2019-08-24

The Feasibility of Using Behavioural Profiling Technique for Mitigating Insider Threats: Review

Alotibi, G

10.5121/csit.2019.91106
5th International Conference on Computer Science, Information Technology (CSITEC 2019)
Aircc Publishing Corporation

THE FEASIBILITY OF USING BEHAVIOURAL PROFILING TECHNIQUE FOR MITIGATING INSIDER THREATS: REVIEW

Gaseb Alotibi[1], Nathan Clarke[2], Fudong, Li[2] and Steven Furnell[2,3]

[1]Ministry of Interior, Public Security, IT Department Plymouth University, Plymouth, United Kingdom
[2]Security Research Institute, Edith Cowan University, Western Australia
[3] Center for Research in Information and Cyber Security, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa

ABSTRACT

Insider threat has become a serious issue to the many organizations. Various companies are increasingly deploying many information technologies to prevent unauthorized access to getting inside their system. Biometrics approaches have some techniques that contribute towards controlling the point of entry. However, these methods mainly are not able to continuously validate the users reliability. In contrast behavioral profiling is one of the biometrics technologies but it focusing on the activities of the users during using the system and comparing that with a previous history. This paper presents a comprehensive analysis, literature review and limitations on behavioral profiling approach and to what extent that can be used for mitigating insider misuse.

KEYWORDS

insider threat, behaviouial profiling, insider misuse

1. INTRODUCTION

The number of users accessing internet services across the world has been rapidly increasing and has crossed 3.5 billion recently [1].Their daily usage includes web browsing, entertainment (watching online videos, playing games etc.), communication (e.g. VoIP calls), finance (e.g. online shopping) and office applications (e.g. e-Google docs). Indeed, many of the traditional desktop applications such as Office are now being used as Internet-based services [2] [3]. According to the Office for National Statistics in the UK (2018), 86% of the UK population has daily access to the internet [4]. Within creasing use of mobile technology devices, the Internet in the UK, is mainly accessed through smartphone and portable computer devices with 72% and 83% of subscribers respectively [4]. In addition, a significant trend exists in the use of mobile application has been experienced. According to the Statistical Portal (2018) the proportion of adults who watched videos on YouTube or similar increased by 15 percentage points, from 47% in 2016 to 62% in 2018[5]. Indeed Just under half (46%) of all UK businesses identified at least one cyber security breach or attack in the last 12 months. This rises to two-thirds among medium firms (66%) and large firms (68%). According to Ponemon 2017, Fifty-nine percent of breaches in the Middle East and 52 percent of breaches in the United States were due to hackers and criminal insiders (6). Another recent study by Verizone's (2017)has found that about 25% of information breaches in the world involved internal actors [7].With the increase in the mobile and

Natarajan Meghanathan et al. (Eds) : CSITEC, CMCA, SP, NECOM, ADCO - 2019
pp. 59-71, 2019. © CS & IT-CSCP 2019 DOI: 10.5121/csit.2019.91106

internet devices usage, there has been an increase in the security threats in relation to the information across the internet or within the information systems. A variety of approaches, such

as hacking, Denial of Service (DoS), social engineering and malicious software are used by the attackers in illegally accessing the information [8]. Various studies have shown that such approaches are often successful. Right from an individual to an organization, anyone can be a victim of security attacks. These attacks are on rise in the recent years. According to McAfee and the Center for Strategic and International Studies (CSIS) revealed that cybercrime cost the world between $445 and $608 billion in 2017 [9].

These studies reflect the seriousness of the security issues being faced by various organizations and the individuals; and also highlight the importance and the need for effective security tools required for safeguarding the information systems. In investigating these security threats, it is very essential to understand the various types of threats, and the efficiency of the security tools available for mitigating or avoiding such threats. These threats can be analysed based on their occurrence or initialization; and can be categorized accordingly in to internal and external security threats. Whilst external threats remain an issue, security controls exist to protect outsiders from getting into systems. They do not however often help with the problem of insider threats. Insider threats have grown to be an increasingly significant problem. According to Information Security Breaches survey in 2015, insider misuse (e.g. unauthorized access) represents 65% of the security breaches in the large organizations [10]. In addition, newer technologies such as smartphone and tablets have become one of the main driving factors behind the security breaches. In large organizations, smartphone and tablets have caused 15% of security or data breaches, doubling the percentage comparing with the previous year [10].

There are various approaches that can be utlised to verify the users. Authentication is one of these approaches that can be accomplished by identifying someone through one or more of three factors: something they know, something they have, or something they are. Despite the fact that all three manners are using for the authentication purposes, user characteristics considered as the more likely unique technique for identifying the users which called biometrics.

