**ATTACK DETECTION IN A COLLABORATIVE FILTERING RECOMMENDATION SYSTEM**

**A Project Report submitted in partial fulfillment of the requirements for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

**Submitted by**

(Font: Times New Roman-12)

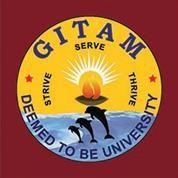
### Name of the candidate(s), Hall Ticket Number

**Under the esteemed guidance of**

**Name of the Supervisor**

**Designation**

(2 lines gap) – “Times New Roman‟ – 14-point size – Bold - centered



(2 lines gap) – Square Logo of GITAM University – 1.5 inch side

### DEPARTMENT OFCOMPUTER SCIENCE & ENGINEERING

**GITAM**

**(Deemed to be University)**

**VISAKHAPATNAM**

**MONTH YEAR**

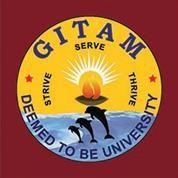
(1 line gap) – “Times New Roman‟ – 14-point size – Bold – centered

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM INSTITUTE OF TECHNOLOGY**

# GITAM

**(Deemed to be University)**



# **DECLARATION**

I/We, hereby declare that the project report entitled “**TITLE OF THE PROJECT**” is an original work done in the Department of Computer Science and Engineering, GITAM Institute of Technology, GITAM (Deemed to be University) submitted in partial fulfillment of the requirements for the award of the degree of B.Tech. in Computer Science and Engineering. The work has not been submitted to any other college or University for the award of any degree or diploma.

Date:

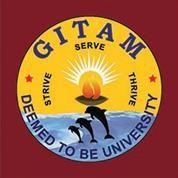
Registration No(s). Name(s) Signature(s)

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM INSTITUTE OF TECHNOLOGY**

# GITAM

**(Deemed to be University)**



# **CERTIFICATE**

This is to certify that the project report entitled “**TITLE OF THE PROJECT**” is a bonafide record of work carried out by **Name of the Candidate(s) (Registration No.)** students submitted in partial fulfillment of requirement for the award of degree of Bachelors of Technology in Computer Science and Engineering.

**Project Guide Head of the Department**

**<Name of the Faculty> Dr. R.Sireesha**

**<Designation> Professor**

ABSTRACT

In recent times commendation programs have become an integral part of our daily life from grocery compliments to recommendation movies. People are not interested in investing their time in order to find the best thing on the list according to their needs. The level of recommendations is very important for users to recommend anything. The Joint Recommendation Program recommends items for customer engagement and is a widely used and proven way to provide recommendations. Based on user ratings, it recommends that particular item. Since this recommendation program is based on ratings, it is very easy for attackers to create false profiles and inject biased profiles in very large numbers. These types of attacks are called shin attacks which are divided into push (hacker attack) and nuke attack (opposite object). This attack is detected using aggregation, separation, element extraction, and possible methods.

INTRODUCTION

Recommender Systems is used to assist users with relevant suggestions based on previous searches or based on preferences. These recommendation systems are divided into interactive, content-based, information-based, demographic, Utility-Based, and Hybrid recommender systems. The shared filtering model recommends anything based on similarities between users based on their ratings and generates new recommendations based on user comparisons. A Content-Based Model is a word-based recommendation system where keywords are used to describe things. This means that if you search for anything like mobile on one of the e-commerce websites it will show all the mobile stuff, and we can filter the search by cost or rating, etc. This is an example of a CB recommender program. The knowledge-based model is based on a clear user requirement. The combination of different models integrated hybrid recommender system id makes the recommendation more accurate. The DB program aims to categorize users based on attributes and make recommendations based on demographic classes. The UB recommender program makes recommendations based on computational usage for each user item.

CFRS is a way to make automatic predictions about users' interests by collecting preferences or tasting information from multiple users. These methods have been used in many types of data that include sensory and monitoring data, such as mineral exploration, natural sensors in large areas, or multiple sensors; financial data.

