DEEP LEARNING-BASED AUTHENTICATION FOR INSIDER THREAT DETECTION IN CRITICAL INFRASTRUCTURE

Arnoldas Budžys, Olga Kurasova, Viktor Medvedev
Vilnius University, Institute of Data Science and Digital Technologies
arnoldas.budzys@mif.stud.vu.lt

ID I-7

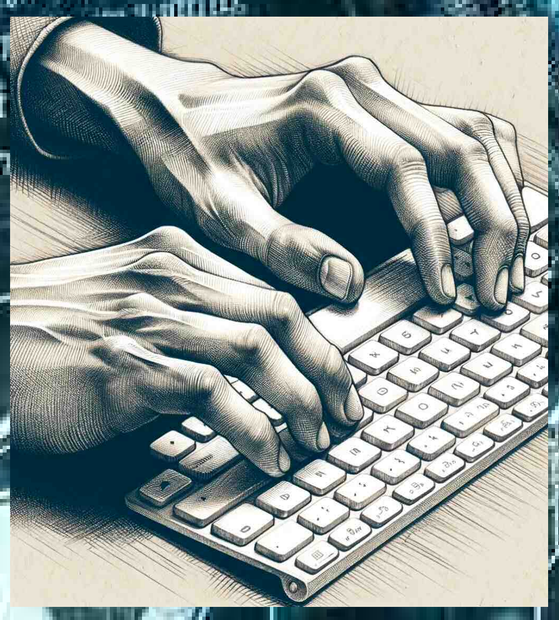
DAMSS DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Introduction

In today's dynamic cyber environment, threats as data breaches, cyber-attacks, unauthorized access threaten national security, and critical infrastructures. This research addresses the challenging task of protecting critical infrastructures from insider threats. To solve this problem, an innovative deep learning-based system for user identification using static authentication is proposed. The study involves the transformation of keystroke data from numerical or non-image data to image data to increase the intrusion detection capability and user authentication accuracy. New GAFMAT (GAbor Filter MAtrix Transformation) method for transforming numerical values into graphical representation using the Gabor filter is presented.

A Siamese neural network with a triplet loss function is used to detect anomalies or unauthorized intrusion into critical infrastructures. Network analyzes the unique characteristics of a user's password typing and compares them with previously typed passwords.

Keystroke Biometrics Transformation



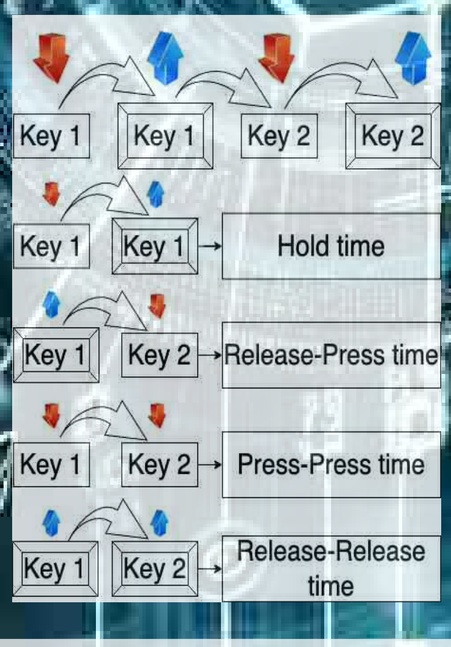


Fig. 1. Vizualizing keystroke dynamics capturing model when user is typing his password

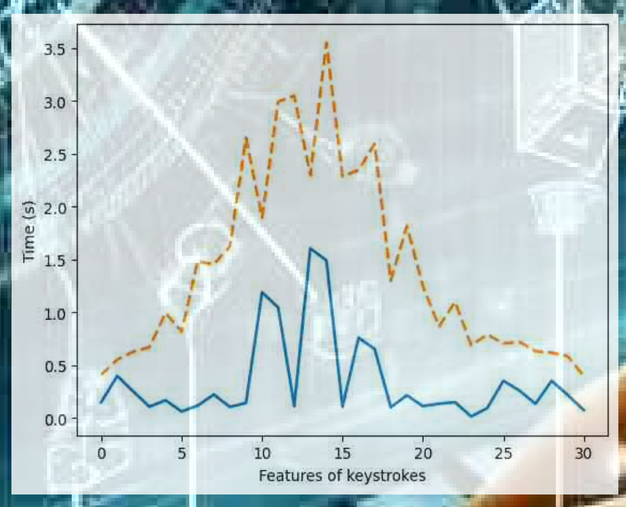


Fig. 2. Emphasizing the time-series features of keystroke dynamics using Gabor filter: blue for the discrete signal, dashed orange for the discrete signal after applying Gabor filter

$$image2D = \begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_nb_1 & a_nb_2 & \cdots & a_nb_n \end{bmatrix}$$

