User Behaviour Analysis Based on Similarity Measures to Detect Anomalies

Arnoldas Budžys, Viktor Medvedev, Olga Kurasova

Vilnius University, Institute of Data Science and Digital Technologies

arnoldas.budzys@mif.stud.vu.lt

Introduction

"Imposter" threats is responsible for 60% of cyber attacks worldwide and are extremely difficult to detect. Whether the imposter is a malicious employee or a contractor whose credentials have been compromised, security teams need to be able quickly and accurately detect, investigate and respond to these potentially attacks.

If user's credentials falls into the hands of malicious parties, all personal, financial or commercial data could be compromised. In order to solve this problem, full attention is being paid to the dynamics of keystrokes (see Figure 1). **The aim of this research** is to detect anomalous behaviour or imposters by dissociating the keystroke behaviour dynamics of a legitimate user from an illegitimate user. When the authenticity of a user is questioned against an already established profile, the system may terminate the session and revert to SA in order to continue working [1].

Summary of the Experiments

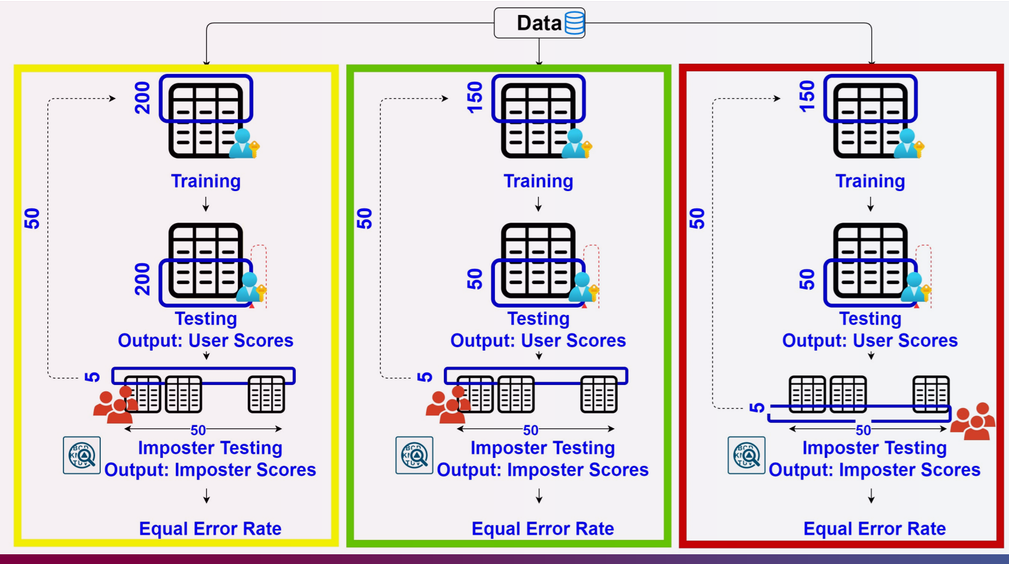

Figure 3. The summary of the experiments

Equal error rate (EER) measure solves the problem of selecting a threshold value partially, and it represents the failure rate when the values of false acceptance rate (FAR) and false reject rate (FRR) are equal [4] (see Figure 4).