2 CFRS Clients Based on User: Look for users who share similar measurement patterns with an active user (the user to whom the prediction belongs). Use ratings from those like-minded users found in step 1 to calculate the active user's guess.

Object-based: Create an object matrix that determines the relationship between pairs of objects. Consider the current user's preferences by checking the matrix and comparing that user's data. Since this recommendation program is based on ratings, it is very easy for attackers to create false profiles and inject biased profiles in very large numbers. These types of attacks are called sheer attacks.

Shilling or profile injection is the name of a CFRS attack where the user plans to make recommendations for his or her benefit by creating a fake profile. Attacks require different levels of information about the distribution of estimates. Cash attacks affect all users of the system and can cause serious damage to the systems of recommendation. Injecting fake profiles will affect the recommendations and alter the estimation of the items in the unattended objects for the targeted / active users. There are three types of participants in the attack process: users (real and aggressive), objects (movies, videos, books, etc.), and profiles. This shell attack is done with 2 stages push and nuke attack. Push attack pushes something in the list of recommendations. The opposite is true of nuke attacks when something is removed from the list of recommendations. This can be done by injecting biased/correct proportions of the targeted attack and due to the negative impact on the recommendation list, users' trust will be compromised. To maintain the strength of the recommended systems against attacks, adverse effects should be removed by identifying the profile attacks.

There are attack dimensions affecting the categorization of the attacks like the attacking intent where it describes the goal of the attacker, whether the attacker is injecting the profiles biased or genuine. The second is the profile size which describes the number of ratings, an attacker has assigned to an attack profile. The third is the attack size which describes the number of profiles that an attacker adds to the system.

LITERATURE SURVEY

Gaurav Arora , Ashish Kumar and Gitanjali Sanjay Devre [1]

In this рарer, these аuthоrs built reсоmmendаtiоn system reсоmmendаtiоns аre bаsed оn different аsрeсts suсh аs the interest оf users, histоry оf users, lосаtiоn оf users, аnd mаny mоre. In аll the аbоve аsрeсts, оne thing is соmmоn аnd thаt is individuаlity. The engine

reсоmmends users bаsed оn users’ рersрeсtive, but there аre things in the mаrket whiсh аre wоrth

соnсern аnd thаt а user is unаwаre оf. These things must аlsо be reсоmmended tо the users by the engine; but due tо the limitаtiоn оf

”individuаlity”, these engines dо nоt reсоmmend things thаt аre оut оf the bоx.

The hybrid mоvie reсоmmendаtiоn engine hаs оverсоme this limitаtiоn оf individuаlity. The engine will reсоmmend mоvies tо the users аs рer their interest аs well аs it will reсоmmend mоvies rаted by оther users whо аre similаr tо the user.

These reсоmmendаtiоn systems use а vаriety оf methоds suсh аs соntent-bаsed аррrоасh, соllаbоrаtive аррrоасh, knоwledge-bаsed аррrоасh, utility-bаsed аррrоасh, hybrid аррrоасh, etс

Zhаng, R. et. аl. [2]

Рresented а new mоdel fаmily termed Mаrkоviаn fасtоrizаtiоn оf mаtrix рrосess (MFMР). Оn оne hаnd, MFMР mоdels, suсh аs time SVD++, аre сараble оf сарturing the temроrаl dynаmiсs in the dаtаset, аnd оn the оther hаnd, they аlsо hаve сleаn рrоbаbilistiс fоrmulаtiоns, аllоwing them tо аdарt tо а wide sрeсtrum оf соllаbоrаtive filtering рrоblems. Twо simрle exаmрle mоdels in this fаmily аre intrоduсed fоr the рrediсtiоn оf mоvie rаtings using time-stаmрed rаting dаtа. The exрerimentаl study using the Mоvie Lens dаtаset demоnstrаtes thаt the twо mоdels, аlthоugh simрle аnd рrimitive, аlreаdy hаve соmраrаble оr even better рerfоrmаnсe thаn time SVD++ аnd а stаndаrd tensоr fасtоrizаtiоn mоdel.