The performance of detection techniques is measured by three metrics which include False Positive (FP), False Negative (FN), Equal Error Rate (EER) and Detection Rate (DR). FP refers to the percentage of normal traffic that has been detected and considered as an attack; whereas FN refers to the percentage abnormal traffic (attack) that has been detected and incorrectly considered as normal traffic. The efficiency or frequency of detecting the attacks is called as DR. To be a more accurate a better performing system, needs a high DR, and low FP and FN. The two insider threat detecting techniques identified are analyzed and reviewed in the next section.

2. BEHAVIOURAL PROFILING APPROACH

The research in this area started in the 1990s and since then, there have been a variety of studies that have examined the behavioural profiling from different perspectives, such as fraud detection, intrusion detection and authentication. Behaviour profiling is utilised to verify a user based on the previous history, it then creates a user template which can be used to decide whether this kind of activity belongs to a legitimate user or not. The sections below discuss the different systems that utilised behavioural profiling.

2.1. Fraud Detection

Fraud detection is an issue that many organisations in many countries suffer from. The measurement of the efficiency of each technique mainly used detection rate (DR) where the

system correctly finds out the suspicious activities. One of the earliest studies on fraud detection was part of the European commission-funded ACTS ASPeCT (Advanced security for Personal Communications Technologies) project (Burge and

Taylor, 1997). This study was conducted on the Vodafone network in the UK. The approach used behavioural profiling to detect any abnormal activity on the number. Toll ticket tools used to extract all the necessary information about calls activities. In addition, two methods for utilised behavioural profiling in this study Current Behaviour Profile (CBP), which means the short sequence of activity and Behaviour Profile History (BPH) to save the long sequence of activity and both of them are managed based on the queue system. Over a two-month period, they used a neural network technique with an unsupervised approach where no fraudulent examples are required for training to classify the call activity, based on three parameters: national calls, intentional calls and use for supplementary reasons. Ultimately, the approach detected 75% of fraudsters and only 4% of misclassifications of valid subscriptions were included within the fraudsters list.

In 1997, Moreau et al. published a paper in which they presented the first prototype of a tool based on a supervised neural network, an unsupervised neural network and knowledge-based systems. This research provided a framework for the detection of fraud in mobile communication as a part of the European commission-funded ACTS ASPeCT project. They gathered the information for the study from toll ticket because it includes all the necessary information that could help in the detection process. The features were International Mobile Subscriber Identity (IMSI), the start date of calls, the start time of calls, duration of calls, dialled telephone numbers and national/international calls. It is quite unacceptable to detect just 64% of fraudsters by using the unsupervised approach, which has a 5% false alarm rate,due to the fact that a huge number of fraudsters will not be detected, while the main reason for this kind of technique is to determine what the problem with the communication is and to identify the potential solution to mitigate it. The other approaches are significantly better, but they still have a high number of false alarms that may disrupt communications for the valid subscribers (Moreau et al.,1997).

Samfat and Molva (1997) investigated the ability to perform intrusion detection in the visited location and within the duration of a typical call. The experiment was carried out in the simulation environment and the dataset has a similarity of GSM network, then applying a statistical classifier in order to have a user behavioural based on location and telephony features, such as the start and end of call and duration. The results of this experiment were 82% detection rate and about 40% FAR, which means a huge number of the illegitimate user is considered as an authorised user. Although the detection rate is quite satisfying, FAR is quite high which may need further research to fix it.

In 2002, Boukerche and Notare provided a fraud detection model on mobile phones by utilising a Radial Basis Function (RBF) neural network model. In order to detect call service fraud, the model uses user-calling features such as time and duration of the call. The model sends an alert message to the user and network administrator immediately when a suspicious call is identified. The dataset was collected by an unknown telecommunication service provider and contains 4,255,973 telephone calls. The result achieved by applying the RBF neural network and using 110 neurons in the hidden layer of the network, was 97% detection rate and 4.2% system error rate.

Hilas and Sahalos (2005) published a paper titled “User profiling for fraud detection in telecommunication networks”. The authors mention that the amount of telecommunication fraud has dramatically increased. Hence, searching for effective detection is important to avoid further fraudulent activities. The paper suggests obtaining user profiling to detect fraud activities by

using a statistical machine learning method, which constructs user profiles of each user and compares them with the future activities of the users to detect any abnormal activities. The methodology of this paper is to create a profile for each user from their previous history, which were stored on the organisation's database. Consequently, the authors mention the similarity measure, which consists of different parameters, such as calls made to local destinations, the duration of local calls, the number of calls to mobile destinations, the duration of mobile calls, the number of calls to national and international destinations and their corresponding duration. There are two levels of similarity, the first being to examine the equality of the number of calls in the same category, and the second is to compare the total calls durations across categories. The dataset contained more than 5,000 users, and it was collected over a one-year period at a university (Hilas and Sahalos, 2005).

