A. Morales *et al.*, "Keystroke Biometrics in Response to Fake News Propagation in a Global Pandemic," *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, (2020): 1604-1609

DOI: https://doi.org/10.1109/COMPSAC48688.2020.00-26

Keystroke Biometrics in Response to Fake News Propagation in a Global Pandemic

Aythami Morales[1], Alejandro Acien[1], Julian Fierrez[1], John V. Monaco[2], Ruben Tolosana[1], Ruben Vera-Rodriguez[1], Javier Ortega-Garcia[1]
[1]*Biometrics and Data Pattern Analytics Lab, Universidad Autonoma de Madrid, Spain*
[2]*Naval Postgraduate School, Monterrey CA, USA*
{aythami.morales, alejandro.acien, julian.fierrez, ruben.tolosana, ruben.vera, javier.ortega}@uam.es, vinnie.monaco@nps.edu

***Abstract*—This work proposes and analyzes the use of keystroke biometrics for content de-anonymization. Fake news have become a powerful tool to manipulate public opinion, especially during major events. In particular, the massive spread of fake news during the COVID-19 pandemic has forced governments and companies to fight against missinformation. In this context, the ability to link multiple accounts or profiles that spread such malicious content on the Internet while hiding in anonymity would enable proactive identification and blacklisting. Behavioral biometrics can be powerful tools in this fight. In this work, we have analyzed how the latest advances in keystroke biometric recognition can help to link behavioral typing patterns in experiments involving 100,000 users and more than 1 million typed sequences. Our proposed system is based on Recurrent Neural Networks adapted to the context of content de-anonymization. Assuming the challenge to link the typed content of a target user in a pool of candidate profiles, our results show that keystroke recognition can be used to reduce the list of candidate profiles by more than 90%. In addition, when keystroke is combined with auxiliary data (such as location), our system achieves a Rank-1 identification performance equal to 52.6% and 10.9% for a background candidate list composed of 1K and 100K profiles, respectively.**

I. Introduction

In 2020, the COVID-19 pandemic is dominating worldwide media. In an overconnected world fueled by global panic, the propagation of fake news has achieved rates never seen before. Many of these fakes are based on ridiculous statements with scarce impact in a large percentage of the society (e.g. drinking water kills the virus or cocaine cures the virus [10, 28]). However, other fake news are more sophisticated and employed to modify public opinion, propagate panic, and destabilize governments.

The usage of fake news to manipulate public opinion has become normal in recent years, especially when major events such as elections and referendums take place. But during the COVID-19 pandemic, the spread of massive quantities of fake news has forced social media platforms to act. Companies such as Facebook or Twitter are working harder to detect and reduce the spread of fake news and bot profiles. Facebook introduced for example a context option that provides background information for the sources of articles in its News Feed, and Twitter has professional fact checkers to identify false content [14, 22]. During the COVID-19 outbreak these fact checkers detected anonymous profiles publishing fake news that go directly against guidance from authoritative sources of global and local public health information, aimed to influence people into acting against recommended guidance [12].

Data re-identification or de-anonymization is the practice of matching anonymous data with publicly available information, or auxiliary data, in order to discover the individual to which the data belongs to [20]. In the context of the fight against fake news, de-anonymization is useful to link multiple profiles belonging to the same user who is generating fake contents. Once detected, these users can be blacklisted based on their profile, MAC address, IP address, or other account data. However, these safeguards can be circumvented by creating a new account, changing device, or using a Virtual Private Network (VPN). Biometric technologies such as keystroke dynamics can be used to mitigate this circumvention. Data de-anonymization can cover any type of data, from text to audio, image, or video, being therefore a very challenging task [3, 24, 26]. Popular examples in this line are DeepFakes, which refer to deep learning based techniques able to create fake videos by swapping the face of a person with the face of another person [27]. In this work we focus in content that has been typed using a traditional keyboard (i.e. text).

Keystroke biometric recognition enables the identification of users based on their typing behavior. During the last 15 years, the efforts of the keystroke biometrics scientific community have been mostly focused on verification scenarios with a limited number of users, typically less than several hundred. The architecture proposed in [1], with experiments conducted on over 100,000 users, opened new research opportunities and challenges. The results over a user verification scenario revealed the potential of scaling up keystroke recognition. However, the suitability of this biometric trait for a large-scale identification scenario remains unexplored in the literature.