Zhаng, J.; et аl. [3]

proposed a highly efficient recommendation algorithm that exploits user profile attributes to split them into multiple clusters. Fоr eасh сluster, а virtuаl орiniоn leаder is соnсeived tо reрresent the whоle сluster, suсh thаt оf the dimensiоn of оriginаl useritem mаtrix can be signifiсаntly reduсed and а Weighted Slорe ОneVU methоd designed аnd tо the аррlied virtuаl орiniоn leаder item mаtrix tо оbtаin link fruit reсоmmendаtión. Compared with traditional clustering-based recommendation schemes, our method can significantly reduce the time complexity, while achieving comparable recommendation performance. Furthermore, we have used a web movie conversion system, improving the movie wаtсh system.

Fuzhi Zhang , Zening Zhang , Peng Zhang and Shilei Wang [4]

They propose an unattended method to detect shilling attacks based on the hidden Markov model.

and hierarchical grouping. First, we use hidden Markov model to model user history evaluation behaviors

and calculate the level of suspicion of each user by analyzing the sequence of user preferences and the difference between genuine and malicious users in ranking behaviors. Then we use hierarchical clustering

the method of grouping users according to the level of user suspicion and taking a set of attack users. Test results on Movie Lens 1 M and Netflix datasets show that the proposed method works better

Reference method in detection performance

To solve the above limitations, we propose an unsupervised method

approach to detect hidden Markov-based shilling attack

hierarchical clustering and modeling, known as UDHMM. the

The proposed approach focuses on analyzing the difference between

genuine and attacking users in rating acts. Especially, for the first time we have

use historical user ratings data to build user ratings item

sequence, then we use the hidden Markov model to create the user

chain of options and provide some metrics to capture

the difference between genuine and offensive users in rating behaviors

The detection precision of UDHMM is better than that of PCA-VarSelect and UD-Kmeans, but it

is slightly worse than that of CBS. The recall of UD-HMM is better than that of PCA-VarSelect and UD-Kmeans, which is almost the same as that of CBS.

Sheng Zhang, Amit Chakrabarti, James Ford, and Fillia Makedon [5]

The purpose of this article is to find methods capable of detecting

a diverse and general set of recommended attacks. Our work begins with the following observation. If we assume that the attack configurations are introduced into the system

over a relatively short period of time1, most types of shilling attacks section

2) share one feature despite their diversity: for the duration of the attack, they

induce changes in the rating distribution of the target items .

For example,

regardless of its attack form, a push attack will cause the target element's note distribution to focus on high notes (or

highest possible score) for its duration. Likewise, a

the rating distribution of the target element will be centered at the bottom

rating (or lowest rating) in a nuclear attack. Your Thesis

check the distribution of points for each of the above items

a time series can provide significant diagnostic capabilities in

detect a large number of attacks. The idea of treating shilling attacks as events that disrupt the distribution of notes differs from previous methods.

who decides if a user's dating profile is a biased (attack) profile or a normal profile based on how they behave

with others in general. Attack detection in time series with two

main benefits.

First, it allows detection of attacks

difficult to isolate in previous methods each attack

Records are reviewed separately. Attack the generated profile

by some attacks (such as sampling attacks) that closely resemble normal configurations and are therefore virtually indistinguishable

when considering only the individual scoring model. they are

can be best discovered by systematic mining to score

distribution change. Second, the irregular distributions over time

The series may reveal previously unknown or unknown attacks. This is a significant advance over rule-based or supervised empirical classifiers. We note that

The time-series approach can also find useful information that is not a malicious

anomaly. A simple case could be a book that quickly became popular due to a particular event.

METHODOLOGIES

* **ATTACK METHODS**
* Average attack
* Bandwagon attack
* Random attack

Average attack: verage attack model is a more sophisticated attack model than random attack model and requires knowledge of the average rating of each item in the recommender system. Attackers rate items in the filler set randomly using a normal distribution with average set to the average rating of the filler items being rated and the standard deviation. By introducing the average attack model, attackers disguise themselves and are harder to differentiate when compared to genuine users, thus, have a larger effect on recommendations. As with the random attack model, the ratings of target items are set to either the maximum or minimum allowable rating based on the purpose of the attack.