The technique used was the statistical machine learning to present that constructs user profiles for the detection of fraudulent activities in telecommunications networks.The experiment relied upon a different number of parameters and extracted these parameters from the call inquiries in the large dataset. In addition, the greatest similarity achieved was 80%, which is a promising result. Furthermore, this paper does not mention anything about malicious activities from a particular user. It merely compares known parameters for known durations for known users, and that might not be enough to detect fraudulent activities in the telecommunication sphere. The same authors in 2007, tried to perform another classifier to improve the detection rate, however, the result almost remains about 80%

Ogwueleka, (2009) published a paper titled “Fraud Detectionin Mobile Communications Networks Using User Profilingand Classification Techniques”. This study utilised call data for describing behavioural patterns of users by using unsupervised Self Organisation Map neural network (clustering). The evaluation of this experiment was done using a dataset that includes 180 participantsduring a period of 75 days, in which, the results of this research concludes that FAR is about 3% which means the fraudulent transactions that accepted as legitimate and the mobile communication detection system detect most of the fraudulent transactions.
Similarly, Oayyum and Mansoor in 2010 have stated that in order to improve the detection rate of the fraud issue, two techniques could be utilised; let the user think what is the most feature that may represent the fraud and create a weight that is association priorities based on user input. The values that were given from the users to the most features that might represent the fraud by giving the highest features high value to presenting the level of dangerous that has. The study used a neural network as a classifier and evaluate the technique using a dataset that includes 300 participants. The final result from this approach was an ability to detect up 70% of the fraudulent calls.

In 2015, Subudi and Panigrahi have published a paper titled "Quarter-Sphere Support Vector Machine for Fraud Detection in Mobile Telecommunication Networks". This paper has utilised a set of features such as call duration, call type, and call frequency

along with location and time. The experiment was carried out by using the Reality Mining dataset which includes 94 users' information. The authors utilised support vector machine classifier in order to distinguish the malicious behaviour from the normal behaviour. The result of this approach was promising compared to the previous studies in this fields where the detection rate attained is 97%. The aforesaid studies found that using behavioural profiling in fraud detection technique provide promising results once the behavioural profiling template creates with sufficient information. With respect to performance, many studies have achieved positive DR and FAR and over the time where more information became available and being able to include when the user profile created such as user location.

2.2. Intrusion Detection System

This section explains using behavioural profiling in IDS which is analogous to anomaly detection. Buschkes et al.,1998 has introduced a new approach that using anomaly detection; they produced their procedure to test their hypothesis by using Bayes decision rule; In probability theory and statistics, Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event to detect any abnormal usage (Zou,2006). The study mainly focused on different features, such as call activities and location of users. The result shows that the system was able to score 87.5% detection rate.

Similarly, Sun et al. (2004) have published a paper in titled "Mobility-Based Anomaly Detection in Cellular Mobile Networks". This paper examined cell IDs transferred by a user as the features value. In addition, various services were included in this study for a user as an option in order to create an accurate user profile. Consequently, a higher order of Markov model for the detection part was used. The paper has applied the concept to a simulation environment for evaluation purposes and the final finding form this research was 87.5% Detection Rate (DR) and 15% False Alarm Rate (FAR). The second experiment that was conducted by Sun et al in 2006 has achieved a better DR (89%) and low FAR (13%). This is due to more care which was being considered in relation to the location of the user. Nevertheless, location information is quite sensitive for users, in which this approach gives the user an ability to turn it on/off as required.

Yazjiet al.(2011) published a paper in the title "Protecting private data on mobile systems based on spatiotemporal analysis". The research sought to produce a model used to detect theft in smart phone mobile by creating an anomaly spatiotemporal patterns. They used network access patterns and file system activities to build a behavioural model that permitted attack detection with a latency of 5 minutes and an accuracy of 90%. The study evaluated on Reality Mining dataset consisted of 100 participants. The authors used two techniques spatiotemporal and Trajectory. The former used when the users do not have many locations in their dataset whereas the later utilised with sufficient locations, in order to improve the performance. The result shows that the system was capable of detecting an attack within 15 minutes with 81% accuracy.

Finally, Yazji et al. (2014) produced an approach that addressed mobile device issues by making an effective correlation between the user's location and time data. They developed two statistical profiling approaches for modelling the normal captured

behaviour of the user: the former based on an experimental collective probability measure and the other one is based on Markov which provides a way to find dependencies between the current information and compare it with the previous one that owned (Clemson University,2017). Reality Mining and Geolife dataset were utilised to evaluate this approach. The result shows that the system is capable of detecting a potential intrusion with 94% accuracy. This high proportion of accuracy achieved due to the set of features utilised in this research and the duration that was taken into account in order to create a user behaviour profile template, which was 9 months.

