www.igi-global.com

Application of Machine Learning to User Behavior-Based Authentication in Smartphone and Web

Manoj Jayabalan
Liverpool John Moores University, UK

ABSTRACT

Authentication is the preliminary security mechanism employed in the information system to identify the legitimacy of the user. With technological advancements, hackers with sophisticated techniques easily crack single-factor authentication (username and password). Therefore, organizations started to deploy Multi-Factor Authentication (MFA) to increase the complexity of the access to the system. Despite, the MFA increase the security of the digital service the usable security should be given equal importance. The user behavior-based authentication provides a means to analyze the user interaction with the system in a non-intrusive way to identify the user legitimacy. This chapter presents a review of user behavior-based authentication in smartphones and websites. Moreover, the review highlights some of the common features, techniques, and evaluation criteria usually considered in the development of user behavior profiling.

Keywords: Behavioral Biometric, Multi-Factor Authentication, Authentication, Continuous Authentication, Usable Security, Transparent Authentication, Biometric Authentication, Behavior Profiling, Non-Intrusive Monitoring, Passive Observation, Privilege Misuse, Intruder, Insider Threat, Trust, Privacy, Security

INTRODUCTION

Digital authentication provides a means to secure access to digital information through various technologies. It acts as a prime component in the access control system to mitigate the risk of unauthorized access (Grassi et al., 2017; Jayabalan, 2020). The traditional and most widely used approach to identify the legitimacy of the user consists of supplying a username and password, a system known as Single Factor Authentication. The password is the oldest and predominant authentication factor that exists in the information security world. It is the simplest method to implement and inexpensive, but it is prone to vulnerabilities such as users using weak passwords that are easily cracked, phishing attacks, and other common hacker techniques (Raza et al., 2012). The technological advancements plethora the usage of digital service that requires several authentication factors to be implemented to prevent malicious users. As such, there is a need for organizations to employ Multi-Factor Authentication (MFA) where increased complexity such as using a combination of two or more independent authentication factors (smart cards, biometrics, and security tokens) offers extra security protection (Andrean et al., 2020).

Three-factor authentication using the combination of the above factors can offer greater privacy and security, but as it is more complex, and organizations also have to maintain acceptable efficiency levels, it is a greater challenge to implement. There is an increase in biometric authentication systems in several organizations since these grant access only after validating a subject's unique characteristics (Memon, 2017). Biometric authentication is broadly classified into physiological and behavioral. The physiological biometrics are based on the subject physical properties such as iris, fingerprint, face, and palm. Whereas behavioral biometrics measures the subject unique behavior or patterns from voice, keystroke, mouse dynamics, gait, and system usage, which can uniquely identify an individual (Aupy & Clarke, 2005; Ferbrache, 2016; Meng et al., 2015; Vielhauer, 2006).

The behavioral biometric strike the balance between security and usability via monitoring the user behavior throughout the active session. According to Global Opportunity Report 2017, "*Behavioral biometrics analyses specific human behavior with intelligent software, adding a new layer of security to verifying identification that is nearly impossible to replicate, without any additional stress for the user. Products and services in this market are moving digital security beyond simple passwords and pin codes, ensuring that as cybercriminals become more advanced, so too do everyday users*" (DNV GL AS, 2017).

The advancement of Artificial Intelligence provides a venue for the information security experts to make an informed decision through gaining insights from the historical user access logs. Access logs are an integral part of the system that collects traces of event that was executed by an individual entity. The logs

are beneficial for experts to identify the deviation that has occurred in the process through monitoring and auditing of the operations. Moreover, logs can be effectively utilized in many ways; process mining is the process of extracting the historical log to identify the cause of business process deviation and to improve the business flow (Claes & Poels, 2014; Jayabalan & Thiruchelvam, 2017). It can be further extended to extract user behavior to perform additional authentication by integrating machine learning algorithms.

The purpose of this chapter is to understand the potential inclusion of user behavior profiling in traditional authentication framework. Moreover, the chapter highlights some of the common features, techniques, and evaluation criteria usually considered in the development of user behavior profiling. The scope of this chapter is limited to user behavior-based authentication in smartphones and websites. This chapter is meant to be useful for identifying trends in user behavior profiling that will allow researchers to focus on areas that needs to be improved and new features that could be beneficial to stakeholders.

At the end of this Chapter, you should be able to:

- Understand the functionality and significance of user behavior authentication.
- Identify the factors that are influencing the utilization of user behavior authentication in the digital information service to protect privacy and security.
- Investigate existing and potential approaches with regards to the application of behavior biometric authentication.
- Determine the possible challenges which might occur while introducing the user behavior authentication in digital service.

ISO 29115:2013

The ISO 29115:2013 provides a detailed framework for entity authentication assurance for the overall process in Information and Communications Technology (ISO, 2013). The standard categories the four authentication factors such as “something you know” (e.g., password, PIN), “something you have” (e.g., smart card, device), “something you are” (e.g., biometric characteristic) and “something you do” (e.g., behavior pattern).

ISO 29115:2013 provides guidance to the four Level of Assurance (LOA) from “control technologies, processes, management activities and assurance criteria for mitigating authentication threats.” Each LOA describes the level of confidence in the authentication processes from Level 1 to Level 4 (Low, Medium, High and Very High). The determination of choosing the appropriate LOAs depends on several factors such

as risk, authentication errors, misuse of credentials, the resultant harm/impact and the likelihood of occurrence. The user behavior-based authentication is suitable for LOA 3 and LOA 4. The requirements and implementation guidance of the LOAs are given in Table 1.

Table 1. Requirements and Implementation Guidance of the LOAs

Level	Requirement(s)	Implementation
Level of Assurance 1 (LOA1)	No specific requirement for this level. This level is used when the minimum risk is associated with the data.	- Simple username and password.
Level of Assurance 2 (LOA2)	This level is used when the moderate risk is associated with the data. Necessary steps to be considered for reducing the eavesdropper, online guessing attacks and action on protecting stored credentials.	- Single-factor authentication.
Level of Assurance 3 (LOA3)	This level is used when a substantial risk is associated with the data. No special requirements for the generation of credentials.	- Multi-factor authentication - Cryptography to be applied to the authentication information exchange and rest.
Level of Assurance 4 (LOA4)	This level is used when the high risk is associated with the data. Should follow LOA3 implementation and requirement for in-person identity proofing for human and the storage of cryptographic keys should be secured with the tamper-resistant hardware.	- Multi-factor authentication - Cryptography to be applied to the authentication information exchange and rest. - Digital certificates for all ICT devices.

