In Loco Identity Fraud Detection Model Using Statistical Analysis for Social Networking Sites: A Case Study with Facebook

Shalini Hanok
Electronics and Communication Engineering,
ATME College of Engineering, Karnataka
shalini.prabhakar@gmail.com

Shankaraiah
Sri Jayachamarajendra College of Engineering (SJCE),
JSS Science and Technology University, Karnataka
shankaraiah@sjce.ac.in

Abstract: *Rapid advancement in internet has made many Social Networking Sites (SNS) popular among a huge population, as various SNS accounts are interlinked with each other, spread of stored susceptible information of an individual is increasing. That has led to various security and privacy issues; one of them is impersonation or identity fraud. Identity fraud is the outcome of illegitimate or secret use of account owner's identity to invade his/her account to track personal information. There are possibilities that known persons like parents, spouse, close friends, siblings who are interested in knowing what is going on in the account owner's online life may check their personal SNS accounts. Hence an individual's private SNS accounts can be invaded by an illegitimate user secretly without the knowledge of the account owner's which results in compromise of private information. Thus, this paper proposes an in loco identity fraud detection strategy that employs a statistical analysis approach to constantly authenticate the authorized user, which outperforms the previously known technique. This strategy may be used to prevent stalkers from penetrating a person's SNS account in real time. The accuracy attained in this research is greater than 90% after 1 minute and greater than 95% after 5 minutes of observation*.

Keywords: *Continuous authentication, in loco, SNS, identity fraud, stalkers.*

Received April 27, 2021; accepted September 22, 2021
https://doi.org/10.34028/iajit/20/2/15

1. Introduction

Today's twenty first century world is completely revolving around the internet as it is playing a vital role in each one's life. Social Networking Sites (SNS) are used by many users' for sharing their personal and delicate information such as photos, videos, friend lists, chats through their SNS accounts such as Facebook, WhatsApp, LinkedIn and other well-known SNS which are used for communication.

People now use social media and microblogs like Twitter and Facebook as a platform and first choice when they want to share information, write about their daily lives, or look for breaking news [1]. Although there are many advantages of using SNS on Online Social Networks (OSNs) few privacy and security risks are also involved in it that have to be solved. One of the major security risks the SNS user's face these days is identity theft or identity fraud or impersonation.

The attackers or stalkers tend to get personal information from SNS accounts of that violates the privacy and security of SNS hence there is a need for trustworthy techniques to authenticate authorized users. There are a number of methods developed to protect the user's identity. For example, Facebook stores the traditional IP address and devices used by each account to login to the site and verifies if the same account is logged in on other IP addresses or device by asking few secret questions [7] or security codes are sent to respective mail or a One Time Password (OTP) is sent to the registered mobile number to authenticate the user. Facebook also sends some suspect messages to the respective user accounts if some malicious activity is detected [6].

Even though there are a number of methods developed to avoid identity fraud, the methods that are present to detect identity fraud using the same device, same network and same IP address i.e., in loco identity fraud are very few. In loco identity fraud can be performed by any of the attackers whom we know, for example parents may use their children's accounts to know their activities' or partners at home or colleagues at the work place can also access user's device. Similarly, friends, relatives, siblings and other persons can access user's account when it is left open on the device such as laptops, mobile phones etc.

In loco identity fraud detection is required where the user tend to save their SNS passwords either on laptop or in their mobile devices for auto login in future, for the purpose of their convenience [17, 27]. Mobile devices are more exposed to strangers or stalkers for identity fraud [35] as SNS such as Facebook, Twitter, Gmail once logged in do not need further authentication for 60 days [10]. Thus, it is clear that if a person knows the user

closely or if the user's mobile device is reachable to someone then getting access to the user's SNS profile by impersonation gets easy which would lead to compromise of their privacy.

In fact, detecting identity fraud on SNS accounts using the same IP address and device is impossible since stalkers utilize the same IP address/device to breach the account owner's privacy without leaving any trace. Since the SNS account is already signed in, account owners are unable to identify and report identity fraud unless definite evidence of tampering is discovered. Wu *et al*. [34] discuss one approach, which is a strategy created to identify in loco identity theft.

In the study provided by Wu *et al*. [34], in situ (in loco) identity theft was discovered using Facebook as a case study among the available SNS sites, with user browsing activity evaluated by examining clicks on newsfeeds, Friend lists, profiles, likes, messages, photos/videos, and comments. To extract, identify, and authenticate the authorized user, a few machine learning methods such as Support Vector Machine (SVM), Smooth SVM (SSVM), and Karush-Kuhn-Tucker (KKT) are utilized on SNS servers. Within 2 minutes, the findings were more than 80% accurate, and after 7 minutes, they were more than 90% accurate.

In this article, we look at the topic of detecting in-person identity theft on social media sites, using Facebook as a case study. The detection method utilized in this work differs significantly from that used by Wu *et al*. [34]. The detection technique is based on statistical analysis, in which account owners' and stalkers' normal browsing behavior is separately monitored and recorded from SNS servers, and then the detection scheme is run, wherein the newly obtained behavior is compared to account owners' and stalkers' behavior.

If the system detects suspicious activity, it prompts the user to enter the account's password; if the password is right, the user is permitted to browse the account; if the password is wrong, surfing is terminated. This protects the user's account by allowing authorized users to access it via an in loco environment. The approach was tested, and the findings produced using statistical methods were found to be superior to those obtained in the publication [34], namely, accuracy of more than 90% after 1 minute and more than 95% after 5 minutes of observation.

Since in loco identity fraud has received little attention in recent years, more study is needed in this area to secure users' sensitive information as the number of users grows fast.

The structure of this document is as follows: Section 2 describes the major contribution of the paper section 3 summarizes the evaluations of relevant earlier works; section 4 examines various behavioral aspects of Facebook users; section 5 discusses the proposed work; Section 6 describes the data collection; section 7 describes the performance evaluation of the implemented detection scheme; section 8 compares the different methods with the proposed method and finally, section 9 concludes the proposed work.

2. Major Contribution of the Paper

- The study uses a statistical model to extract features from Facebook users and store them as training data in a database.
- The test features are compared using statistical analysis and the F-test.
- F-test is used to determine the critical value; if a match is discovered, the F-test value equals the critical value; otherwise, the system re-authenticates the user using their password.
- The features derived in this research are solely based on observations of individuals' activity within their accounts, as well as observations of stalkers' activity in other accounts.