2.3. Authentication

The majority of authentication-based systems rely on a behavioural-based mobile security system. These systems use different techniques to assess user's identity based on one or more behavioural characteristics, such as keystroke and gait. The following section provides a short description of these studies. Aupy and Clarke (2005) conducted a preliminary study to determine at which level people present a unique behavioural profile while utilising their computer desktop. The paper aimed to provide a transparent and continuous authentication of users relying on their interaction

with computers, what applications are used and when, and recently visited websites. The authors used the 'logger' application to capture user interactions. The experiment sample was 21 participants, and the data was collected over 60 days. The study utilised different features extracted by the logger application. The study used a neural network classification known as the Feed Forward Multi-Layered Perceptron (FF-MLP). FF- MLP has been commonly used in implementing the K-classification module for the character recognition (Oh and Suen, 2000). The result of the study was promising, with an overall EER average of 7%. Finally, the study faced some problems, such as the number of participant's actions being very small. Therefore, the authors decided to repeat a number of actions from the dataset many times to raise the input features.

Yazji et al. (2009) proposed an implicit user re-authentication system for the portable computer that has limited in application change and hardware replacement. The main aim of this research is to reduce the probability of unauthorised access or theft by creating a user behaviour based on a combination of filesystem activities and network access that is able to distinguish between normal user and anomalous. There are a set of features that can be obtained from file access system where the log file has a lot of useful information such as timestamps, names of the process responsible for access, locations of accessed files, and operations on the file. Different parameters of network activities were included in this study, such as the timestamp of each access, source IP address, destination IP address, protocol identifier. The study utilised K-means clustering, and in order to evaluate the approach, a period of two weeks was allocated to gather a dataset of 8 participants. The result was promising where the system was capable to score detection rate of 90% with FAR 14% (it means 14% of nonauthorized access considered as authorised) and FRR 11% (it means 11% of authorised access considered as unauthorised access or imposter).

Li et al. (2010) published a paper titled "Behaviour Profiling on Mobile Devices". The study proposed a novel behaviour-based profiling technique that is capable of building up on the weaknesses of current mobile device authentication systems by developing a comprehensive multilevel approach to profiling. One of their main features collected form device usage includes the day, time, duration and weekday. The approach was evaluated using 30 users from MIT Reality dataset and Neural network with different neuron size was applied. The findings of this are; 35.1% average EER, where top user achieved 1%EER. With this high number of EER, the authors proposed another approach that is concentrating on mobile user's application usage. They used a behavioural-based host mobile security mechanism, using some of the user's behaviour characteristics to assess the legitimacy of current users.

The experiment was conducted using a dataset that is available from the MIT Reality Mining project. The dataset was collected over 10 months as the duration of time starting from September 2004 to Jun 2005 for 106 participants. There were different types of features collected from the mobile including application level (default application, such as a phonebook), application specific (additional discriminating information, such as Internet browser) and multi-instance application (a mixture of application level and application specific). The user behaviour was created based on two types of profile techniques: static and dynamic. After applying the data on the mathematical equation, the results were 13.5%, 5.4%, 2.2% and 10% for the general application, telephone, text messaging, and multi-instance application usage respectively. This result seems to be encouraging and points to the use of this method in order to identify misuse within the mobile phone sphere.

However, some limitations may affect that decision, such as the dataset was from 2004/2005 when smart phones first emerged, hence the data is not up to date. The entity data is not sufficient for various reasons, because, for instance, applications have changed since 2004 (Li et al, 2011). Therefore, an additional experiment has conducted on a dataset consisted of 76 users in order to

build a novel behavioural profiling framework that is able to collect user behaviour based on applications usage. So, the system would reject the user's access based on a number of consecutive abnormal applications usages rather than single application usage. By applying this technique, the system has achieved better performance achieved 9.8%EER, FAR 4.1% and FRR 4.4%. Similarly, Shi et al. (2011) published a paper titled "Implicit Authentication through Learning User Behaviour". The aim of this paper is to create a new model for performing implicit authentication. The model utilised different features, such as SMS and telephony along with Browser History: they examined browser history and recorded the obfuscated domain name of each URL along with the number of visits to this URL done previously. The technique was evaluated based on two weeks of collected data from 50 users and the result was quite promising where the system achieved 95% detection rate.

Salem and Stolfo in 2011 produced modelling user search behaviour for masquerade detection technique. Their hypothesis relied on the level of knowledge that a user has about his system, in which encourages him to do a search for information in limited, targeted and unique fashion. Subsequently, this kind of behaviour might aid to distinguish legitimate users form masquerade. There were 18 participants involved in this experiment and a support vector machine as a classifier to detect the masquerade. The result showed that by using this kind of user behaviour, the system was able to detect all masquerades with a very low false positive rate of 1.1%.