USER BEHAVIOR PROFILING IN AUTHENTICATION

This section discusses the results obtained from analyzing the existing studies on user behavior profiling based on the application and system usage. Authentication is one of the important factors for any level of digital service that requires validating user legitimacy and ensures user confidentiality. With the gradual surge in the number of security breaches across digital services in diverse industries such as healthcare, banking, military etc., organizations boost their security by using MFA that increases the complexity of the access to the system. The design of usable security should be given equal importance to reduce the hindrance level of users. Usability is one of the key drivers that makes a system good enough to be acceptable to the end-user and other stakeholders (Vasudavan et al., 2016).

Biometric user authentication overcomes the issue of transferability of credentials, in which knowledge and possession of the credential are not belonging naturally to the owner. This means the biometric properties of an individual are distinct from one another and difficult to be transferred to another person. Behavioral biometric authentication considers the properties of an individual pattern captured during the interaction with the information system and use it as a mechanism to identify the legitimacy. Therefore, significant data loss can be avoided through the early detection of unusual behavior. User behavior profiling different from a traditional intrusion detection system in which user behavior is utilized to detect anomalies rather than tracking system or device behavior.

The user behavioral profiling implication is demonstrated in the general Java Authentication and Authorization Service (JAAS) classes that are utilized to securely authenticate the client. It provides a modular framework allowing the applications to remain independent from underlying authentication technologies. Hence, providing a framework to customize based on the organization needs to implement the authentication factors. Figure 1 demonstrates the user behavior profiling in JAAS.

The client-side application acquires user login credentials and environmental conditions as input and sends those parameters to the login module. The web logic server container (for example, RMI, EJB etc.) passes the parameters received from the clients to the web logic server. It sends the parameters to authentication providers to verify the credentials. A meanwhile, the environmental conditions are sent to the decision logic for measuring the similarity of data access and the decision logic decides whether to demand additional authentication based on the organization policy. The patterns are generated by the “behavior profile generator” from the user access log and stored into the “behavior profile data store” for the decision logic to classify the future user behavior.

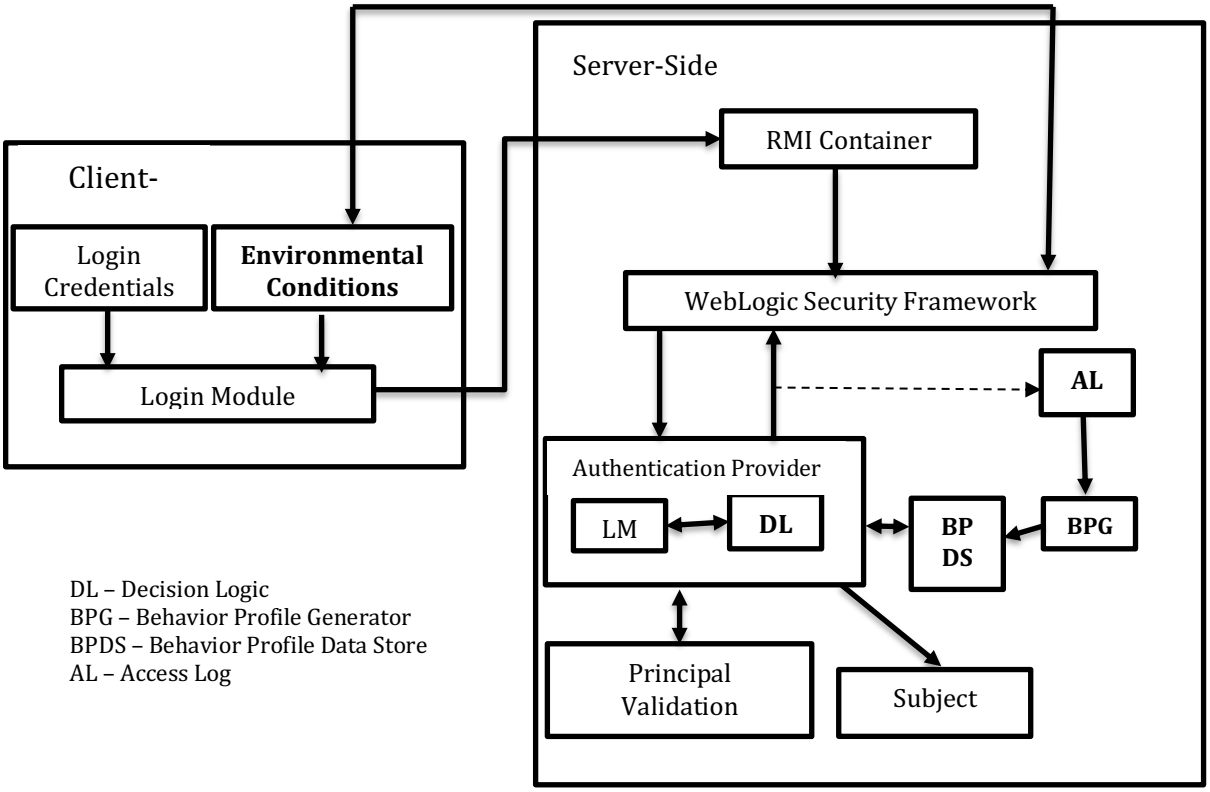

Figure 1: User Behavior Profiling Mapping in JAAS

The below subsection will discuss the most commonly used features in the development of user behavior authentication, machine learning models and evaluation criteria to measure the performance of the model.

Features

The features are the important measures that are required for constructing user behavior profiling, which could identify future user behavior. Text analysis performed on the reviewed articles to identify the most commonly utilized features along with their relationships. Cluster analysis (k = 3) was performed with the extracted keywords to find the Jaccard's coefficient based on the agglomerative order. Figure 2 shows the Dendrogram for the feature, which resulted in two clusters and one single word removed from the cluster.