3. Literature Review

This section reviews the various security measures implemented in SNS and the need for improvement. Considering Facebook is the most extensively and frequently utilized SNS on OSN as well as Mobile Social Network (MSN), security is critical. Even though Facebook has built-in security features such as two-factor authentication and login alerts [13], there is always a need for further protection.

As mentioned before, a few algorithms were created in the previous year to monitor fraud detection. Fraud detection is the most common use for analyzing user clickstream data. The natural and most comparable actions of distinct users were examined utilizing these clickstream data [15, 19, 26], using various approaches such as Markovchains [2, 16] and clustering [32, 33]. Clickstream data analysis has aided in the examination of various user acceptances [22] and offers this analysis for future research [36].

The first bot identification technique on Twitter, developed by Varol *et al*. [29], cross-validated publicly available data with an accuracy rate of 86%.

With an accuracy of 72%, Nuakoh and Anwar [21] employed the Artificial Immune System (AIS) approach to distinguish fraudulent or fictitious users from legitimate users.

Concepcin-Snchez *et al.* [5] employed a text mining and fuzzy logic technique to identify identity theft by examining the total words and emotions that appear often but are not used by the user, as well as the location from which the user signs in to his account. In the application level, there are a variety of behavioral biometrics available; one of the ways that has prompted researchers to consider using them is keystroke biometric [23].

The study work of Morales *et al*. [20] is the first to apply a keystroke biometric approach to identify authorized users. This technique authenticates an

individual based on the way and rhythm with which they type characters. The system's performance was assessed, and an Equal Error Rate (EER) of 5.32% was attained, with accuracy deterioration of less than 1%.

In their study, Daribay *et al.* [8] employed keystroke dynamics to create a multimodal authentication system with the use of machine learning/deep learning classification methods, achieving a 90. 91% accuracy.

By detecting rapid changes in user behavior, Egele *et al*. [9] presented a unique way to detect stalkers utilizing the account owner's profile using statistical modeling and anomaly detection approaches. COMPA is a technique they created to recognize valid user activity, and it was tested on big datasets of Twitter and Facebook conversations.

For online sub-communities, or communities that exist within bigger communities (like Facebook groups or Sub-reddits), Tsikerdekis [28] suggests a proactive strategy that makes use of social network data and is concentrated on identity deception prevention. The technique delivers excellent accuracy when spotting fraudulent accounts when an effort is made to join a sub community and may be used with a variety of Social Media Platforms (SMPs). Social Networking Analysis (SNA) was employed for detection, and the results showed a 73% accuracy rate.

A method that has been effectively used to identify false accounts made by bots or computers was proposed by Walt and Eloff [31]. These machine learning models have to use constructed variables, including the "friend-to-follower's ratio," in the case of bots. These elements were created using attributes from SMP account profiles, such "friend-count" and "follower-count," which are readily available. The study covered in this paper uses these similar engineering qualities to a collection of fictitious human accounts in an effort to further the accurate identification of fictitious human identities on SMPs. The technique resulted in accuracy of 85.91%.

Chaudhary *et al*. [4] used deep learning to analyze the topological properties of social networks in their study to use deep learning to identify anomalies in email networks and twitter networks. They demonstrated a model called the Graph Neural Network, which is used to analyze social link networks for anomalies. Their approach yields a 97.56% accuracy rate.

To evaluate the data, Hu *et al.* [14] employed logistic regression as a machine learning approach. The overall percentage of correct classification using the National Crime Victimization Survey Identity Theft Supplement (NCVS-ITS) conducted by the Bureau of Justice Statistics (BJS) in 2012, 2014, and 2016 is found to be 91.3 %.

Data mining approaches are highlighted by Raja and Raja [25] in order to preserve original user information and spot malicious accounts on social networking sites. According to the evaluation of Publicly Privacy Protection System (3PS), this work uses the malicious account detection technique in OSN dependent on the deceitful person's many shared posts in a day and newest activity and behaviors. This activity includes setting up OSN accounts for experiments, checking into the most current postings, comments, and images, running web searches, etc., Using this method, it is discovered that the accuracy is 94.6 %.

The User Credibility (UCred) model is suggested by Verma *et al.* [30] as a way to distinguish between false and legitimate user accounts. This model combines the outputs of Robustly optimized BERT, Bi-LSTM (RoBERT) (Bidirectional LSTM), and Random Forest (RF) to classify profiles. The output produced by all three methods is fed into the voting classifier to increase classification precision. 98.96% accuracy is provided by the suggested UCred model.

Wu *et al.* [34] suggested a continuous authentication approach to identify in situ (in loco) identity fraud episodes involving the same accounts, device, and IP address. The social network utilized for the case study is Facebook. SVM, a widely used machine learning algorithm for binary classifications, is employed in this detection strategy, followed by SSVM to speed up the training process. It also employs the KKT optimization approach to transform SVM to an unconstrained minimization issue. The accuracy of the results was 96.2%, with a 90% True Positive Rate (TPR) and a 4.5% False Positive Rate (FPR).

Wu *et al*. [34] conducted research that is quite similar to the work done in this publication, except that their detection approach is different from the detection system used in this paper, which yields a better result.

The detection techniques used in the literature survey so far have all predicted the behavior of the impostor if he/she is accessing a user's account on a different device, network, and IP address, but the detection technique used by Wu *et al.* [34] addresses a scenario where the account owner's and stalkers are accessing the SNS account on the same device, same IP address, and same network.

With an accuracy of 99.7%, TPR and FPR of 98% and 4.8%, respectively, the identification technique used in this study was statistically tested to identify legitimate users from stalkers. The results obtained in this research are superior to those obtained by Wu *et al.* [34].

4. Behavior Analysis on Facebook

The scientific study of the principles of learning and behavior is known as behavior analysis. The goal of this branch of study is to describe, analyze, predict, and change behavior. Individuals of any species, including humans, engage with one another for good or evil through social behavior. The goal of behavior analysis is to comprehend human social behavior. The Association for Behavior Analysis International (ABAI) defines behavior analysis as "the use of natural science

to understand individual expression." The proposed work investigates the social behavior of individuals on Facebook.

4.1. Data Collection from Facebook

To investigate various elements of Facebook users' behavior, some of them were asked to take part in an experiment that involved their visiting their own Facebook accounts as account owners and stalking the accounts of others.