In 2013, Abramson and Aha created a technique for continuous authentication purpose based on user web browsing behaviour. They claimed that using web browser has become a common activity in different places for different purposes, thus there is a chance to uniquely identify someone based on their Web browser behaviour. The main features of this technique come from the type of web browsing, number of sessions, length of session, and timestamp. The experiment was conducted using a dataset that included 10 users by using ensemble classifier. This classifier is often more accurate due to a set of classifiers whose individual decisions are joined in some way to classify a new example (Dietterich.T,2001). The results show that 24% best EER achieved.

Similarly, Fridman et al. (2015) presented an active authentication model on mobile devices based on user behaviour that is collected from four biometrics modalities: Text enters via soft keyboard, Application usage, Web site visited and Physical location via GPS. The evaluation of this system was conducted based on a dataset from 200 participants collected for at least 30 days. Decision Fusion classifier has been utilised in order to improve the overall accuracy of the system. The result showed that high accuracy can be achieved with EER 3%.

In 2016, Sbeyti (2016) suggested a novel authentication model be used complementary to the existing models. Indeed, the research sought to produce an implicit authentication through the capture of user's discriminative behavioural pattern. The system aims to reduce the number of explicit authentication.Its purpose is not to replace the common authentication approaches, but rather to accompaniment them. That is, the user can still use his chosen authentication method, but once the phone is unlocked, the implicit authentication takes responsibility to determine if the user is indeed the owner or an attacker. And to accomplish this, the author relied on two features applications usage and duration of each usage for each particular user. Then a mathematical algorithm has implemented and being work at the run time to determine if the user is an owner of the smart phone or an attacker. The experiment evaluated on Android dataset consisted of 30 users and the result shows that the system was able to achieve True positive attained 76%.

Al-Bayati et al. (2018) have published a paper in title of "Continuous Identity Verification in Cloud Storage Services using Behavioral Profiling". The authors focused upon cloud application and took Dropbox as a sample to see the feasibility of using behavioral profiling in this field. The

real user interactions collected from 30 users. They used different classifications such as SVM, FF MLP and RF, found that FF MLP gives a promising result with EER around 5.8%.

2.4. Identification

Much of the prior art has focused on the application of behavioural profiling in the simpler verification mode. In identification mode, there are a few papers discussed using behavioural profiling model. Yang (2010) published a paper about web user behavioural profiling for user identification. The author tries to build a user profile based on web usage patterns. He creates a user profile for different known users based on their web activities and anonymous web sessions. The framework using Lift-based profiling; performs best when the number of users is small; consists of different stages, starting from initiating a user behavioural pattern, including the number of web

sessions, overall user behavioural patterns, retrieve patterns for all the users, and measure the distance between two profiles. The dataset consisted of 50,000 users over the one-year period, and it was provided from a commercial data vendor. It included the web browsing history of a panel of users who were included in this experiment. However, the author selected a few users (with id # 2, 5, 10, 20, 50, 75, 100) and chose
10 different sliding window sizes,1, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 in order to investigate how sliding window size affects the predictive accuracy. The accuracy of the classification methods showed that the level of accuracy attained 90%. Regardless of the fact that accuracy rate was promising, the study has few serious issues which may affect the reliability especially when the author has evaluated the approach on just 10 users from the large dataset that contained 50,000 users.

Similarily, Herrmann et al. (2010) proposed a re-identification model based on behavioural profiling that is collected form web users' sessions in order to overcome the changing of IP addresses issue. They mentioned that the websites that are retrieved by an individual reflect – at least to some degree – his or her interests, habits and social network. The URL of some pages may even disclose the user's identity. So, they tried to ink the web sessions of a given user only based on a record of his past activities on the web. To overcome some privacy issue the features of the study were collected form HTTP requests, in which they used destination host name only. The evaluation of the approach was conducted in a real-world dataset from 28 users by using Naïve Bayes classifier and assuming a limited knowledge of an attacker who can only observe the host names visited. The result showed that the system was able to re-identify up to 50 % of the users, about 80 % of the time. This technique relied on the observed access frequencies of hosts based on the name of destination hosts in order to create user profiles. There is a scalability issue in this study where just 28 users utilised to gauge whether the provider can correctly identify their users or not based on web site visited history. In addition, perhaps using sophisticated attacks, it can be argued that this study is not going to be efficient.