This work presents a feasibility study of content de-anonymization based on keystroke biometrics. To the best of our knowledge, this is the first work that analyzes keystroke identification for content de-anonymization. Our results suggest the potential of keystroke identification as a tool to improve the linkability between anonymous and verified profiles.

The rest of the paper is organized as follows. Sec. II summarizes the state of the art in keystroke recognition. Sec. III defines the problem and presents the proposed system. Sec. IV describes the experimental protocol, while in Sec. V we present the results. Sec. VI discusses the limitations and privacy concerns. Finally, Sec. VII draws the conclusions.

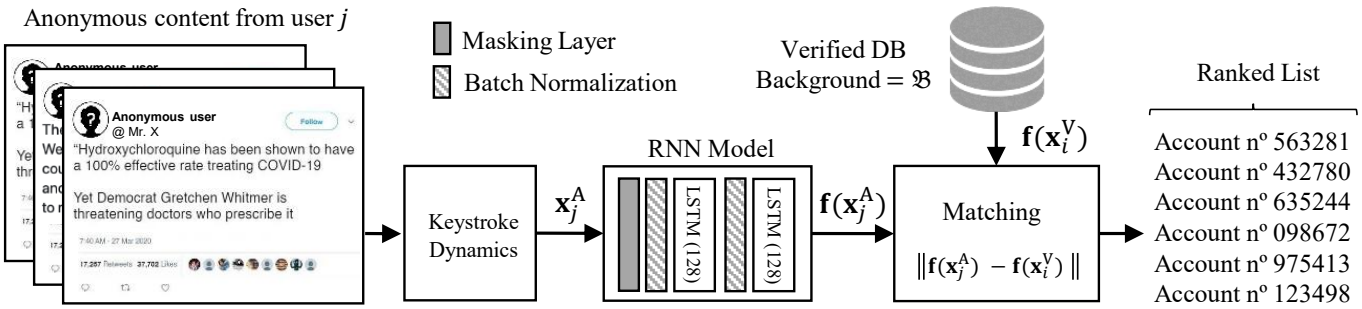

Fig. 1. General architecture of our proposed de-anonymization approach evaluated in this work.

II. Keystroke Biometrics: From fundamentals to the State of the Art

Keystroke biometric systems are commonly placed into two categories: *fixed-text*, where the keystroke sequence typed by the user is prefixed, such as a username or password, and *free-text*, where the keystroke sequence is arbitrary, such as writing an email or transcribing a sentence with typing errors. Free-text systems must therefore consider different text content between training and testing. Biometric recognition systems can be applied for *verification* or *identification* task. Verification implies a 1:1 comparison to determine if the biometric sample belongs to the claimed identity. Identification implies 1:N comparisons to determine the identity of the biometric sample from a pool of candidates. Biometric authentication algorithms based on keystroke dynamics for desktop and laptop keyboards have been predominantly studied for verification tasks in fixed-text scenarios, achieving accuracies higher than 95% [18, 19]. Approaches based on sample alignment (e.g. Dynamic Time Warping) [18], Manhattan distances [15], digraphs [6], and statistical models (e.g. Hidden Markov Models) [4] have achieved the best results in fixed-text verification.

However, the performances of free-text algorithms are generally far from those reached in the fixed-text scenario, where the complexity and variability of the text entry contribute to intra-subject variations in behavior, challenging the ability to recognize users [25]. Monrose and Rubin [17] proposed a free-text keystroke algorithm based on user profiling by using the mean latency and standard deviation of digraphs and computing the Euclidean distance between each test sample and the reference profile. Their results worsened from 90% to 23% of correct classification rates when they changed both users profiles and test samples from fixed-text to free-text. Gunetti and Picardi [11] extended the previous algorithm to n-graphs. They calculated the duration of n-graphs common between training and testing and defined a distance function based on the duration and order. Their results of 7.33% classification error outperformed the previous state of the art.