The subjects were invited to perform the roles of account owner and stalker for three rounds of 30 minutes each, as illustrated in Figures 1 and 2, and the behavioral traits gained were recorded. For varied time limits, similar tests were carried out.

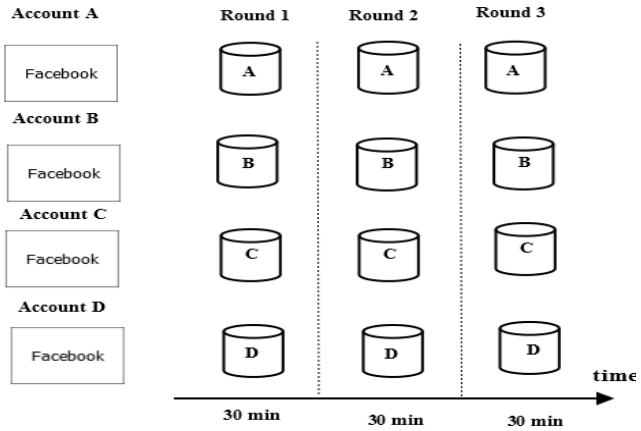


Figure 1. A, B, C, D are 4 different Facebook accounts. The user A browses his own account in three rounds for 30 minutes, similarly actions are done by user B, user C and user D and the datas are recorded.

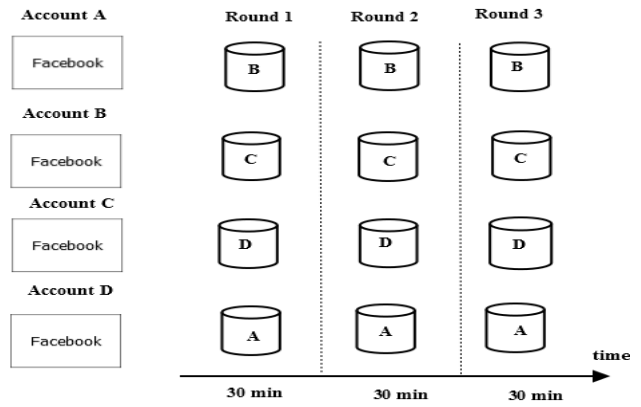


Figure 2. A, B, C, D are 4 different Facebook accounts. User A stalks the account of user B, user B stalks the account of user C, user C stalks the account of user D and user D stalks the account of user A for 30 minutes in 3 rounds and the datas are recorded.

For the purposes of this study, 50 frequent Facebook users who use their personal Facebook account on a daily basis were evaluated. The subjects were free to do anything they wanted in each round, including like or sharing photographs, videos, writing on friends' timelines, commenting on posts, changing their profile information, and so on, but external content surfing was limited to one minute.

4.2. Defining Facebook Features

The experiment data was examined, and a few similar aspects were tallied for all users, as shown in Table 1. In the table, there are account-related, page-related, or both activities recorded from the trials, and the user discovered 11 common traits.

Table 2 shows an example of a timestamp acquired for each user's action. Each action taken by the user includes a timestamp, the type of action taken, and a target individual, which can be either a friend or a stranger.

Table 1. Common user action collected from experiments on Facebook.

Actions	Account Related	Page Related
Likes	✓	
Comments	✓	✓
Posts	✓	✓
Friends with	✓	
Wrote on Timeline	✓	✓
Videos	✓	✓
React	✓	
Share	✓	✓
Tagged	✓	
Replied	✓	
Update Profile	✓	

Table 2. Examples of user action collected from Facebook.

Time Stamp	Actions	Target Person
1566289806	Group Page	
1566211499	Shares	Friend A
1566032456	Likes	Friend B
1566211498	Shares Page Post	
1566233996	Friend with	
1566235076	Reacted	Friend C

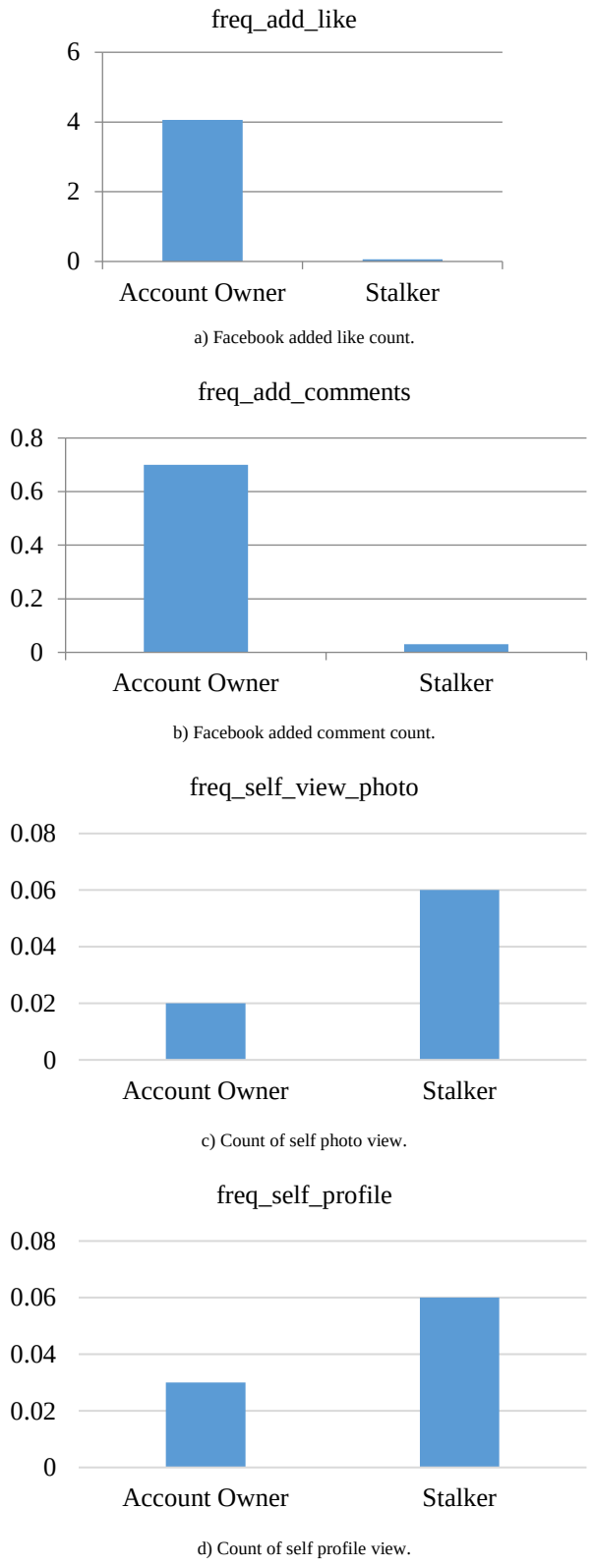
4.3. Various Behavioral Aspects of Users on SNS

The survey was done on a group of active Facebook users, and the results were analyzed using data from Figures 1, 2, and a few user patterns from Figure 3, which are briefly addressed in this section. The behavioral patterns of users on social media sites are studied to see if there is a distinction between stalkers and account owners. The graphs collected from the survey are shown in Figure 3-a). The survey details can be used for statistical modeling. The following are discussions on various users' behavior trends.