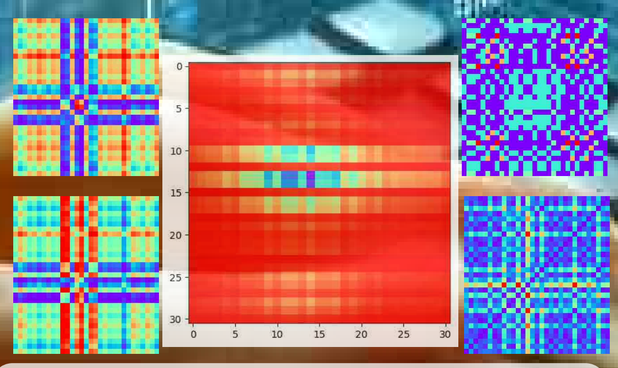
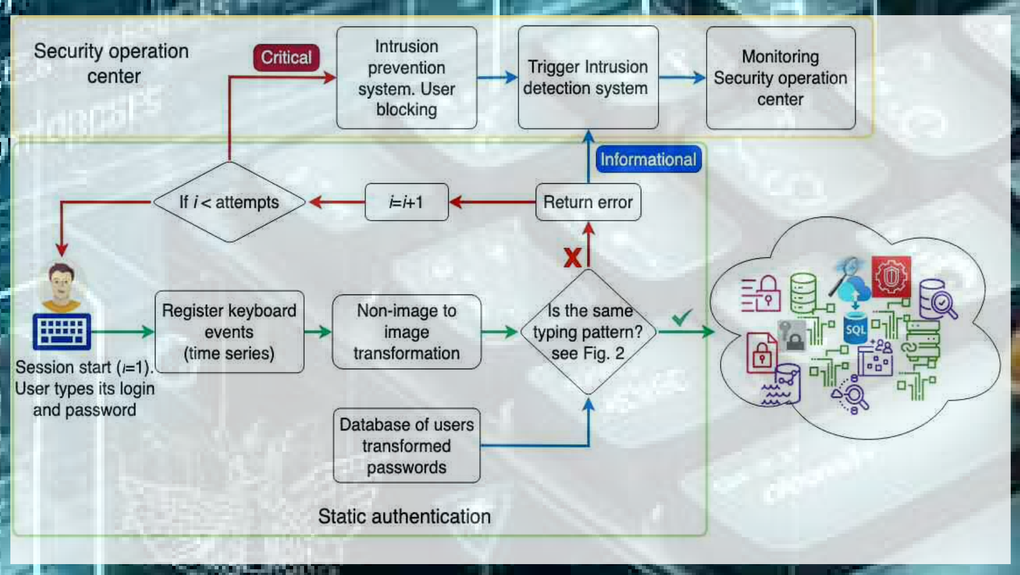


Fig. 3. The result of transforming the features of keystroke dynamics into an image using GAFMAT, RP, MTF, GASF, GADF methods

User Authentication Methodology



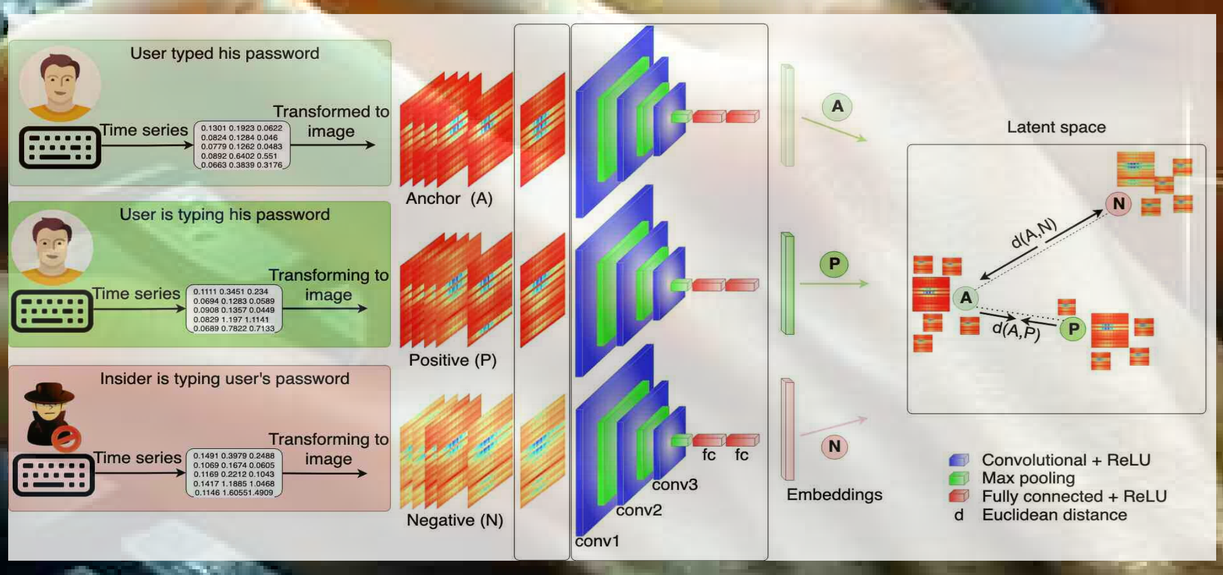
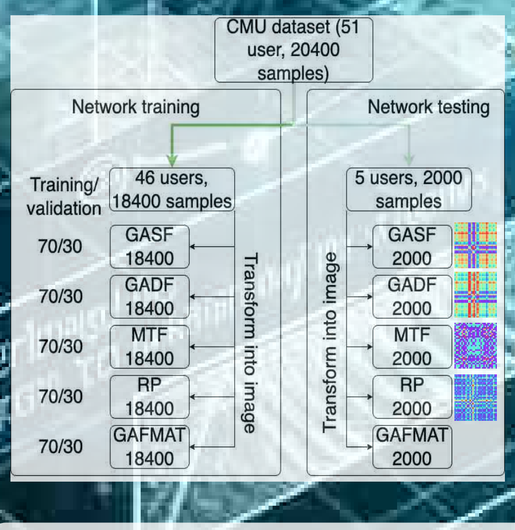