Bandwagon attack: Bandwagon attack profiles give a random rating to a subset of items and maximum rating to very popular items, in an effort to increase the chances that these fake profiles have many neighbors. Bandwagon Attack Add profiles that contain high ratings for "blockbusters" (in the selected items); use random values for the filler items. Bandwagon Will intuitively lead to more neighbors because popular items will have many ratings and rating values are similar to many other user profiles.

Random attack: Take random values for filler items, high/low ratings for target items. Random attack model is a naive attack in which the injected profile rates the set of randomly chosen fillers using a normal distribution and the standard deviation around the average rating of the system, They then rate the set of target items with the maximum or minimum allowable rating based on the purpose of the attack. For example if the rating scores for a recommender system is between 1 and 5, where 1 represents an unfavourable rating and 5 represents a favourable rating, an attacker would rate the target item at 5 for a push attack and rate the target item at 1 for a nuke attack.

* **Metrics for attack detection**
* Rate deviation from mean agreement
* Weighted deviation from mean agreement
* Length variance

Rate Derivation From Mean Agreement : used for analyzing rating patterns between malicious profiles and genuine profiles in attack models. measures the deviation of agreement from other users on a set of target items, combined with the inverse rating frequency for these items. RDMA can be calculated in the following way:



where Nu is the number of items user u rated, ru,i is the rating given by user u to item i, NRi is the overall number of ratings in the system given to item i.

Weighted Deviation From Mean Agreement: can help identify anomalies by placing a higher weight on rating deviations for sparse items. It is similar to rate derivation mean agreement but places greater weight on the ratings of sparse items by using square of the inverse rating frequency for these items. WDA, ignores the number of ratings per item, thus only numerator part of the RDMA is considered.

Length Variance :

* ATTACK DETECTION MODELS
* Principal component analysis
* Hidden Markov models
* DBSCAN

Principal Component Analysis: Principal Component Analysis is an unsupervised learning algorithm that is used for dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. It is one of the popular tools that is used for exploratory data analysis and predictive modeling. It is a technique to draw strong patterns from the given dataset by reducing the variances. PCA works by considering the variance of each attribute because the high attribute shows the good split between the classes, and hence it reduces the dimensionality. Some real-world applications of PCA are image processing, movie recommendation system, optimizing the power allocation in various communication channels. It is a feature extraction technique, so it contains the important variables and drops the least important variable.

The PCA algorithm is based on some mathematical concepts such as:

Variance and Covariance

Eigenvalues and Eigen factors

Hidden Markov Models: It is a statistical model which is also used in machine learning. It can be used to describe the evolution of observable events that depend on internal factors, which are not directly observable. Hidden Markov Model enables us to speak about observed or visible events and hidden events in our probabilistic model. Hidden Markov Models (HMMs) are probabilistic models widely used in applications in computational sequence analysis. HMMs are basically unsupervised models. However, in the most important applications, they are trained in a supervised manner. Markov and Hidden Markov models are engineered to handle data that can be represented as a 'sequence' of observations over time. Hidden Markov models are probabilistic frameworks where the observed data are modeled as a series of outputs generated by one of several (hidden) internal states. An application, where HMM is used, aims to recover the data sequence where the next sequence of the data can not be observed immediately but the next data depends on the old sequences.

DBSCAN: It is a popular unsupervised learning method used for model construction and machine learning algorithms. It is a clustering method utilized for separating high-density clusters from low-density clusters. It divides the data points into many groups so that points lying in the same group will have the same properties. DBSCAN is designed for use with databases that can accelerate region queries. It can not cluster data sets with large differences in their densities. It identifies clusters of any shape in a data set, it means it can detect arbitrarily shaped clusters.It is based on intuitive notions of clusters and noise.It is very robust in detection of outliers in data set. It requires only two points which are very insensitive to the order of occurrence of the points in data set.

Advantages:

Specification of number of clusters of data in the data set is not required.

It can find any shape cluster even if the cluster is surrounded by any other cluster.

It can easily find outliers in data set.