4.3.1. Sessions Controlled by Account Owners

Rather than their own personal facts, account owners are more interested in learning about their friends. The common account owner's actions in their Facebook account include adding comments, adding likes, adding reactions, adding reply, updating their own profile, adding comments on timeline, sharing some videos/photos, adding friends, adding reply, and adding posts in their profile some of them are depicted in Figure

3-a) and Figure 3-b). As a result, account owners have the highest values for actions performed in their account, while stalkers have the lowest values.



a) Facebook added like count.

b) Facebook added comment count.

c) Count of self photo view.

d) Count of self profile view.

Figure 3. Documentation of behavioral patterns of users on facebook.

4.3.2. Sessions Controlled by Stalkers

Stalkers are more interested in seeing self (account owner's) page wall, viewing self-profile, viewing some self-photos, growing and observing friends list or visiting friends page and viewing non friends page than performing some activity in their account some of them are depicted Figure 3-c) and Figure 3-d). As a result, stalkers place a higher importance on viewing the content of the account owner's account than on actions conducted in the account.

5. Proposed Method

The Figure 4 depicts the identity fraud detection technique discussed in this section. When a user first accesses the SNS server, the user's features are extracted and submitted to a detection model, which assesses if the user is authentic or not. If suspicious or fraudulent conduct is detected, a few authentication settings are triggered, requiring the user to provide a password to re-authenticate the legitimate user.

The features retrieved from the user are statistically examined with the recorded features in the database by computing the F-Values from F-test, as shown in Figure 5. The user is identified as an authentic user if the F-Value is equal to the crucial value; otherwise, the user is considered a stalker.

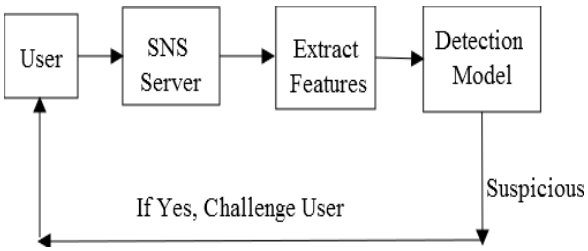


Figure 4. Principal flow of the fraud detection scheme.

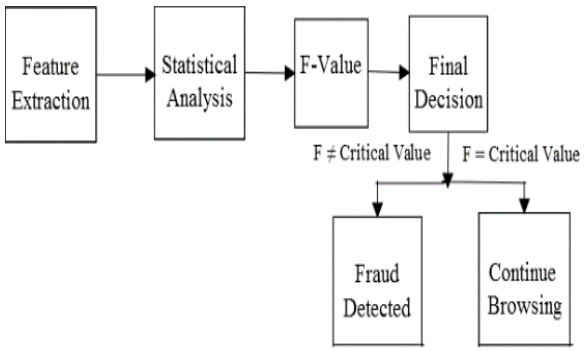


Figure 5. Proposed flow of detection model.

5.1. Modeling the Dataset

The F-Value from statistical F-test analysis is generated by comparing the test sample and training set samples to create the model. If the F-Value is 1, the test sample is determined to be an authorized user; otherwise, if the F-Value is not 1, the test sample is determined to be a stalker, and the user is prompted to re-authenticate using the password specified by the authentic user for the account owner's account.

5.1.1. Calculating F-Value Using F-Test

Any statistical test with an F-distribution under the null

hypothesis is known as an F-test. It is most commonly used when comparing statistical models that have been fitted to a data set to determine which model better matches the population from which the data were sampled. Fisher created the F-distribution to investigate the behavior of two variances from random samples drawn from two different normal populations. The F-test is used to confirm or deny the null hypothesis.

The Equation (1) calculates the F-Value for test sample and training samples.

$$F = \sum_{i=1}^{K} \frac{n_i\ (\bar{X}_i - \bar{X})^2}{K-1} \quad (1)$$

Where,

$\bar{X}_i$ denotes the sample mean in the *i*-th group.
n_i The number of observations in the *i*-th group.
$\bar{X}$ denotes the overall mean of the data.
K denotes the number of groups.

Algorithm (1) is used to find the F-Values of test data and training set data.

5.2. Access Control

The sessions predicted in Figure 3 aid in the execution of subsequent activities. If the SNS server detects a suspicious session, it prompts the user to verify with their login password; if the password is correct, the user is allowed to continue browsing; otherwise, the user's access is refused and surfing is halted, and the stalker is detected.

Algorithm 1: Calculating Critical Value matrix using F-test

Step 1: *Feature extracted from SNS*

$$Xij = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \ldots\ldots\ldots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \ldots\ldots\ldots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \ldots\ldots\ldots & X_{3n} \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ X_{m1} & X_{m2} & X_{m3} & \ldots\ldots\ldots & X_{mn} \end{bmatrix}$$

Step 2: *Calculated F-Value matrix using F-test*

$$Fij = \begin{bmatrix} F_{11} & F_{12} & F_{13} & \ldots\ldots\ldots & F_{1n} \\ F_{21} & F_{22} & F_{23} & \ldots\ldots\ldots & F_{2n} \\ F_{31} & F_{32} & F_{33} & \ldots\ldots\ldots & F_{3n} \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ F_{m1} & F_{m2} & F_{m3} & \ldots\ldots\ldots & F_{mn} \end{bmatrix}$$

Step 3: *if Fij=1 then*
AUTHENTIC USER
else
STALKER
end if

6. Dataset Collected

This section describes the dataset that was gathered for testing purposes. Before we go into the data, we should acknowledge that the stalker behavior replicated in the program may not be identical to that of a real-world stalker, and that there are certain behavioral distinctions to be found. In theory, a stalker in the real world acts in secret, leaving no evidence for the account owner, and operates on a time schedule because the account owner can return to his or her device at any moment. We proceed to both the stalkers' and account owner's activity, believing that the data set so collected may fulfill the stalkers' behavior to some level.