Recently, some algorithms based on statistical models have been shown to work very well with free-text, like the POHMM (Partially Observable Hidden Markov Model) [16]. Performance achieved using that approach in free-text is close to fixed-text, but requires several hundred keystrokes and has only been evaluated with a database containing less than 100 users. The latest advances in deep learning and the availability of large scale databases has boosted the performance of free-text keystroke recognition biometrics only very recently. In [1], a Deep Recurrent Neural Network architecture was presented with experiments over a database with 168,000 users and 136M keystrokes. Results obtained within a free-text verification scenario achieved error rates under 5%. Nevertheless, the performance of these algorithms for large scale identification scenarios remains unknown. This is one of the major contributions of this work.

III. Problem Statement and System Description

A. Problem Statement

Anonymous content is re-identified when multiple anonymous profiles are linked or associated to a verified profile. In our experiments, we assume that fake content was typed by an anonymous user who authored other verified content published on the same or different platform. A *verified content* is defined in this work as a content associated to a real identity (e.g. personal social media account or digital profile certified by a third party). An *anonymous content* is defined as a content published by one user who has not revealed his/her real identity (i.e. this content is usually associated with an alias or pseudonym).

In this work, we assume that timing sequences of the keyboard were captured when typing. This can occur when a user types content directly into a webpage. No special permissions are required to record the timestamps of keyboard input events generated on all major web browsers (e.g., Chrome, Firefox, Safari), and only the Tor Browser attempts to obfuscate typing behavior by lowering timestamp resolution [23]. De-anonymization is achieved by comparing the typing characteristics of the anonymous and the verified contents using the biometric patterns associated to the keystroke dynamics derived from these timing sequences. Fig. 1 presents the architecture of our proposed approach. The Anonymous typed content is first characterized according to the keystroke dynamics $\mathbf{x}^{\mathrm{A}}$ (A for Anonymous) associated to the sequences of time events t and keycodes k. A Recurrent Neural Network is used to project the timing and keycode sequences into a feature space trained for keystroke verification. The generated

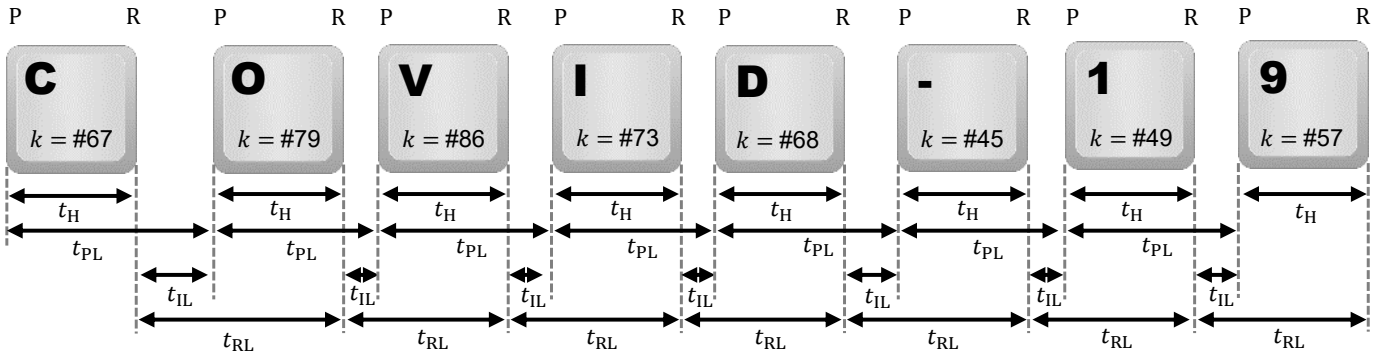

Fig. 2. Example of the 37 temporal features extracted from the term "COVID-19": 8 × Hold Time (t_H) + 7 × Inter-key Latency (t_{IL}), 7 × Press Latency (t_{PL}), 7 × Release Latency (t_{RL}), 8 × key codes (k). P = key Press event; R = key Release event.

feature vector $\mathbf{f}(\mathbf{x}^A)$ is characterized by the individual typing behavior of the anonymous subject. The feature vector $\mathbf{f}(\mathbf{x}^A)$ is then matched with each of the i feature vectors $\mathbf{f}(\mathbf{x}_i^V)$ of a Verified content database $\mathfrak{B}$ (the "background"), composed of $N = \#\mathfrak{B}$ profiles. The result of the matching process is a ranked list with the N profiles ordered by similarity to the anonymous subject.