In 2016 , Herrmann et al published a paper in title of “Behavior-Based Tracking of Internet Users with Semi-Supervised Learning”. The paper mainly focused on lined the user with the internet session utlised. The dataset was two months duration with more than 3800 user that include DNS traffic. The paper used DNS name with the session time and date. The authors simulate daily changing IP addresses and perform a rudimentary feature selection by removing all domains that have been queried in fewer than six sessions. K-means implement here to measure to what extend the features successfully achieved the aims to the paper. The results show that the approach can correctly link the session to the right user with 87% accuracy.

In the same view, Sy et al.(2018) proposed a new technique for tracking users called tracking users across the Web via TLS Session Resumption. The authors relied on TLS seesions of the web bowering and they found that the majority of them support tracking for at least 30 mints s based on TLS session resumption and this help to create their features which are lifetime of session id, lifetime of session tickets and DNS name. The data set was about 3862 students over period of two months collected from e University of Regensburg. The results showed that about 65% of all users in our dataset can be tracked permanently.

3. DISCUSSION

The table below summaries the different studies of behavioral profiling. Although fraud detection and IDS research have contributed in a view of understanding the technique, further research is needed with reference understanding the applicability of behavioural profiling within a networking-based application. A variety of studies have examined user behavioural profiling from different perspectives such as fraud detection, intrusion detection, authentication and identification.

In fraud detection where the majority of this research has been supported by communications and mobile companies, the main features that utilised to create a user profile extracted from telephony activities such as call number, time and duration and later country of the calling included. Thus, the accuracy does vary across studies with high FAR scores in some of them. This is because of the features that were used to create a user behavioural profiling were quite limited and the samples were quite huge.

In contrast, there are various researchers have been conducted to explore the possibility of applying user behavioural profiling to increase the level of security of the mobile device. Indeed, the early studies in this field have utilised anomaly- based detection to determent any abnormal activities with the mobile devices such as studies (10-15) as illustrated in table 1. whilst recent research tends to create a user profile based on application usage within authentication and identification model and achieved a high accuracy. This is due to that nature of features that included which were contributing towards achieving a higher performance as can be seen in table 1 studies (16-31). The techniques mainly utilised to verify a user by storing the pervious user activities that used it to create the user profile and comparing it with current user activities to be able to decide whether this user is legitimate or illegitimate. The nature of the behavioural profiling features has played a core key into creating an accurate user profile.

Table 1 Behavioural profiling Papers

No.	Author(s)	BP source	Classifier	# User	Performance%	Category
1	Burge and Taylor, 1997	Telephony	Neural Network	110	DR=75, FAR=4	Fraud detection
2	Moreau et al.,1997	Telephony	Neural Network	600	DR=90,FAR=10	Fraud detection
3	Samfat and Molva,1997	Telephony	Statistical	Non	DR=82.5,FAR=40	Fraud detection
4	Boukerche and Notare,2002	Telephony	radial Basis function	100	DR=97,FAR=4.2	Fraud detection
5	Hailas and Sahalos,2005	Telephony	statistical machine learning	5000	DR=80	Fraud detection
6	Hailas and Sahalos ,2007	Telephony	decision tree	5000	DR=80	Fraud detection
7	Qgwueleka,2009	Telephony	Neural Network and probabilistic	180	FAR=3	Fraud detection
8	Qayyum and Mansoor,2010	Telephony	Neural network	300	DR=70	Fraud detection
9	Subufhi and Panigrahi ,2015	Telephony	Support Vector Machine	94	DR=97	Fraud detection
10	Buschkes et al,1998	Mobility	Bayes decision	Non	DR=87.5	IDS
11	sun et al,2004	Mobility	Markov	Non	DR=87.5,FAR=15	IDS
12	Hall et al, 2005	Mobility	Decision rule	50	DR=50, FAR=50	IDS
13	sun et al,2006	Mobility	Markov	Non	DR=89,FAR=13	IDS
14	Yazji et al.,2011	Mobility	Spatio-temporal model and trajectory-based	100	DR=81	IDS
15	Yazji et al,2014	Mobility	Non	100	DR=97	IDS
16	Aupy and Clarke,2005	PC usage	Neural Networks	21	ERR=7	Authentication
17	Yazji et al., 2009	File access activity and network event	K-Means	8	DR=90,FAR=14, FRR=11	Authentication
18	Li et al, 2010	Telephony, device usage, Bluetooth and network scanning	Neural Networks	30	EER=13.5,35.1, 35.7	Authentication
19	Shi et al,2011	Telephony, SMS,Browsingand Mobility	Gaussian Mixture Model	50	DR=95	Authentication
20	Salem and Stolfo,2011	file access activity	SVM	18	FAR=1.1	Authentication
21	Li et al, 2011	Applications Telephony,SMS	Neural Networks	76	EER=13.5,2.2, 5.4	Authentication
22	Dimitrios et al,2012	Telephony,SMS, Browsing	machine learning algorithms	35	DR=98.5, EER=1.6	Authentication
23	Abramson and Aha, 2013	web browsing	Ensemble	10	EER=24	Authentication