The behaviors of the account owners were monitored for varied time periods during repeated sessions and the common acts of the account owners were noticed and documented, just as stalkers have their own distinct behavioral patterns.

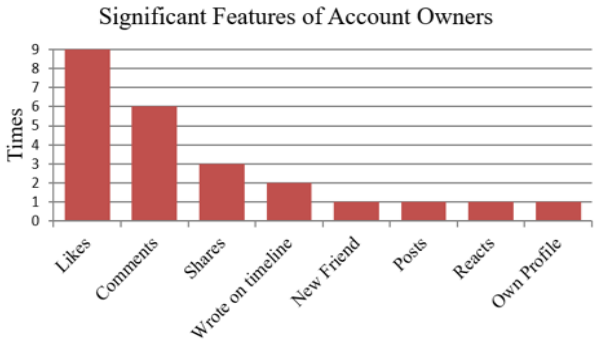


Figure 6. Significant features of account owner's.

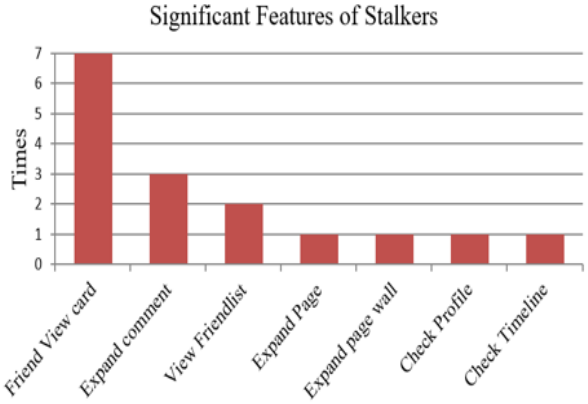


Figure 7. Significant features of stalkers.

The different activities that were recorded are depicted in Figure 6. The amount of Likes, comments, wrote on timeline, new friend, posts, responds, and changing their own profile for various time limitations are among the key features thus documented.

As shown in Figure 7, significant stalker characteristics are also detected and documented. Among the characteristics usually observed in stalkers are buddy view card, expand remark, view friend list, expand page, expand page wall, check profile, and check timeline.

6.1. Screenshots using R-Studio IDE

The simulations are run in the R-Studio IDE, which uses

the R programming language. Initially, the data of 20 users is collected and stored in a database, which is referred to as training data, as shown in Table 3.

The F-test is performed between the test data and the training data, and if the test data and training data are identical, the criticle value using the F-test will be 1, as shown in Figure 8. If the test and training data are not identical, the criticle value obtained using the F-test will be a nonequal number, as seen in Figure 8.

Table 3: Data Collected from 20 different Facebook Account with genuine actions (3 Months back data).

Features	P1	P2	P3	P4
Likes	4	9	9	9
Comments	4	6	1	0
Posts	1	1	0	0
New Friends	0	0	0	0
Others Timeline	0	0	0	3
Reacts	0	0	0	0
Shares	0	0	0	0
Tagged	0	0	0	0
Replied	0	0	0	0
Mean	1	1.78	1.11	1.11
Standard Deviation	1.73	3.34	2.97	2.42
Variance	3	11.19	8.86	5.86

P1-P1

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	0.7777778	0.777778
Variance	1.9444444	1.944444
Observations	9	9
df	8	8
F	1	
P(F<=f) one-tail	0.5	
F Critical one-tail	0.2908582	

P3-P4

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	1.111111111	1.111111
Variance	8.861111111	5.861111
Observatic	9	9
df	8	8
F	1.511848341	
P(F<=f) on	0.286149035	
F Critical o	3.438101233	

Figure 8. F-test Conducted on different variables between test data and training data (Green color is for genuine user and red color is for stalker).

Each set of data has an F-Value generated for it, which is organized into an adjacent matrix. The R-program is performed in R-Studio, and the CSV file generated from the adjacent matrix is used as input, for which the ROC curve and accuracy curve are presented in Figures 9 and 10.

7. Performance Evaluation

This section assesses the suggested detection model. This model outperforms the other technique developed by Wu *et al*. [34] in the same in loco scenario for detecting fraud on SNS services

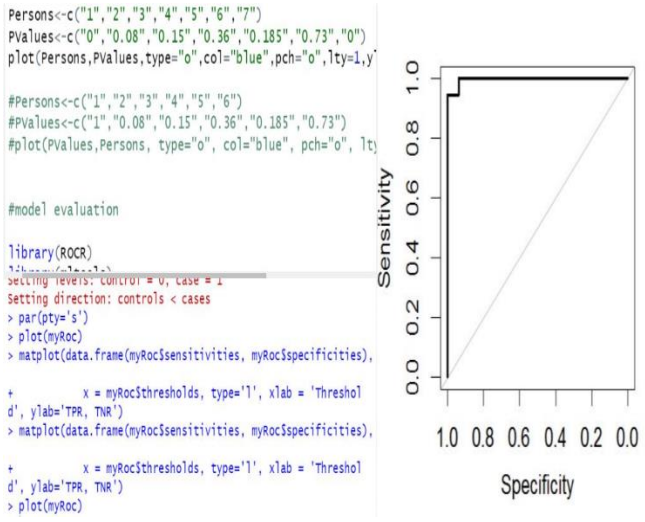


Figure 9. ROC curve at different thresholds.

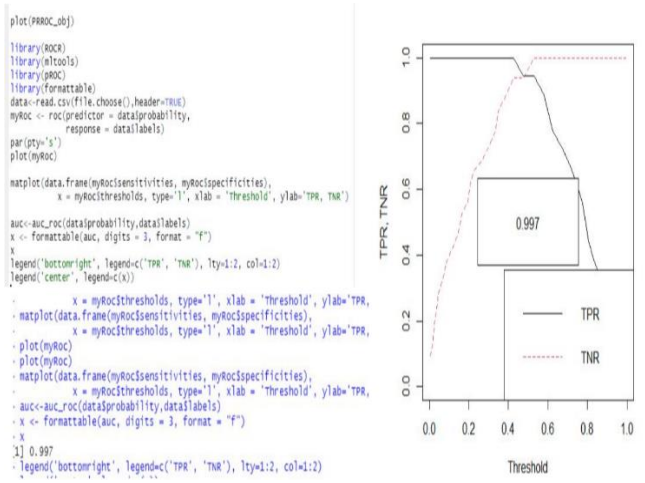


Figure 10. Accuracy curve at different thresholds.

