommender systems with high precision can be felt more than before and they play an important role in keeping the users interested. The accuracy of the prediction and recommendation is the most important factor in RS. If the system does not have good precision, its popularity and the profit will decrease and the users will not have any more trust in the recommendations, so they will choose another online shopping sites. When a fake user attacks the system, the Top-K recommendation is affected. Therefore, the recommendation list will contain items which are not relevant to the user preference. So researchers have to design RS approaches resistant to different attack types and evaluate them with datasets containing fake users. In this article, we have designed a tool that injects different attack strategies and types containing their features to the desired database. So, the researchers could use this tool for creating the dataset containing different attack types and strategies for evaluating their system to detect the attackers.

REFERENCES

[1] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM,* vol. 35 1992, ,pp. 61-70.

[2] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *ACM SIGIR Forum*, 2017, pp. 227-234.

[3] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*, ed: Springer, 2011, pp. 1-35.

[4] R. Burke, B. Mobasher, and R. Bhaumik, "Limited knowledge shilling attacks in collaborative filtering systems," in *Proceedings of 3rd International Workshop on Intelligent Techniques for Web Personalization (ITWP 2005), 19th International Joint Conference on Artificial Intelligence (IJCAI 2005)*, 2005, pp. 17-24.

[5] R. Burke, B. Mobasher, R. Bhaumik, and C. Williams, "Collaborative recommendation vulnerability to focused bias injection attacks," in *International Conference on Data Mining: Workshop on Privacy and Security Aspects of Data Mining (ICDM 2005)*, 2005.

[6] R. Burke, B. Mobasher, R. Bhaumik, and C. Williams, "Segment-based injection attacks against collaborative filtering recommender systems," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 2005, p. 4 pp.

[7] R. Burke, B. Mobasher, R. Zabicki, and R. Bhaumik, "Identifying attack models for secure recommendation," *Beyond Personalization,* vol. 2005.

[8] S. K. Lam and J. Riedl, "Shilling recommender systems for fun and profit," in *Proceedings of the 13th international conference on World Wide Web*, 2004, pp. 393-402.

[9] M. O'Mahony, N. Hurley, N. Kushmerick, and G. Silvestre, "Collaborative recommendation: A robustness analysis," *ACM Transactions on Internet Technology (TOIT),* vol. 4, pp. 344-377, 2004.

[10] M. P. O'mahony, N. J. Hurley, and G. C. Silvestre, "An evaluation of neighbourhood formation on the performance of collaborative filtering," *Artificial Intelligence Review,* vol. 21, pp. 215-228, 2004.

[11] F. Rezaimehr and C. Dadkhah, "A survey of attack detection approaches in collaborative filtering recommender systems," *Artificial Intelligence Review,* 2020, pp. 1-56.

[12] T. Srikanth and M. Shashi, "New Metrics for Effective Detection of Shilling Attacks in Recommender Systems," *International Journal of Information Engineering & Electronic Business,* vol. 11, 2019.

[13] Y. Cai and D. Zhu, "Trustworthy and profit: A new value-based neighbor selection method in recommender systems under shilling attacks," *Decision Support Systems,* vol. 124, 2019, pp. 113112.

[14] F. Yang, M. Gao, J. Yu, Y. Song, and X. Wang, "Detection of shilling attack based on bayesian model and user embedding," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2018, pp. 639-646.