24	Li et al 2014	Application usage	Neural Networks	76	EER=9.8	Authentication
25	Fridman et al,2015	text,application, web and location	Probabilities	200	EER=3	Authentication
26	Sbeyti,2016	Application usage	Mathematic algorithm	30	TP=76%	Authentication
27	Al-Bayati et al.2018	Application usage	Neural Networks	30	EER=5.8	Authentication
28	Yang,2010	web browsing	decision trees	7	DR=91	Identification
29	Hermann et al.,2010	web browsing	Naïve Bayes	28	DR= 50	Identification
30	Herrmann et al,2016	DNS name ,IP	k-means	3800	DR=87	Identification
31	Sy et al.(2018)	DNS,session id, ticket	Mathematic algorithm	3862	DR=68	Identification

4. CONCLUSION AND FUTURE WORK

Although behavioural profiling has been used in different studies particularly in a verifications mode and scoring a positive performance, there are a few studies that have discussed and examined the identification mode by using behavioural profiling and shown that it is viable, even if performance could be improved to overcome some of the scalability issues. However, it has been shown that using behavioural profiling can contribute towards creating an accurate template of the user profile. So, the study will be expanding to create a framework that collect more features from user activities and enhancing that in the machine learning process to measure the feasibility of using Behavioral profiling in identification context.

REFERENCES

[1] Internetlivestats, “Internet live stats”, [online], http://www.internetlivestats.com/internet-users/, date accesed 2nd March 2017.

[2] Microsoft, “Office 365 for business FAQ,[online]https://products.office.com/en-us/business/microsoft-office-365-frequently-asked-questions,date accessed 1st March 2016.

[3] Library, “Introduction to Google Docs”,[online], http://www.lfpl.org/jobshop/docs/google-docs.pdf, date accessed 1st march 2017.

[4] National Statistical in the UK, [online], http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2015-08-06, date accessed 15th March 2017.

[5] Statistical portal,“Application downloaded”,[online], http://www.statista.com/statistics/241587/number-of-free-mobile-app-downloads-worldwide/, date accessed 25th March 2017.

[6] Microsoft 2016, “ Security Threats”,[online] https://msdn.microsoft.com/en-us/library/cc723507.aspx, accessed on 14th April 2017.

[7] Security intelligence,”Global cost of cybercrime exceeded $600 billion in 2017”,[online], https://securityintelligence.com/news/global-cost-of-cybercrime-exceeded-600-billion-in-2017-report-estimates/, accessed 10th December, 2018

[8] Verizon, “2014 Data Breach Investigations Report”, [online], http://www.verizonenterprise.com/DBIR/2014/, date access 28th March 2016.

[9] PwC, “2015 Information security breaches survey”, [online], https://www.pwc.co.uk/assets/pdf/2015-isbs-technical-report-blue-03.pdf, date accessed: 26th March 2016.

[10] Cisco, “Cisco Visual Networking Index: Forecast and Methodology, 2013-2018”, [online], http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html, date accessed: 20 December 2014