To assess the detection, the model is simulated over a variety of time periods, and the results are compared to the detection strategy suggested by Wu *et al.* [34].

Consider the time duration of the model are N minutes, the behavioral features are extracted for various timing sessions such as N=5, 10, 15, 20, 25, 30 minutes respectively and compared.

7.1. Performance Evaluation Characteristics

A Receiver Operating Characteristic (ROC) curve is presented to show the performance of the detection model at N=30 minutes. Plotting the Genuine Accept Rate (GAR)/True Positive Rate (TPR) on the y-axis and the False Accept Rate (FAR)/ False Positive Rate (FPR) on the x-axis yields the ROC curve [3].

1. *False Accept Rate (FAR)/ False Positive Rate (FPR)*: the FAR, is a measure of the risk that a biometric security system may accept an unauthorized user's access attempt incorrectly. The FAR of a system is commonly calculated by dividing the number of incorrect acceptances by the number of identification tries [24].

$$FAR\ \% = \frac{FP}{FP + TN} X100 \quad (2)$$

Where, FP=Imposter Score Exceeding Threshold.

False Positive (FP)+True Negative (TN)=Total number of attempts.

2. *False Rejection Rate (FRR)/True Negative Rate (TNR)*: the FRR is defined as the ratio of instances rejected incorrectly to the total number of tries [32].

$$\text{FRR } \% = \frac{\text{FN}}{\text{FN} + \text{TP}} \text{X100} \quad (3)$$

Where, FN=Genuine scores Exceeding threshold.

False Negative (FN)+True Positive (TP)=Total number of attempts.

3. *Genuine Acceptance Rate (GAR)/ True Positive Rate (TPR)*: the GAR is defined as correctly accepted users by the system. It is given by

$$\text{GAR\%=100-FRR\%} \quad (4)$$

7.2. Performance at 30 Minutes

The total number of features selected for experimentation are 15, F-Values for each feature with each subject is found and it is observed that the ratio of positive instances to negative instances in the dataset D is 1.5:1.

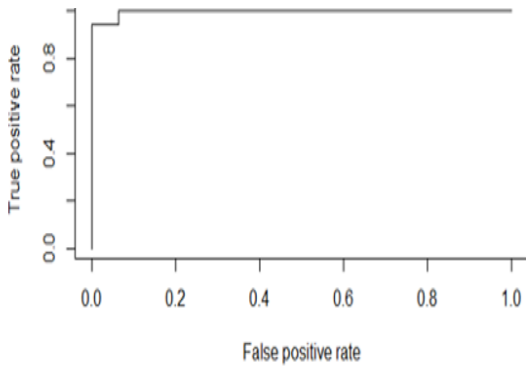


Figure 11. ROC Curve for 30 minutes at various threshold levels.

The ROC curve of the model with 98 % TPR and an FPR of 4.8 % is shown in Figure 11

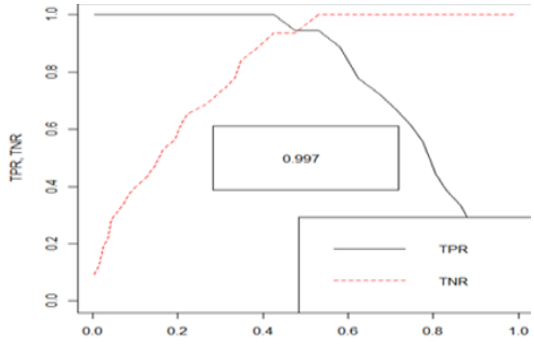


Figure 12. Accuracy calculated from TPR, TNR plotted for different threshold values.

The TPR and TNR with regard to different thresholds are plotted on the graph given in Figure 12, where it is seen that at 0.5 threshold value, 98% TPR, 95.2 % TNR and 4.8% FPR, are attained with 99.7% accuracy.

7.3. Performance Detection at Various Time Periods

To avoid the compromise of sensitive information on the same device, the statistical technique is used for different time periods, altering the timing from N=5, 10, 15, 20, 25, 30, 30 minutes, and training the detection model for each session with 15 feature selection and 50 F-Values. The accuracy attained by the various timing models is shown in Figure 13; it can be shown that after N=5 minutes, the result is stable with an accuracy of 97.9%, which is greater than the accuracy rate reached by Wu *et al*. [34], i.e., 90% accuracy rate when N=7 minutes.

As a result, it can be inferred that the suggested statistical technique outperforms the machine learning method given by Wu *et al*. [34] in terms of accuracy and stability. It's also important to note that the accuracy rate is greater at 1 minute (95%) than at 2 minutes (80%), as proposed by Wu *et al*. [34].

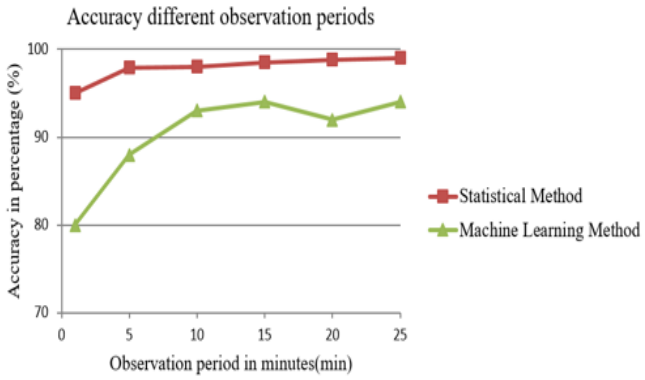


Figure 13. Accuracy of the detection scheme with various observation periods. The graph shows the detection model with good accuracy at 5 minutes. The graph also shows the comparison between In-situ detection method and the proposed statistical method.