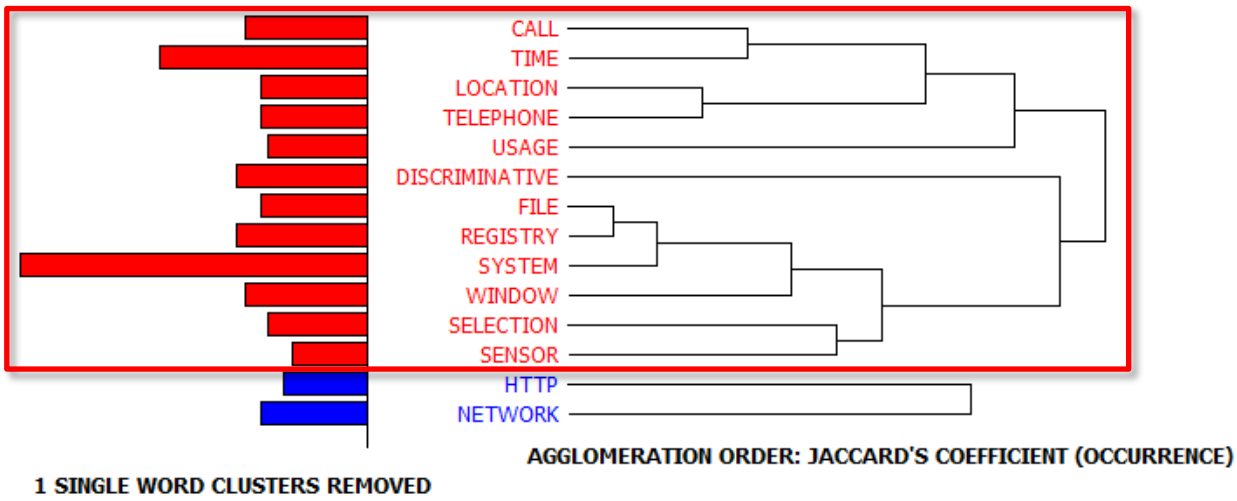

Figure 2. Dendrogram for the keyword Feature

In a typical web-related application, general features such as browsing sequence, time, date, Internet Protocol (IP) address are usually considered in user behavior profiling. The parameter for mobile devices constitutes, the mobile sensor, spatial and the general usage of varied applications are usually considered during the development of profiles. The user call and time are directly linked, and the time parameter has a strong link between the location in which the user accessed (usage) the file and the system.

The user behavior profiling pertaining to the system and application interactions generates an enormous amount of dense data. The significant challenges arise due to an increase in model training time and accuracy in detecting legitimate users. In addition, utilizing the dense data generated from the different sensors and access logs may not produce useful behavior, thus reducing the quality of user profiling. The application of dimensionality reduction techniques over dense data can overcome the issues. Researchers considered the smoothing function to reduce the noise and extract the most usable behavior from the dense data using additive smoothing, moving average which yields better accuracy (Albayram et al., 2013; Li et al., 2014).

User Behavior Learning Methods

User behavior learning is a process of understanding the human interactions with information systems and different means to extract profiles for identifying future user behavior. The dynamic behavior propels the obstructive user authentication in client machines such as mobile devices, desktops, laptops, cloud computing and the Internet of Things (IoT).

The advancement in the mobile device provides a multifaceted approach towards user behavior profiling with the increase in quality of in-built sensors and capabilities to process different applications with ease of access (Ismael et al., 2020). The locking/unlocking of the mobile device provides inconvenience, thus causing the user not to adopt secure authentication. Therefore, one study has shown the possibility of 3-dimensional sensors in the verification as soon as the unlock event action initiated by the user to detect anomaly (Buriro et al., 2017). For the readers to understand the user behavior perspective of locking/unlocking the smartphone, refer to the article (Mahfouz et al., 2016).

Social network usage is increasing at a rapid rate through the use of a smartphone, leading to utilize them in continuous authentication. TrackMasion is a behaviometric analytics platform to monitor social network usage to identify user behavior and utilize it in mobile authentication (Anjomshoa et al., 2016). Further, Radial Basis Function Neural Network applied on the short messaging service to create a linguistic profile that can be used to determine the user behavior and perform continuous authentication (Saevanee et al., 2011).

One study trained the model using n-gram and utilized the perplexity method to predict the abnormal/normal behavior. The spatial and temporal parameters usage in the construction of user behavior generates a significant number of instances. As such, the additive smoothing method was considered to extract the instances for training (Albayram et al., 2013). In another study, the system-level user behavior model proposed to collect data from the registry, file system, as well as general actions performed in the system such as creation and deletion. The features were selected through the fisher method and multivariate Gaussian mixture were utilized to build the model (Yingbo Song et al., 2013).

Cloud computing offers several benefits to an individual and organization without requiring the user to have any knowledge of the infrastructure used by the service providers. Further, virtualization in cloud computing provides an opportunity to increase or decrease IT resources as needed to meet the demands. However, privacy and security are a major concern in storing the organization sensitive information in the third-party server (Kubbo et al., 2016). Thus, several researchers focus on the incorporation of user behavior analysis for anomaly detection and misuse of the service. The user profiling system using Fuzzy and genetic algorithms to monitor the usage pattern and detect suspicious activity in the system (Sahil et al., 2015). Further, research proposed user behavior analysis for the cloud users through analyzing the application usage and multi-algorithmic approach (Adaptive classifier) implemented for each service to achieve better performance (Al-bayati et al., 2016). In addition, to the general discussion on the different learning methods discussed earlier, this section further introduces three classifications of user behavior learning methods based on their applications.

Steering behavior

The user behavior profiling generated based on the predefined set of sequences can be effectively utilized in analyzing the behavior such as web page navigation. One study (Alswiti et al., 2016), proposed a k-NN algorithm for building the classifier based on the user navigation historical data. Another study constructed the user profiles based on the unigram Markov model, which allows to construct of a logical sequence. The utilization of the entire web class leads to a higher false-positive rate, thus only the top k/2 web classes are considered along with the browsing time and classes of web pages (Zhao et al., 2016). However, the researchers do not consider a logical relationship between the web pages.

The user interacting with a web application is considered as a web language through which user actions are modeled as words. The n-gram was utilized to predict user behavior based on past usage patterns (Milton & Memon, 2016). It was performing better in binary classification when compared with multi-classification. Moreover, the performance of the model entirely dependent on the keyword abstraction and even a slight change affects the ability to detect the anomaly and requires high performance computing environment. Most of the web browsing sequences reported the need for a greater number of instances to increase the accuracy of the model (Milton & Memon, 2016; Zhao et al., 2016). Interested readers to understand the process involved in weblog mining can refer to this article (Pabarskaite & Raudys, 2007).