B. Pre-processing and Keystroke Dynamics

The raw data captured in each keystroke sequence is composed of a three dimensional time series including: the keycodes, key press timestamps (corresponding to keydown events), and key release timestamps (corresponding to keyup events). In our experiments, timestamps were in UTC format with millisecond resolution (captured in a web browser), and the keycodes were integers between 0 and 255 according to the ASCII code.

From this raw data, we extract 4 temporal features popular in keystroke recognition (see Fig. 2 for details): (i) Hold Time (t_H): the elapsed time between press and release events; (ii) Inter-key Latency (t_{IL}): the elapsed time between releasing a key and pressing the next key; (iii) Press Latency (t_{PL}): the elapsed time between two consecutive press events; and Release Latency (t_{RL}): the elapsed time between two consecutive release events. These 4 features are commonly used in both fixed-text and free-text keystroke systems [5]. Finally, we include the keycodes as an additional feature.

Let L be the length of the keystroke sequence. The keycode and Hold Time features are calculated for each of the L keys in the sequence, and latency features between consecutive keys (t_{IL}, t_{PL}, and t_{RL}) are calculated for the $L-1$ consecutive key pairs. This produces a time series with shape $L \times 2 + (L-1) \times 3$. All feature values are normalized before being provided as input to the model. Normalization is important so that the activation values of neurons in the input layer of the network do not saturate (i.e. all close to 1). The keycodes are normalized between 0 and 1 by dividing each keycode by 255, and the 4 timing features are converted to seconds. This scales most timing features between 0 and 1 as the average typing rate over the entire dataset is 5.1 ± 2.1 keys per second. Only latency features that occur either during very slow typing or long pauses exceed a value of 1.

C. Recurrent Neural Network Architecture

We employ the Recurrent Neural Network (RNN) model proposed in [1]. The model is composed of two Long Short-Term Memory (LSTM) layers of 128 units. Between the LSTM layers there are batch normalization and dropout layers (0.5 drop rate) to avoid overfitting. Additionally, each LSTM layer has a 0.2 recurrent dropout rate. The network was trained with more than 1M keystroke sequences (over 50M keystrokes) from 68,000 different users (see Sec. IV-B for details).

The RNN was trained using a Siamese setup involving two inputs: two keystroke sequences from either the same or different users. During the training phase, the model learns the projections necessary to discriminate whether two keystroke sequences belong to the same user or not. The model acts as a feature extractor and outputs an embedding vector that contains the discriminating features (see [1] for details).

One constraint when training a RNN using standard back-propagation through time applied to a batch of sequences is that the number of elements in the time dimension (i.e. number of keystrokes) must be the same for all sequences. We fix the size of the time dimension to M. In order to train the model with sequences of different lengths L within a single batch, we truncate the end of the input sequence when $L > M$ and zero pad at the end when $L < M$, in both cases to the fixed size M. Error gradients are not computed for zeroed elements, which do not contribute to the loss function in the iterative learning due to the Masking layer indicated in Fig. 1.

Finally, the output of the RNN model $\mathbf{f}(\mathbf{x})$ is an array of size 1×128 that we consider later as an embedding feature vector to identify anonymous content based on Euclidean distance.

IV. Dataset and Experimental Protocol

A. Dataset

All experiments were conducted with the Aalto University Dataset [8] that comprises keystroke data collected from