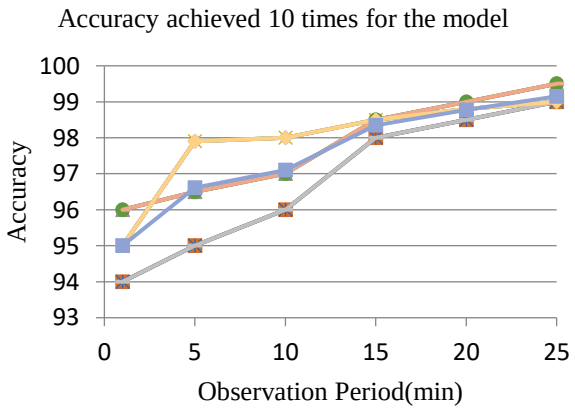


Figure 14. Accuracy for randomly permuting the model 10 times with 10-fold cross validation.

The model's resilience is tested by randomly permuting D for 10 times with 10-fold cross validation [11, 12, 18], which trains one model for each of the 10 permutations. The mean accuracy and standard deviation of the 10 models are displayed in Figure 14.

As can be seen in Figures 13 and 14, the model's performance improves when cross validations are used, implying that the suggested statistical detection scheme's performance is relatively stable.

8. Comparative Analysis

The technique used in the proposed work is compared with the outcomes of various research works carried out by researchers using different methodologies in Figure 15. And Table 4.

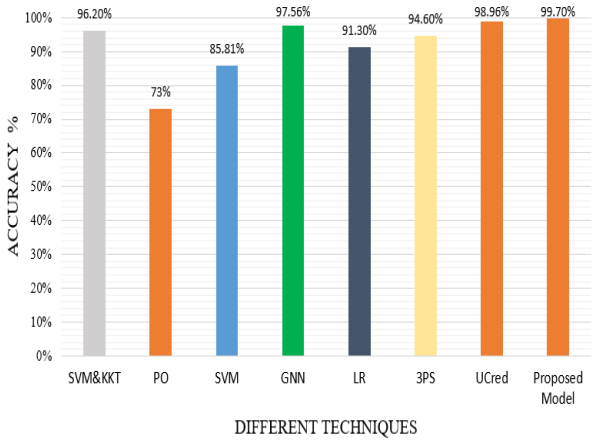


Figure 15. Comparative Analysis of Proposed Technique with other techniques with respect to accuracy.

Table 4. Comparison of the accuracy of several approaches and the proposed technique.

SL.No.	Author	Technique	Accuracy (%)
1	Wu *et al.* [34]	Support Vector Machine(SVM) & Karush-Kuhn-Tucker (KKT)	96.20
2	Tsikerdekis [28]	Proactive Approach(PO)	73
3	Walt and Eloff [31]	Support Vector Machine(SVM)	85.81
4	Chaudhary *et al.* [4]	Graph Neural Network(GNN)	97.56
5	Hu *et al.* [14]	Logistic Regression(LR)	91.30
6	Raja and Raja [25]	Publicly Privacy Protection System(3PS)	94.60
7	Verma *et al.* [30]	User Credibility (UCred)	98.96
8	Proposed Method	Statistical Analysis with F-test	99.7

The proposed technique in this paper performs better than other methods, such as Support Vector Machine (SVM) and Karush-Kuhn-Tucker (KKT) [34], Proactive Approach (PO) [28], Support Vector Machine (SVM) [31], Graph Neural Network (GNN) [4], Logistic Regression (LR) [14], Publicly Privacy Protection System (3PS) [36], User Credibility (UCred) [33], used by the authors in their respective research.

9. Conclusions

This research presents an in loco detection approach, which involves visiting SNS accounts using the same device, network, and IP address. In this research, using Facebook as a case study, a statistical analysis approach is applied in the detection scheme to distinguish between stalker and account owner browsing habits. This suggested approach outperforms the previously constructed model in terms of accuracy, TPR, and FPR. For a one-minute observation time, the detection strategy suggested in this study beats the previously existing technique. The accuracy rate acquired using statistical analysis approach is 99.7%, with a TPR of 98 % and an FPR of 4.8%, which is higher than the result obtained by Wu S H in their article. The average accuracy attained during a 1-minute observation period is 95%, indicating that the statistical analysis approach outperforms the machine learning technique used by Wu *et al.* [34].

The detection model created in this paper can be applied to a variety of other Online Social Networking (OSN) websites (such as Instagram, Gmail, WhatsApp, and Twitter) that demand login information and authenticate legitimate users in order to identify real users and stop stalkers from accessing account owners' accounts.

References

[1] Alruily M., "Issues of Dialectal Saudi Twitter Corpus," *The International Arab Journal of Information Technology*, vol. 17, no. 3, pp. 367-374, 2020.

[2] Benevenuto F., Rodrigues T., Cha M., and Almeida V., "Characterizing user Behavior in Online Social Networks," *in Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, Chicago Illinois USA, pp. 49-62, 2009.

[3] Cann B. and Ross A., "Relating ROC and CMC Curves via the Biometric Menagerie," *in Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems*, Washington, 2013.

[4] Chaudhary A., Mittal H., Arora A., "Anomaly Detection using Graph Neural Networks," *in Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*, Faridabad, pp. 346-350, 2019.

[5] Concepcin-Snchez J., Molina-Gil J., Caballero-Gil P., and Santos-Gonzlez I., "Fuzzy Logic System for Identity Theft Detection in Social Networks," *in Proceedings of the 4th International Conference on Big Data Innovations and Applications IEEE Computer Society*, Barcelona, pp. 65-70, 2018.

[6] Constine J., "Facebook Asks Every User for a Verified Phone Number to Prevent Security Disaster," http://techcrunch. com/2012/06/14/facebook-security-tips/, Last Visited, 2021.

[7] Constine J., "Facebook has Users Identify Friends in Photos to Verify Accounts, Prevent Unauthorized" http://www. inside facebook. com/2010/07/26/facebook-photos-verify/, Last Visited, 2021.

[8] Daribay A., Obaidat S., and Krishna P., "Analysis of Authentication System Based on Keystroke Dynamics," *in Proceedings of the International Conference on Computer Information and Telecommunication Systems*, Beijing, pp. 1-6, 2019.

[9] Egele M., Stringhini G., Kruegel C., and Vigna C., "COMPA: Detecting Compromised Accounts on Social Networks," *in Proceedings of the Symposium Network and Distributed System Security*, San Diego, pp. 83-91, 2013.

[10] Facebook Removal of Offline-Access Permission http://developers.facebook.com/roadmap/offlinea ccess removal/, Last Visited, 2021.