Trust behavior

Trust is an important notion to believe an entity is a legitimate person accessing the system without any malicious intent (Jayabalan, 2020). The trust of a user is calculated using several parameters such as the number of transactions, credibility of feedback, transaction context and community context. The trust vector generated using these criteria are applied with association rule mining to generate user behavior. The obtained patterns are applied with a Bayesian classifier to determine whether the given user access is trustworthy or untrustworthy (D'Angelo et al., 2016). Similarly, (Brosso et al., 2010) proposed continuous authentication through analyzing the user behavior, which is computed using the measure of confidence on the various environmental factors and scores are evaluated using the Neuro-Fuzzy to determine the trust level.

In (Kent & Liebrock, 2013), proposed user authentication for a large-scale enterprise to model the behavior using graphs and the characteristics of the graphs are utilized to build the logistic regression model. The

concept of graphs only benefited in providing basic insights into potential credential mixing risks within the network. Another study proposed an adaptive authentication for Malaysia government e-service, which combines multiple applications with single sign-on capabilities (Bakar & Haron, 2014). The user behavior profiling is generated based on the frequency of attribute values and the approach does not find the correlation between them. It lacks predictive capability, high variance in certain attributes, and does not adapt to the most recent changes.

One study proposed the use of an “Interactive Dichotomiser 3” algorithm to characterize the behavior of the user authentication and utilizes the Random Petri network model to analyze the credibility (Lu & Xu, 2014). The credibility degree is computed on normalized user behavior and assigned different levels of trust score to access the data. However, the authenticated users are allowed to directly access the resources based on the roles and user behavior is analyzed at a later stage. This approach needs to compromise on a certain amount of data loss before the anomaly is being identified. There are additional problems in characterizing the user behavior, for instance, only the attributes with the highest entropy are selected and the remaining attributes are not utilized when the instances are correctly classified with fewer attributes, thus, leading to an overfitting problem. Secondly, the approach does not perform better when there are limited instances.

The mobile phone operates with limited resources leads to the computational overhead of processing the trust score within the device considering the entire user behavior activity of different applications with spatial and temporal parameters. Hence, the cloud platform provided an efficient infrastructure to process the trust score using the probability to determine the legitimacy of a user (Chow et al., 2010).

Trial Behavior

Generating a challenge question based on the previous transactions dating back over a decade (B. & Venkataram, 2007). With the recent era of Big Data and its technologies, the possibilities to generate user behavior with the huge volume of data from different sources leads to the prospects of constructing challenge questions in authentication framework (Ibrahim & Ouda, 2016). The knowledge of historical user transactions is the key factor to identify the legitimacy of the user based on the challenge questions.

The questions are usually generated based on the predefined features mapped with the user transactions to measure the frequency of actionable items (Skračić et al., 2017). Notwithstanding, mobile misuse is a major challenge and the researchers' utilized mobile application usage for building the user profiles. The rule-based classifier is used for determining the probability of the event and neural networks for analyzing the

call history (Li et al., 2014).

The recommender system analyzes user needs and preferences by finding the correlation between the user, items, rating or reviews. The recommender system has been implemented to identify the top selling items, products, customer demographics, past buying behavior, search history and can also consider social connections of the specific user (Katarya and Verma, 2016; Rana and Jain, 2012; Tarus et al., 2017). Further, the researchers have shown the possibilities of generating user behavior profiles and dynamic challenge questions based on past transactions with the help of collaborative filtering (Ibrahim & Ouda, 2017).

Evaluation Criteria

The models are built on the annotated data should generalize well on future unseen data (Raykar & Saha, 2015). A decent estimate of the model performance is an important characteristic that usually computed through measuring accuracy in order to detect the future predicted behavior. The performance evaluation metrics are broadly classified into the threshold, probability, and ranking metrics. These metrics are the scalar group method that presents the classifier performance in a single score value, thus making it easier to compare and contrast the results with other metrics. In most cases, these types of metrics are employed in three different evaluation applications (Hossin & Sulaiman, 2015).

- **Generalization:** In this evaluation, the metrics were used to measure the generalizability and quality of the summary on the trained classifier. The common metrics utilized for this evaluation consist of accuracy and error.
- **Model Selection:** The best classifier among the different trained classifiers are selected based on the performance of the test set.
- **Discriminator:** The evaluation metrics are employed to discriminate and select the optimum classifier during the validation.

In order to measure the performance of generalization and model selection, all the three discussed evaluation metrics (threshold, probability and ranking) can be employed to measure the effectiveness. However, only certain types of metrics from the three categories utilized for discriminating the classifier such as A Receiver Operating Characteristic Curve (ROC), confusion matrix etc. (Caruana & Niculescu-Mizil, 2004; Han et al., 2012; Marcot, 2012). The commonly used evaluation methods for user behavior

profiling are listed below (Pisani et al., 2016).

- False Acceptance Rate (FAR) measures how often a classifier falsely identifies an impostor as a genuine user by calculating false matches over total impostor match attempts.
- False Rejection Rate (FRR) measures how often a classifier falsely identifies a genuine user as an impostor by calculating false rejection over total genuine match attempts.
- Equal Error Rate (EER) measures the threshold point between FAR and FRR.
- Accuracy rate measures correct classification obtained by the classifier in percentage; and
- Integrated error measures the portion of the area resulted by plotting FAR and FRR together.

DISCUSSION AND FUTURE DIRECTION

This section presents the discussion and future directions of user behavior-based authentication. The behavioral biometrics authentication uniquely identifies legitimate users from the adversary based on the behavioral trail. The concept of behavioral biometric dates back to over a century and was even utilized in World War II to uniquely identify the telegraph operators based on the keystroke dynamics. The approach was termed as "Fist of the Sender" to uniquely identify and validate the sender message by analyzing the typing rhythm, pace, and syncopation of the telegraph keys (Banerjee & Woodard, 2012). Behavioral biometrics such as keystrokes and mouse dynamics, which are usually captured under static and controlled conditions. These approaches are vulnerable to replay attacks, human interaction simulation and advanced malware injections. However, the behavioral biometrics are trained as the user operates the system which is difficult to be mimic by the robots due to the invisible challenge and improve security with a cognitive fingerprint of the user (Ferbrache, 2016; Turgeman & Zelazny, 2017).