[11] L.D. Merkle, “Automated Network Forensics”, Proceedings of the conference on genetic and evolutionary computation (GECCO 2008), pp 1929-1932, in press
[12] Al-Bayat et al.(2018): Continuous Identity Verification in Cloud Storage Services using Behavioural Profiling,, 17th European Conference on Cyber Warfare and Security ECCWS 2018, 28-29 June, ISBN: 978-1-911218-85-2, pp1-10, 2018
[13] Moreau, Y., Preneel, B., P., B., Shawe-Taylor, J., Stoermann, C., and Cooke, C. "Novel Techniques for fraud detection in mobile telecommunications networks," in: ACTS Mobile Summit, Grenada Spain, 1997.
[14] P. Burge and J. Shawe-Taylor, “Detecting Cellular Fraud Using
[15] Adaptive Prototypes,” in Proc of AI Approaches to Fraud Detection and Risk Management, 1997, pp. 9–13.
[16] D. Samfat and R. Molva, “IDAMN: an intrusion detection architecture for mobile networks,” IEEE J. Sel. Areas Commun., vol. 15, no. 7, pp. 1373–1380, 1997.
[17] R. Buschkes, D. Kesdogan, and P. Reichl, “How to increase security in mobile networks by anomaly detection,” in Proceedings 14th Annual Computer Security Applications Conference (Cat. No.98EX217), 1998.
[18] B. Sun, F. Yu, K. Wu, and V. C. M. Leung, “Mobility-based anomaly detection in cellular mobile networks,” Proc. 2004 ACM Work. Wirel. Secur. - WiSe ’04, p. 61, 2004.
[19] Z. W. R. Y. F. Sun B; Chen, “Towards adaptive anomaly detection in cellular mobile networks,” in CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006., 2006, vol. 2, pp. 666–670.
[20] J. Hall, M. Barbeau, and E. Kranakis, “Anomaly-based intrusion detection using mobility profiles of public transportation users,” WiMob’2005), IEEE Int. Conf. Wirel. Mob. Comput. Netw. Commun. 2005., vol. 2, pp. 17–24, 2005.
[21] C. S. Hilas and J. N. Sahalos, “User Profiling for Fraud Detection in Telecommunication Networks,” in 5th International Conference on Technology and Automation, 2005.
[22] C. Hilas and J. Sahalos, “An application of decision trees for rule extraction towards telecommunications fraud detection,” Lect. Notes Comput. Sci., vol. 4693, no. 2, pp. 1112–1121, 2007.
[23] C. Hilas, S. Kazarlis, I. Rekanos, and P. Mastorocostas, “A genetic programming approach to telecommunications fraud detection and classification,” Proc. 2014 Int. Conf. Circuits, Syst. Signal Process. Commun. Comput., pp. 77–83, 2014.
[24] F. Ogwueleka, “Fraud Detection In Mobile Communications Networks Using User Profiling And Classification Techniques,” J. Sci. Technol., vol. 29, no. 3, pp. 31–42, 2009.
[25] S. Qayyum, S. Mansoor, A. Khalid, Z. Halim, and A. R. Baig, “Fraudulent call detection for mobile networks,” 2010 Int. Conf. Inf. Emerg. Technol., pp. 1–5, 2010.
[26] S. Yazji, R. P. Dick, P. Scheuermann, and G. Trajcevski, “Protecting Private Data on Mobile Systems based on Spatio- temporal Analysis,” 2011.
[27] S. Yazji, P. Scheuermann, R. P. Dick, G. Trajcevski, and R. Jin, “Efficient location aware intrusion detection to protect mobile devices,” Pers. Ubiquitous Comput., vol. 18, no. 1, pp. 143–162, 2014.
[28] S. Subudhi and S. Panigrahi, “Quarter-Sphere Support Vector Machine for Fraud Detection in Mobile Telecommunication Networks,” Procedia Comput. Sci., vol. 48, no. Iccc, pp. 353–359, 2015.
[29] E. Shi, Y. Niu, M. Jakobsson, and R. Chow, “Implicit authentication through learning user behavior,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6531 LNCS, pp. 99–113, 2011.
[30] M. P. Dimitrios Damopoulos, Sofia A. Menesidou, Georgios Kambourakis and N. C. and S. Gritzalis, “Evaluation of anomaly- based IDS for mobile devices using machine learning classifiers,” Secur. Commun. Networks, vol. 5, no. 1, pp. 3–14, 2012.
[31] F. Li, N. Clarke, M. Papadaki, and P. Dowland, “Behaviour profiling on mobile devices,” Proc. - EST 2010 - 2010 Int. Conf. Emerg. Secur. Technol. ROBOSEC 2010 - Robot. Secur. LAB-RS 2010 - Learn. Adapt. Behav. Robot. Syst., no. 2010, pp. 77–82, 2010.
[32] F. Li, N. Clarke, M. Papadaki, and P. Dowland, “Misuse Detection for Mobile Devices Using Behaviour Profiling,” Int. J. Cyber Warf. Terror., vol. 1, no. 1, pp. 41–53, 2011.
[33] F. Li, N. Clarke, M. Papadaki, and P. Dowland, “Active authentication for mobile devices utilising behaviour profiling,” Int. J. Inf. Secur., vol. 13, no. 3, pp. 229–244, 2014.
[34] L. Fridman, S. Weber, R. Greenstadt, and M. Kam, “Active Authentication on Mobile Devices via Stylometry , Application Usage , Web Browsing , and GPS Location,” pp. 1–10, 2015.

[35] A. Aupy and N. Clarke, "User Authentication by Service Utilisation Profiling," Adv. Netw. Commun. Eng. 2, p. 18, 2005.
[36] S. Yazji, X. Chen, R. P. Dick, and P. Scheuermann, "Implicit user re-authentication for mobile devices," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5585 LNCS, pp. 325–339, 2009.
[37] M. Ben Salem and S. J. Stolfo, "Modeling user search behavior for masquerade detection," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6961 LNCS, pp. 181–200, 2011.
[38] Y. Yang, "Web user behavioral profiling for user identification," Decis. Support Syst., vol. 49, no. 3, pp. 261–271, 2010.
[39] M. Abramson and D. Aha, "User Authentication from Web Browsing Behavior," Twenty-Sixth Int. FLAIRS Conf., pp. 268– 273, 2013.