[11] Gitte V. Hendrik B., "On Estimating Model Accuracy with Repeated Cross-Validation," *in Proceedings of the 21st Belgian-Dutch Conference on Machine Learning*, Ghent, pp. 39-44, 2012.

[12] Hastie T., Tibshirani R., Friedman J., *Elements of Statistical Learning: Data Minuteing, Inference, and Prediction*, Springer, 2019.

[13] https://www.facebook.com/help/2353532535059 47, Last Visited, 2021.

[14] Hu X., Zhang X., and Lovrich N., "Forecasting Identity Theft Victims: Analyzing Characteristics and Preventive Actions through Machine Learning Approaches," *An International Journal of Evidence-based Research, Policy, and Practice*, vol. 16, no. 4, pp. 465-494, 2020.

[15] Liu Z., Wang Y., Dontcheva M., Hoffman M., Walker S., and Wilson A., "Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 321330, 2017.

[16] Lu L., Dunham M., and Meng Y., "Mining Significant Usage Patterns from Clickstream Data," *in Proceedings of the 7th International Conference on Knowledge Discovery on the Web: Advances in Web Mining and Web Usage Analysis WebKDD05*, Chicago, pp. 1-17, 2005.

[17] Mah P., "Stored Passwords Add to Mobile Security Risks" http://www. It business edge. com/cm/blogs/ mah/stored passwords-add-to-mobile-security-risks/?s=47183, Last Visited, 2021.

[18] McLachlan G., Do K., Ambroise C., *Analyzing Microarray Gene Expression Data*, Wiley, 2004.

[19] Montgomery A., Li S., Srinivasan K., and Liechty J., "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, vol. 23, no. 4, pp. 579595, 2004.

[20] Morales A., Ferrez J., Tolosana R., Garcia J., Galbally J, Barrero M., Anjos A., and Marcel S., "Keystroke Biometrics Ongoing Competition," *IEEE Access*, vol. 4, pp. 7736-7746, 2016.

[21] Nuakoh E., and Anwar M., "Detecting Impersonation in Social Network Sites (SNS) Using Artificial Immune Systems (AIS), SoutheastCon, St. Petersburg, 2018.

[22] Park Y., OHare N., Schifanella R., Jaimes A., and Chung C., "A Large-Scale Study of User Image Search Behavior on The Web," *in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems CHI*, Seoul, pp. 985-994, 2015.

[23] Peacock A., Ke X., and Wilkerson M., "Typing patterns: A Key to User Identification," *IEEE Security Privacy*, vol. 2, no. 5, pp. 40-47, 2004.

[24] Prabhakar S. and Shankaraiah., "Multimodal Biometric Decision Fusion Security Technique to Evade Immoral Social Networking Sites for Minors," *Applied Intelligence*, vol. 53, no. 2, May 2022.

[25] Raja M. and Raj L., "Detection of Malicious Profles and Protecting Users in Online Social Networks" *Wireless Personal Communications*, vol. 127, pp. 107-124, 2021.

[26] Senecal S., Kalczynski P., and Nantel J., "Consumers Decision- Making Process and Their Online Shopping Behavior: A Clickstream Analysis," *Journal of Business Research*, vol. 58, no. 11, pp. 1599-1608, 2005.

[27] Technologies C Phone Data Makes 4. 2 Million Brits Vulnerable to Id Theft [Online]. Available: http://www. credant. com/news-a-events/press-releases/69-phone- data-makes-42-million-britsvulnerable-to-id-theft. html, Last Visited, 2021.

[28] Tsikerdekis M., "Identity Deception Prevention using Common Contribution Network Data," *Ieee Transaction on Information Forensics and Security*, vol. 14, no. 8, 2015.

[29] Varol O., Ferrara E., Davis C., Menczer F., and Flamminutei A., "Online Human-Bot Interactions: Detection, Estimation, and Characterization," *Proceedings of the International AAAI Conference on Web and Social Media*, Montreal, pp. 280-289, 2017.

[30] Verma P., Agrawal P., Madaan V., and Gupta C., "UCred: Fusion of Machine Learning and Deep Learning Methods for user Credibility on Social Media," *Social Network Analysis and Mining*, vol. 12, no. 54, 2022.

[31] Walt E. and Eloff J., "Using Machine Learning to Detect Fake Identities: Bots Vs Humans," *IEEE Access*, vol. 6, pp. 6540 -6549, 2016.

[32] Wang G., Konolige T., Wilson C., Wang X., Zheng H., and Zhao B., "You Are How You Click: Clickstream Analysis for Sybil Detection," *in Proceedings of the 22Nd Usenix Conference on Security SEC13 Usenix Association* Washington, pp. 241-256, 2013.

[33] Wang G., Zhang X., Tang S., Zheng H., and Zhao B., "Unsupervised Clickstream Clustering for User Behavior Analysis," *in Proceedings of the CHI Conference on Human Factors in Computing Systems*, San Jose California, pp. 225-236, 2016.

[34] Wu S., Chou M., Tseng C., Lee Y., and Chen K., "Detecting in Situ Identity Fraud on Social Network Services: A Case Study with Facebook," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2432-2442, 2017.

[35] Yu R., Lost Cellphones Added up Fast in 2011 http://usatoday30.usatoday.com/tech/news/story/2012-322/lost-phones/53707448/1, Last Visited, 2021.

[36] Zhao T., Hu M., Rahimi R., and King I., "It's About Time! Modeling Customer Behaviors as the Secretary Problem in Daily Deal Websites," *in Proceedings of the International Joint Conference on Neural Networks*, Anchorage, pp. 36703679, 2017.

Shalini Hanok is Assistant Professor in the department of Electronics and Communication Engineering, at ATME college of Engineering, Mysore. She had completed her PhD in the field of wireless communication and network security, her research interest is in artificial intelligence, machine learning, data science, cybersecurity, image processing using social network analysis. She has 6 years of teaching experience and 4 years of research experience.

Shankaraiah is currently working as a Professor in the Department of Electronics and Communication in Sri Jayachamarajendra college of Engineering, JSS Science and Technology University, Mysore, India. He obtained his Ph.D.(Electrical communication engineering), from Indian Institute of Science(IISc), Bangalore. He has 27 years of teaching experience, His Research Interests are Wireless networking, Hybrid wireless networks, Context aware computing, Ubiquitous networks, Security in mobile communication, Context-aware Trust issues in Ubiquitous Computing, Privacy issues in WSN.