The researchers' major perseverance to adopt the user behavior analytics in authentication is to detect insider threats, prevent misuse and usable security. The system level attacks are well planned, and several security tools are utilized to monitor and prevent external threats to organization wide networks. Nevertheless, the insider threat and misuse are a major concern to the organizations where co-workers or imposters steal credentials and access the sensitive information, which able to be detected through user behavior profiling (Al-bayati et al., 2016; Li et al., 2014; Yingbo Song et al., 2013). The 2017 Verizon Data Breach Investigations Report says, "Insider misuse is a major issue for the Healthcare industry; in fact, it is the only industry where employees are the predominant threat actors in breaches*"*. Just over half of the incidents with confirmed healthcare data disclosure analyzed were due to privilege misuse and misdelivery

(Verizon, 2017).

With the Health Insurance Portability and Accountability Act (HIPAA) and ISO22600-1:2014 requiring healthcare organizations to boost security by using MFA that increases the complexity of the access to the system, the design of usable security should be given equal importance (ISO, 2014; Jayabalan & O'Daniel, 2016; Tipton et al., 2016). As such, healthcare practitioner behavior and the nature of their interaction with security features should be considered as an important characteristic at the design stage (Jayabalan & O'Daniel, 2019; Realpe-Munoz et al., 2016). In studies conducted to identify usability issues in electronic health record authentication, the major concerns among healthcare practitioners were revealed to be efficiency and availability (Ferreira et al., 2011; Wang & Jin, 2008). It was further noted that practitioner acceptance and attitude depend on electronic health records usability (Kaipio et al., 2017).

According to Gartner, "Affiliated physicians are not employees of the healthcare delivery organization but have an elective relationship. Obliging the affiliated physician to use an OTP hardware token may sour and even curtail that relationship. Adopting contextual/analytic and adaptive capabilities can minimize the burden of higher-trust authentication on physicians by limiting its use to only those instances where the level of risk demands it" (Mahdi et al., 2016).

A semi-structured interview for the physiological needs for privacy and security in smartphones resulted in a low response from the participants (Kraus et al., 2017). Since the individual expectations are beyond the need for general authentication. However, this might not be the case for an organization to adapt user behavior profiling. As the behavioral patterns constructed based on the application usage in a continuous manner (intrusive monitoring) to ensure the verification process is carried out in a user-friendly way without any additional efforts from the user.

The trail behavior discussed in the previous section focused on generating the challenge questions based on the historical transactions. It might be suitable for industries such as social networks, e-commerce, banking and finance. However, the information security experts should consider the users' age as an important factor before deciding to adopt this variant. Because older people face age-related impairments which might affect their ability to recall their historical transactions (Vasudavan et al., 2016). The trust behavior variant focuses on calculating the risk associated with user authenticity and applies a mathematical formula to compute a trust score or rank. This method of authentication is more suitable in different areas such as handheld devices, IoT, and dynamic industries. Cloud computing, National Security and Intelligence, military, and healthcare works in a unique operating environment, and high impact of threats that requires additional mechanisms to protect privacy and security (Jayabalan, 2020). For instance, the cloud service provider

offers the organization to manage their service which requires dynamic threat assessment (Ehsan Rana et al., 2017). Thus, user behavior profiling through its implementation can assess the user risk and trust using the vulnerability of the current environment, threats and integrity of user with the historical user behavior.

People tend to exhibit certain uniqueness in the level of interaction to the system which can change gradually over the course of time, thus pattern aging is one of the root causes to influence false positives or error rates (Clarke, 2011). Accuracy can be improved by dynamically adopting the most recent changes in user behavior. However, renewing the template might include the illegitimate usage which an imposter might be accepted by the system over time as the genuine user (Al-bayati et al., 2016). One article considers this issue and addressed using the change point detection with the fixed sliding window for the number of instances using time series (Al Solami et al., 2010). Further studies required in identifying illegitimate usage while renewing the template.

Another major challenge to information security experts in user behavior-based authentication is to overcome the cold start problem for the new users, which is not addressed in the existing studies. The new users without having any access trails will most likely not be selected for continuous authentication, which is referred to as “cold start”. However, it can be overcome by using the general access templates for individual role-based profiles.

The general hypothesis in authentication factors is “a successfully authenticated subject is a truthful owner accessing the information”, thus naïve to authorization mechanism allowing an intruder to take for granted. Further research can consider the access policies (XML, Web Ontology Language) that represent the semantic meaning of every object and its relationships based on the user roles to monitor along with the user behavior (Jayabalan & Oadaniel, 2018). Thus, a combination of authentication mechanisms can be tailored based on the consumption of different sensitive data. For instance, fingerprint authentication is required to access highly sensitive data from certain locations and single factor authentication (username and password) is sufficient to access highly sensitive data from the trusted region and device.

There are two perspectives of privacy risk in user behavior profiling, first order and second order; the leakage of single information is known as first order privacy risk. The second order privacy risk arises due to the user profiling and data mining techniques that are applied to individual data access (Bal et al., 2015). Hence, access confinement and distorting data are methods used to protect sensitive data. At the user data profiling depository phase, encryption techniques such as Identity-Based Encryption and Attribute-Based Encryption are well-known apart to protect while data stored in the cloud vendor or server. The privacy-preserving techniques are mainly acquired in the data processing step of big data analytics. Data

anonymization, also known as data masking or data desensitization, is used to obfuscate or conceal any sensitive data about an individual, thus limiting the person's re-identification (Rajendran et al., 2017). Further research needed in virtue of overcoming second order privacy risks through the application of cryptographic and privacy preserving techniques.

CONCLUSION

Authentication is a fundamental security mechanism to protect user privacy and security in digital services. There are several methods proposed in the existing studies to secure data with multifactor authentication and usability is always a concern. Transparent and continuous authentication provide a better tradeoff between security and usability. Employing user behavior-based authentication to the existing multi-factor authentication framework will provide additional security to the system without user intervention. There is a need for continuous authentication to be performed in the industries managing sensitive data through analyzing the user behavior towards their digital services to detect the potential threats.