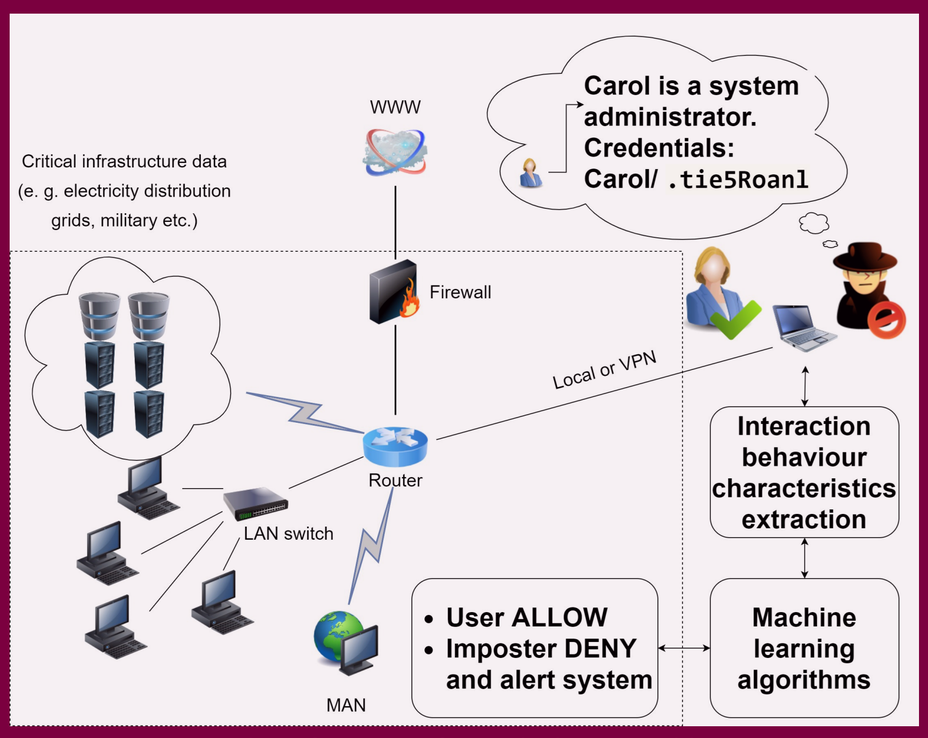

Figure 1 . Identification of user behavioural biometrics

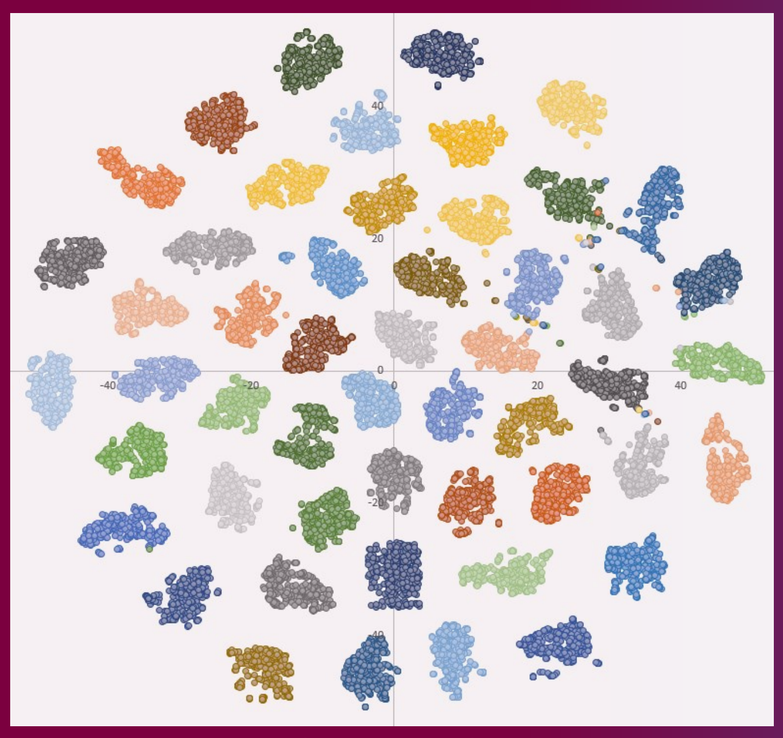

Figure 2. Visualizing users data using t-SNE

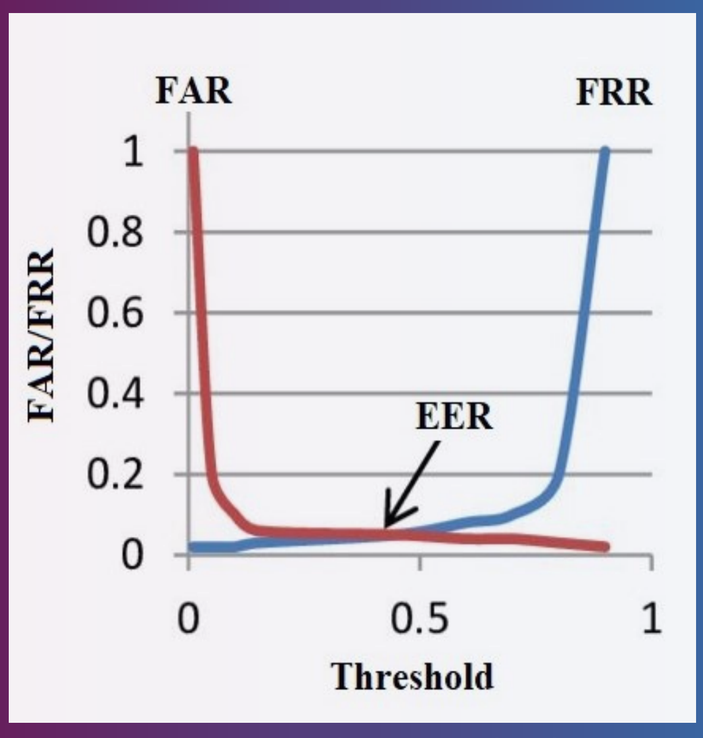

Figure 4. Equal error rate (EER)

Datasets

Users keyboard's typing datasets [2]: GREYC keystroke Benchmark and GREYC12 Static Keystroke Dynamics Benchmark Datasets, Si6, Rhu Keystroke Dataset, Clarkson University Dataset, CMU Dataset, The Sapientia University Dataset.

The CMU dataset [3] was used for this experiment. The data consists of 51 users (see Figure 2) from the Carnegie-Mellon University community. All users/subjects entered the same password, and each subject typed the password 400 times over 8 sessions (50 repetitions per session). The password (.tie5Roanl) was chosen to match a strong 10-character password.

Results

Methods	Average-EER		
Nearest Neighbour (Mahalanobis)	0.3795	0.4548	0.5013
Euclidean	0.1693	0.1863	0.2346
Manhattan	0.1503	0.1622	0.2032
Manhattan (Scaled)	0.0945	0.0986	0.1291
Manhattan (Filtered)	0.1253	0.1399	0.1886
Mahalanobis	0.1596	0.1987	0.2338
Euclidean (Normed)	0.2107	0.2308	0.2483
Mahalanobis (Normed)	0.1996	0.2686	0.3083
Outlier (Counting)	0.1031	0.1060	0.1687
k Means	0.1533	0.1721	0.2238
SVM	0.1205	0.1077	0.1478

Methods