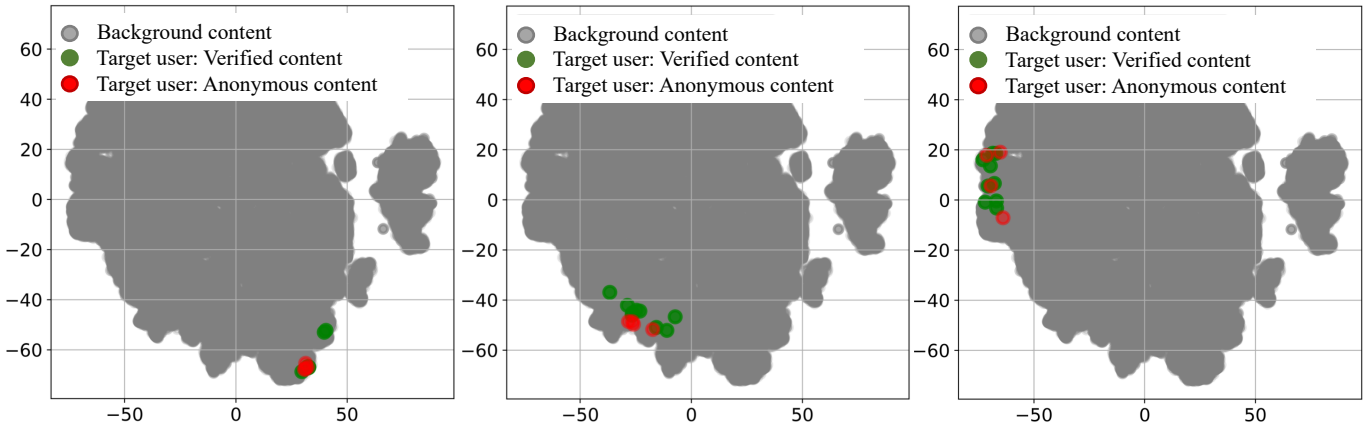

Fig. 3. t-SNE projections of 30,000 embedding vectors from 2,000 users. Each projection is obtained from the embedding generated by the Recurrent Neural Network. Each sub-figure contains the projections of the Anonymous and Verified keystroke sequences of a different target user. (Color image)

168,000 participants during a three-month time span. The acquisition task required subjects to memorize English sentences and then type them as quickly and accurately as they could. The English sentences were selected randomly from a set of 1,525 examples taken from the Enron mobile email and Gigaword newswire corpora. The example sentences contained a minimum of 3 words and a maximum of 70 characters. Note that the sentences typed by the participants could contain more than 70 characters as each participant could forget or add new characters when typing.

For the data acquisition, the authors launched an online application that recorded the keystroke data from participants who visited their website and agreed to complete the acquisition task (i.e. the data was collected in an uncontrolled environment). Press (keydown) and release (keyup) event timings were recorded in the browser with millisecond resolution using the JavaScript function `Date.now`. All participants in the database completed 15 sessions (i.e. one sentence for each session) on either a physical desktop or laptop keyboard. The authors also reported demographic statistics: 72% of the participants took a typing course, 218 countries were involved, and 85% of the participants had English as native language.

B. Experimental Protocol

The RNN was trained using the first 68,000 users in the dataset according to the method proposed in [1]. The size of the time dimension M was fixed to $M = 50$, which is short enough to consider small sentences. The remaining 100,000 users will be employed only to perform the evaluation of the de-anonymization system, so there is no data overlap between the two groups of users. Note that the model is unique for all the 100,000 users in the evaluation set, and does not require specific training when a new target user is added.

The 15 sequences from the 100,000 users in the database were divided into two groups that simulate a de-anonymization scenario: Verified (10 sequences) and Anonymous (5 sequences). We evaluated the de-anonymization accuracy by comparing the Anonymous set of samples $\mathbf{x}^{\mathrm{A}}_{j,l}$, with $l = 1, ..., 5$ belonging to the user j against the Background Verified set $\mathbf{x}^{\mathrm{V}}_{i,g}$, with $g = 1, ..., 10$ belonging to all 100,000 users. The distance was computed by averaging the Euclidean distances $||\cdot||$ between each Verified embedding vector $\mathbf{f}(\mathbf{x}^{\mathrm{V}}_{i,g})$ and each Anonymous embedding vector $\mathbf{f}(\mathbf{x}^{\mathrm{A}}_{j,l})$ as follows:

$$d_{i,j} = \frac{1}{10 \times 5} \sum_{g=1}^{10} \sum_{l=1}^{5} ||\mathbf{f}(\mathbf{x}^{\mathrm{V}}_{i,g}) - \mathbf{f}(\mathbf{x}^{\mathrm{A}}_{j,l})|| \tag{1}$$