User behavior-based profiles are created based on the pertinent information from the historical access log. The confidence levels are computed based on the similarity between the real-time factors with the existing patterns to determine the legitimacy of the user. The user behavior-based authentication was demonstrated using the Java Authentication and Authorization Services for the information security experts and developers to understand the implementation details. Further, the taxonomy of the user behavior learning methods was introduced in this chapter such as trail behavior, trust behavior, and steering behavior. The application of machine learning and natural language processing was dominant in trail behavior and steering behavior. Whereas trust behavior is an amalgamation of the aforementioned techniques with probability and statistics. This chapter also presented the most common issues to be dealt with whilst the organizations adopt user behavior-based authentication to protect privacy and security. Moreover, the chapter highlighted some of the research gaps with a lack of empirical studies.

ACKNOWLEDGMENT

I would like to thank Dr. Thomas O'Daniel from Asia Pacific University of Technology and Innovation, Malaysia for sharing his valuable suggestions during the research.

REFERENCES

Al Solami, E., Boyd, C., Clark, A., & Islam, A. K. (2010). Continuous Biometric Authentication: Can It Be More Practical? *2010 IEEE 12th International Conference on High Performance Computing and Communications (HPCC)*, 647–652. https://doi.org/10.1109/HPCC.2010.65

Al-bayati, B., Clarke, N., & Dowland, P. (2016). Adaptive Behaviroal Profiling for Identity Verification in Cloud Computing: A Model and Preliminary Analysis. *GSTF International Journal on Computing (JoC Vol.3 No.2)*, *5*(1), 21–28. https://doi.org/10.5176/2251-3043

Albayram, Y., Kentros, S., Ruhua Jiang, & Bamis, A. (2013). A method for improving mobile authentication using human spatio-temporal behavior. *2013 IEEE Symposium on Computers and Communications (ISCC)*, 000305–000311. https://doi.org/10.1109/ISCC.2013.6754964

Alswiti, W., Alqatawna, J., Al-Shboul, B., Faris, H., & Hakh, H. (2016). Users Profiling Using Clickstream Data Analysis and Classification. *2016 Cybersecurity and Cyberforensics Conference (CCC)*, 96–99. https://doi.org/10.1109/CCC.2016.27

Andrean, A., Jayabalan, M., & Thiruchelvam, V. (2020). Keystroke Dynamics Based User Authentication using Deep Multilayer Perceptron. *International Journal of Machine Learning and Computing*, *10*(1), 134–139. https://doi.org/10.18178/ijmlc.2020.10.1.910

Anjomshoa, F., Catalfamo, M., Hecker, D., Helgeland, N., Rasch, A., Kantarci, B., Erol-Kantarci, M., & Schuckers, S. (2016). Mobile behaviometric framework for sociability assessment and identification of smartphone users. *2016 IEEE Symposium on Computers and Communication (ISCC)*, *2016-Augus*, 1084–1089. https://doi.org/10.1109/ISCC.2016.7543880

Aupy, A., & Clarke, N. (2005). User Authentication by Service Utilisation Profiling. *Advances in Network and Communications Engineering 2*, *2*, 18.

B., S. B., & Venkataram, P. (2007). An Authentication Scheme for Personalized Mobile Multimedia Services: A Cognitive Agents Based Approach. *Future Generation Communication and Networking (FGCN 2007)*, 167–172. https://doi.org/10.1109/FGCN.2007.57

Bakar, K. A. A., & Haron, G. R. (2014). Adaptive authentication based on analysis of user behavior. *2014 Science and Information Conference*, 601–606. https://doi.org/10.1109/SAI.2014.6918248

Bal, G., Rannenberg, K., & Hong, J. I. (2015). Styx: Privacy risk communication for the Android smartphone platform based on apps' data-access behavior patterns. *Computers and Security*, *53*(69), 187–202. https://doi.org/10.1016/j.cose.2015.04.004

Banerjee, S. P., & Woodard, D. L. (2012). *Biometric Authentication and Identification using Keystroke Dynamics : A Survey*. *7*, 116–139. https://doi.org/10.13176/11.427

Brosso, I., La Neve, A., Bressan, G., & Ruggiero, W. V. (2010). A Continuous Authentication System Based on User Behavior Analysis. *International Conference on Availability, Reliability, and Security, 2010. ARES '10*, 380–385. https://doi.org/10.1109/ARES.2010.63

Buriro, A., Crispo, B., & Zhauniarovich, Y. (2017). Please hold on: Unobtrusive user authentication using smartphone's built-in sensors. *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, *February*, 1–8. https://doi.org/10.1109/ISBA.2017.7947684

Caruana, R., & Niculescu-Mizil, A. (2004). Data mining in metric space: an empirical analysis of supervised learning performance criteria. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 69–78. https://doi.org/1-58113-888-1/04/0008

Chow, R., Jakobsson, M., Masuoka, R., Molina, J., Niu, Y., Shi, E., & Song, Z. (2010). Authentication in the clouds. *Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop - CCSW '10*, 1. https://doi.org/10.1145/1866835.1866837

Claes, J., & Poels, G. (2014). Merging event logs for process mining: A rule based merging method and rule suggestion algorithm. *Expert Systems with Applications*, *41*(16), 7291–7306. https://doi.org/10.1016/j.eswa.2014.06.012

Clarke, N. (2011). Transparent User Authentication. In *Transparent User Authentication*. Springer London. https://doi.org/10.1007/978-0-85729-805-8

D'Angelo, G., Rampone, S., & Palmieri, F. (2016). Developing a trust model for pervasive computing based on Apriori association rules learning and Bayesian classification. *Soft Computing*. https://doi.org/10.1007/s00500-016-2183-1

DNV GL AS. (2017). *Global Opportunity Opportunity*.

Ehsan Rana, M., Kubbo, M., & Jayabalan, M. (2017). Privacy and Security Challenges Towards Cloud Based Access Control in Electronic Health Records. *Asian Journal of Information Technology*, *16*(2), 274–281. https://doi.org/10.36478/ajit.2017.274.281

Ferbrache, D. (2016). Passwords are broken – the future shape of biometrics. *Biometric Technology Today*, *2016*(3), 5–7. https://doi.org/10.1016/S0969-4765(16)30049-2

Ferreira, A., Cruz-Correia, R., & Antunes, L. (2011). Usability of authentication and access control: A case study in healthcare. *Proceedings - International Carnahan Conference on Security Technology*, 1–7. https://doi.org/10.1109/CCST.2011.6095873

Grassi, P. A., Garcia, M. E., & Fenton, J. L. (2017). *Digital identity guidelines: revision 3*. https://doi.org/10.6028/NIST.SP.800-63-3

Han, J., Kamber, M., & Pei, J. (2012). *Data mining : Concepts and Techniques* (3rd ed.). Elsevier.