Fig. 4. User authentication process with Intrusion detection and Intrusion prevention systems based on user typing behavior using a Siamese neural network with triplet loss function

Fig. 5. Schematic representation of the proposed framework for time series transformation into images and training process of Siamese neural network with CNN branching to further improve the user authentication process

Data Pre-Processing



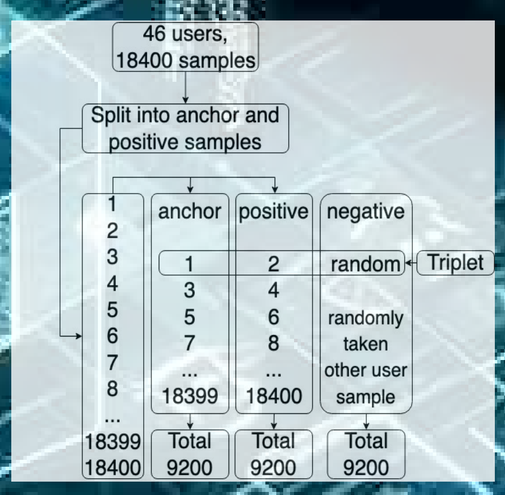


Fig. 6. The process of preparing CMU[1] data for model training/validation and for testing the trained network

Fig. 7. Splitting data into an anchor and positive samples for each transformed dataset for triplet preparation

Results

Table 1. Results of different metrics for passwords from GREYC-NISLAB[2] on a validation dataset when transforming time series features of keystroke dynamics into GAFMAT

Metrics	Passwords (GREYC-NISLAB)				
	leonardo dicaprio	the rolling stones	michaell schumacher	red hot chilli peppers	united states of america
Accuracy↑	0.97656	0.98698	0.99219	0.97778	0.99220
EER↓	0.07552	0.04688	0.0651	0.04444	0.04688
AUC↑	0.97824	0.98667	0.98771	0.98272	0.98847
AP_ED↓	0.44736	0.43986	0.39958	0.45165	0.39566
AN_ED↑	1.55644	1.61202	1.48864	1.63478	1.61275
AP_STD↓	0.24318	0.21992	0.20467	0.21505	0.19676
AN_STD↓	0.40601	0.37381	0.38351	0.38917	0.38013
AN_CS↓	0.49905	0.48703	0.52795	0.47839	0.4979
AP_CS↑	0.77632	0.78007	0.80021	0.77417	0.80217

Table 2. Results of image transformation methods on keystroke dynamics data from the CMU Dataset using GADF[3], GASF[3], RP[3], MTF[3], and GAFMAT

Metrics	Non-Image to Image Transformation Methods				
	GADF	GASF	RP	MTF	GAFMAT
Accuracy↑	0.99077	0.98473	0.98331	0.94744	0.98935
EER↓	0.04794	0.0554	0.05327	0.12074	0.04545
AUC↑	0.98612	0.9829	0.98394	0.94862	0.98668
AP_ED↓	0.44127	0.47255	0.43633	0.56487	0.486
AN_ED↑	1.72784	1.71689	1.68884	1.59469	1.76378
AP_STD↓	0.27487	0.29295	0.28245	0.36906	0.31383
AN_STD↓	0.32888	0.34455	0.34881	0.40005	0.31295
AN_CS↓	0.45772	0.45264	0.46871	0.46011	0.43755
AP_CS↑	0.77936	0.76373	0.78183	0.71756	0.757

References: 1. K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in 2009 IEEE/IFIP International Conference on Dependable Systems & Networks. IEEE, 2009, pp. 125–134. **2.** Y. Li, B. Zhang, Y. Cao, S. Zhao, Y. Gao, and J. Liu, "Study on the BeiHang keystroke dynamics database," in 2011 International Joint Conference on Biometrics (IJCB). IEEE, 2011, pp. 1–5. **3.** A. Estebsari and R. Rajabi, "Single residential load forecasting using deep learning and image encoding techniques," Electronics, vol. 9, no. 1, p.