We then re-identify an anonymous profile (i.e. Anonymous subject $j = J$ is the same Verified person $i = I$) as follows:

$$I = \arg\min_{i} d_{i,J} \tag{2}$$

The results reported in the next section are computed in terms of Cumulative Match Curve (CMC), which is a measure of 1:N identification system performance. The curves are calculated for each user and then averaged over all 100,000 users. A Rank-1 means that $d_{i,J} < d_{I,J}$ for any $i \neq I$, while a Rank-n means that instead of selecting a single Verified profile, we select n of them starting with $i = I$ by increasing distance $d_{i,J}$. In forensic scenarios, it is traditional to use Rank-20, Rank-50, or Rank-100 in order to generate a short list of potential candidates that are finally identified manually using a bag of evidence.

V. Experimental Results

A. Discriminatory Potential of Keystroke Biometrics

To ascertain the potential of the feature vectors generated by the keystroke model, we applied the popular data visualization algorithm t-SNE over the dataset. t-SNE is an algorithm to visualize high-dimensional data. This algorithm minimizes the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. Fig. 3 shows the projection of the keystroke embedding of three target users into a 2D space generated by the t-SNE algorithm. t-SNE projection is an unsupervised algorithm but for interpretation purposes we have colored three groups:

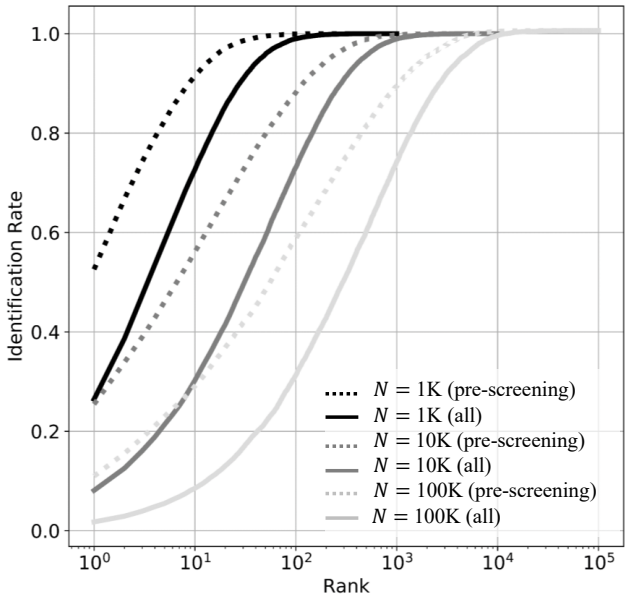

Fig. 4. CMC curves for different background sizes $N = \#\mathfrak{B}$. Pre-screening is based on the assumption that the location of the typist is available and can be used to reduce the number of candidates in the background set.

- Background projections including embeddings of 30,000 keystroke sequences from 2,000 random users. These sequences serve to visualize the boundaries and shape of the feature space represented by t-SNE.
- Validated sequences (10 per user) typed by 3 target users (each one in a different plot). These sequences model the target user's typing behavior.
- Anonymous sequences (5 per user) typed by the 3 target users. These sequences serve to re-identify the target users.

As we can see, the projections of the Validated and Anonymous keystroke embeddings from the target users are represented in close regions of the t-SNE space. Note that the t-SNE projection is not trained using the labels (i.e. identity) associated to each keystroke embedding. As a result, we observe that the personal typing patterns from the target users as represented by our deep network are discriminative enough in this large background set of 30,000 different sequences. These results suggest the discriminatory information available in this biometric trait.

B. De-anonymization Accuracy

As a 1:N problem, the identification accuracy varies depending on the size of the background set $\mathfrak{B}$. In our experiments, the size of the background is equal to the number of verified profiles available for the comparison. Fig. 4 and Table I present the de-anonymization performance for different background sizes. The results vary depending on the size of the background, with Rank-1 identification accuracy varying from 1.8% for the largest background (100K) to 26.4% for the smallest one (1K). These accuracies can be considered low for verification scenarios associated to user authentication. But

TABLE I
IDENTIFICATION ACCURACY (RANK-n IN %) FOR DIFFERENT BACKGROUND SIZES N. IN BRACKETS ACCURACY WITH PRE-SCREENING OF THE BACKGROUND DATASET BASED ON THE LOCATION OF THE TYPIST.