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, *5*(2), 1–11. https://doi.org/10.5121/ijdkp.2015.5201

Ibrahim, A., & Ouda, A. (2017). A hybrid-based filtering approach for user authentication. *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–5. https://doi.org/10.1109/CCECE.2017.7946830

Ibrahim, A., & Ouda, A. (2016). Innovative Data Authentication Model. *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 1–7. https://doi.org/10.1109/IEMCON.2016.7746268

Ismael, A. A., Jayabalan, M., & Al-Jumeily, D. (2020). A study on human activity recognition using smartphone. *Journal of Advanced Research in Dynamical and Control Systems*, *12*(5 Special Issue), 795–803. https://doi.org/10.5373/JARDCS/V12SP5/20201818

ISO. (2013). *BS ISO/IEC 29115:2013: Information technology. Security techniques. Entity authentication assurance framework*.

ISO. (2014). BS EN ISO 22600-1:2014: Health informatics. Privilege management and access control. Overview and policy management. In *British Standards Institute.*

Jayabalan, M. (2020). Towards an Approach of Risk Analysis in Access Control. *2020 13th International Conference on Developments in ESystems Engineering (DeSE)*, 287–292. https://doi.org/10.1109/DeSE51703.2020.9450772

Jayabalan, M., & Oadaniel, T. (2018). Continuous and transparent access control framework for electronic health records: A preliminary study. *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, *2018-Janua*, 165–170. https://doi.org/10.1109/ICITISEE.2017.8285487

Jayabalan, M., & O'Daniel, T. (2016). Access control and privilege management in electronic health record: a systematic literature review. *Journal of Medical Systems*, *40*(12), 261. https://doi.org/10.1007/s10916-016-0589-z

Jayabalan, M., & O'Daniel, T. (2019). A study on authentication factors in electronic health records. *Journal of Applied Technology and Innovation*, *3*(1), 7–14. https://jati.apu.edu.my/

Jayabalan, M., & Thiruchelvam, V. (2017). A design of patients data transparency in electronic health records. *2017 IEEE International Symposium on Consumer Electronics (ISCE)*, 9–10. https://doi.org/10.1109/ISCE.2017.8355532

Kaipio, J., Lääveri, T., Hyppönen, H., Vainiomäki, S., Reponen, J., Kushniruk, A., Borycki, E., & Vänskä, J. (2017). Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. *International Journal of Medical Informatics*, *97*, 266–281. https://doi.org/10.1016/j.ijmedinf.2016.10.010

Kent, A. D., & Liebrock, L. M. (2013). Differentiating User Authentication Graphs. *2013 IEEE Security and Privacy Workshops*, 72–75. https://doi.org/10.1109/SPW.2013.38

Kraus, L., Wechsung, I., & Möller, S. (2017). Psychological needs as motivators for security and privacy actions on smartphones. *Journal of Information Security and Applications*, *34, Part 1*, 34–45. https://doi.org/https://doi.org/10.1016/j.jisa.2016.10.002

Kubbo, M., Jayabalan, M., & Rana, M. E. (2016). Privacy and Security Challenges in Cloud Based Electronic Health Record : Towards Access Control Model. *Third International Conference on Digital Security and Forensics (DigitalSec)*, 113–121.

Li, F., Clarke, N., Papadaki, M., & Dowland, P. (2014). Active authentication for mobile devices utilising behaviour profiling. *International Journal of Information Security*, *13*(3), 229–244. https://doi.org/10.1007/s10207-013-0209-6

Lu, X., & Xu, Y. (2014). An User Behavior Credibility Authentication Model in Cloud Computing Environment. *2nd International Conference on Information Technology and Electronic Commerce.*, 271–275. https://doi.org/10.1109/ICITEC.2014.7105617

Mahdi, D. A., Ant, A., & Singh, A. (2016). Market Guide for User Authentication. *Gartner Reprint*, *November*, 1–15.

Mahfouz, A., Muslukhov, I., & Beznosov, K. (2016). Android users in the wild: Their authentication and usage behavior. *Pervasive and Mobile Computing*, *32*, 50–61. https://doi.org/10.1016/j.pmcj.2016.06.017

Marcot, B. G. (2012). Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling*, *230*, 50–62. https://doi.org/10.1016/j.ecolmodel.2012.01.013

Memon, N. (2017). How Biometric Authentication Poses New Challenges to Our Security and Privacy [In the Spotlight]. *IEEE Signal Processing Magazine*, *34*(4), 196–194. https://doi.org/10.1109/MSP.2017.2697179

Meng, W., Wong, D. S., Furnell, S., & Zhou, J. (2015). Surveying the development of biometric user authentication on mobile phones. *IEEE Communications Surveys and Tutorials*, *17*(3), 1268–1293. https://doi.org/10.1109/COMST.2014.2386915

Milton, L. C., & Memon, A. (2016). Intruder detector: A continuous authentication tool to model user behavior. *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, 286–291. https://doi.org/10.1109/ISI.2016.7745492

Pabarskaite, Z., & Raudys, A. (2007). A process of knowledge discovery from web log data: Systematization and critical review. *Journal of Intelligent Information Systems*, *28*(1), 79–104. https://doi.org/10.1007/s10844-006-0004-1

Pisani, P. H., Giot, R., de Carvalho, A. C. P. L. F., & Lorena, A. C. (2016). Enhanced template update: Application to keystroke dynamics. *Computers & Security*, *60*, 134–153. https://doi.org/10.1016/j.cose.2016.04.004

Rajendran, K., Jayabalan, M., & Ehsan Rana, M. (2017). A Study on k-anonymity, l-diversity, and t-closeness Techniques focusing Medical Data. *IJCSNS International Journal of Computer Science and Network Security*, *17*(12), 172–177. http://paper.ijcsns.org/07_book/201712/20171225.pdf

Raykar, V. C., & Saha, A. (2015). Data Split Strategiesfor Evolving Predictive Models. In A. Appice, P. P. Rodrigues, V. Santos Costa, C. Soares, J. Gama, & A. Jorge (Eds.), *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I* (pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-23528-8_1