Methods for anomaly detection [3] identifying abnormal user behaviour used in experiments: Nearest Neighbour (Mahalanobis), Manhattan (Scaled), Outlier (Counting) Mahalanobis, Mahalanobis (Normed), Manhattan (Filtered), Manhattan, Euclidean, Euclidean (Normed), Support Vector Machine (SVM), k Means. The summary of the experiments carried out is presented in Figure 3.

Generalisation

The experimental results reveal which methods have the lowest error rate when identifying unusual user behaviour. Preliminary results show which methods have the lowest error rate in identifying possible anomalous user behaviour. The identified methods will be used in further research using other datasets. The initial set-up of the training and testing data sets, as well as the imposter data sets, increases the threshold of equal error rate.

References: [1] Krishnamoorthy, S., Rueda, L., Saad, S., Elmiligi, H. (2018). Identification of user behavioral biometrics for authentication using keystroke dynamics and machine learning. In *Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications* (pp. 50-57). **[2]** Raul, N., Shankarmani, R., & Joshi, P. (2020). A comprehensive review of keystroke dynamics-based authentication mechanism. In *International Conference on Innovative Computing and Communications* (pp. 149-162). **[3]** Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. In *2009 IEEE/IFIP International Conference on Dependable Systems & Networks* (pp. 125-134). IEEE. **[4]** Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.