Raza, M., Iqbal, M., Sharif, M., & Haider, W. (2012). A survey of password attacks and comparative analysis on methods for secure authentication. *World Applied Sciences Journal*, *19*(4), 439–444. https://doi.org/10.5829/idosi.wasj.2012.19.04.1837

Realpe-Munoz, P., Collazos, C. . A., Hurtado, J., Granollers, T., & Velasco_Medina, J. (2016). An Integration of Usable Security and User Authentication into the ISO 9241-210 and ISO/IEC 25010:2011. In T. Tryfonas (Ed.), *Human Aspects of Information Security, Privacy, and Trust* (pp. 65–75). Springer International Publishing Switzerland 2016. https://doi.org/10.1007/978-3-319-39381-0

Saevanee, H., Clarke, N., & Furnell, S. (2011). *SMS Linguistic Profiling Authentication on Mobile Device*. 224–228.

Sahil, Sood, S., Mehmi, S., & Dogra, S. (2015). Artificial intelligence for designing user profiling system for cloud computing security: Experiment. *2015 International Conference on Advances in Computer Engineering and Applications*, 51–58. https://doi.org/10.1109/ICACEA.2015.7164645

Skračić, K., Pale, P., & Kostanjčar, Z. (2017). Authentication approach using one-time challenge generation based on user behavior patterns captured in transactional data sets. *Computers and Security*, *67*, 107–121. https://doi.org/10.1016/j.cose.2017.03.002

Tipton, S. J., Forkey, S., & Choi, Y. B. (2016). Toward Proper Authentication Methods in Electronic Medical Record Access Compliant to HIPAA and C.I.A. Triangle. *Journal of Medical Systems*, *40*(4), 1–8. https://doi.org/10.1007/s10916-016-0465-x

Turgeman, A., & Zelazny, F. (2017). Invisible challenges: the next step in behavioural biometrics? *Biometric Technology Today*, *2017*(6), 5–7. https://doi.org/10.1016/S0969-4765(17)30114-5

Vasudavan, H., Jayabalan, M., & Ramiah, S. (2016). A preliminary study on designing tour website for older people. *2015 IEEE Student Conference on Research and Development, SCOReD 2015*. https://doi.org/10.1109/SCORED.2015.7449423

Verizon. (2017). 2017 Data Breach Investigations Report Tips on Getting the Most from This Report. *Verizon Business Journal*, *1*, 1–48. https://doi.org/10.1017/CBO9781107415324.004

Vielhauer, C. (2006). *Biometric User Authentication for IT Security* (Vol. 18, Issue 0). Springer-Verlag. https://doi.org/10.1007/0-387-28094-4

Wang, Q., & Jin, H. (2008). Usable Authentication for Electronic Healthcare Systems. *Proceedings of the Symposium On Usable Privacy and Security (SOUPS)*.

Yingbo Song, Ben Salem, M., Hershkop, S., & Stolfo, S. J. (2013). System Level User Behavior Biometrics using Fisher Features and Gaussian Mixture Models. *2013 IEEE Security and Privacy Workshops*, 52–59. https://doi.org/10.1109/SPW.2013.33

Zhao, P., Yan, C., & Jiang, C. (2016). Authenticating Web User's Identity through Browsing Sequences Modeling. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 335–342. https://doi.org/10.1109/ICDMW.2016.0054

ADDITIONAL READING

Andrean, A., Jayabalan, M., & Thiruchelvam, V. (2020). Keystroke Dynamics Based User Authentication using Deep Multilayer Perceptron. *International Journal of Machine Learning and Computing*, *10*(1), 134–139. https://doi.org/10.18178/ijmlc.2020.10.1.910

Clarke, N. (2011). Transparent User Authentication. Springer London. https://doi.org/10.1007/978-0-85729-805-8

Ferbrache, D. (2016). Passwords are broken – the future shape of biometrics. *Biometric Technology Today*, *2016*(3), 5–7. https://doi.org/10.1016/S0969-4765(16)30049-2

Halunen, K., Häikiö, J., & Vallivaara, V. (2017). Evaluation of user authentication methods in the gadget-free world. *Pervasive and Mobile Computing*, *40*, 220–241. https://doi.org/10.1016/j.pmcj.2017.06.017

Jain, A. K., Nandakumar, K., & Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, *79*, 80–105. https://doi.org/10.1016/j.patrec.2015.12.013

Jayabalan, M., & O'Daniel, T. (2016). Access control and privilege management in electronic health record: a systematic literature review. *Journal of Medical Systems*, *40*(12), 261. https://doi.org/10.1007/s10916-016-0589-z

Jayabalan, M., & O'Daniel, T. (2019). A study on authentication factors in electronic health records. *Journal of Applied Technology and Innovation*, *3*(1), 7–14. https://jati.apu.edu.my/

Jayabalan, M., & O'Daniel, T. (2017). Continuous and transparent access control framework for electronic health records: A preliminary study. *2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 165–170. https://doi.org/10.1109/ICITISEE.2017.8285487

Kraus, L., Wechsung, I., & Möller, S. (2017). Psychological needs as motivators for security and privacy actions on smartphones. *Journal of Information Security and Applications*, *34, Part 1*, 34–45. https://doi.org/https://doi.org/10.1016/j.jisa.2016.10.002

Ring, T., & Wilk, R. (2015). Behavioural analytics: Sifting good users from bad actors. *Biometric Technology Today*, *2015*(11), 8–11. https://doi.org/10.1016/S0969-4765(15)30173-9

KEY TERMS AND DEFINITIONS

Access Policy: A list of roles and resources to which the access permissions are defined for an individual role.

Cloud Computing: On demand availability of computing power and data storage capacity.

Continuous Authentication: A verification method aimed to provide identity confirmation and cybersecurity protection on an ongoing basis.

Intruder Detection: A software application or device to monitor the organization network for unusual activity.

Keystroke: The pressing of a single key on a keyword.

Mouse Dynamics: A tiny patterns and variation in the mouse and/or pointer movements while the user interacts with the screen.

Transparent Authentication: A verification method aimed to assess the user behavior in a non-intrusive way to identify the legitimacy.

Usable Security: A process to ensure the security products and services are usable by those who need them.