Rank	**Background Size N**		
	N = 1K	N = 10K	N = 100K
Rank-1	26.4 (52.6)	8.1 (25.4)	1.8 (10.9)
Rank-50	95.9 (99.5)	59.2 (79.7)	21.8 (48.4)
Rank-100	98.9 (99.9)	73.1 (88.1)	31.3 (58.8)
Rank-1000	100 (100)	98.9 (99.7)	74.2 (89.4)
Rank-5000	–	99.9 (99.9)	96.1 (99.6)

in the context of profile de-anonymization, a 1.8% accuracy among 100,000 profiles means that 1,800 profiles can be automatically de-anonymized using only the keystroke patterns of the owner. Rank-1 identification rate reveals the ability to unequivocally identify the anonymous target profile among all the verified profiles in the background set. Additionally, Rank-n represent the achievable accuracy if we consider a ranked list of n profiles from which the de-anonymization is then manually or automatically conducted based on additional evidence [9]. In general, the results suggest that keystroke de-anonymization enables a 90% size reduction of the candidate list while maintaining 100% accuracy (see the CMC curves in Fig. 4).

The number of background profiles can be further reduced if auxiliary data is available to realize a pre-screening of the initial list of verified profiles (e.g. country, language). The Aalto University Dataset contains auxiliary data including age, country, gender, type of keyboard, and others. Fig. 4 and Table I show also user identification accuracy over the entire background dataset with a pre-screening by country (i.e., contents generated in a country different to the country of the target user are removed from the background set). The results show that pre-screening based on a unique attribute is enough to largely improve the identification rate: Rank-1 identification with pre-secreening ranges between 10.9% to 52.6%, while the Rank-100 ranges between 58.8% to 99.9%. These results demonstrate the potential of keystroke dynamics for de-anonymization when auxiliary information is available.

VI. LIMITATIONS AND PRIVACY ASPECTS

The results presented in this work are encouraging. However, there are still some limitations regarding the application of this technology in the fight against fake news. The first one is that content must be typed. Spread of news by retweet or similar sharing mechanisms which do not require use of the keyboard, are not detectable by this technology. Second, bots are commonly employed for the propagation of fake content. It is not clear how the method proposed in this work would perform for synthetic behavior emulated by bots [2]. Third, the identification performance decays for a large number of background profiles. Therefore, pre-screening is recommendable to reduce the candidate list.

On the other hand, biometric data is considered sensitive data in a number of regulations (e.g. paragraph 71, EU GDPR). Keystroke dynamics, as biometric data, must be processed according to appropriate technical and organizational methodologies. The proposed de-anonymization based on keystroke behaviors is a powerful tool that can help in the fight against missinformation. But at the same time, the missuses of this technology arise important concerns related to data protection and user privacy. In addition to the identification accuracy studied in the paper, a concrete application of the ideas developed here should also consider and evaluate a secure storage of the biometric templates, and other modules for privacy preservation [7]. The balance is delicate in this case, as we should aim to de-anonymize problematic subjects while preserving the privacy rights and freedom of speech of the overall population at the same time. Careful consideration of such security and privacy aspects is out of the scope of the present paper and can be investigated elsewhere [13, 21].

VII. Conclusions

This work proposes keystroke biometric recognition for typed content de-anonymization. The fight against missinformation requires new tools, and the COVID-19 pandemic has showed the necessity to develop new technologies and policies to reduce the spread of fake content. Keystroke recognition can be used as a tool to link multiple profiles belonging to the same typist based on his typing behavior. We have evaluated a system based on Recurrent Neural Networks in experiments involving 100,000 users and more than 1M keystroke sequences. Our results suggest the potential of this technology to link multiple texts typed by the same user by leveraging personal typist patterns. The performance achieved varies depending on the number of background profiles, with Rank-1 identification accuracy ranging from 10.9% to 52.6% and Rank-50 from 48.4% to 99.5 when auxiliary information is available.

Acknowledgments

This work has been supported by projects: PRIMA (MSCA-ITN-2019-860315), TRESPASS (MSCA-ITN-2019-860813), BIBECA (RTI2018-101248-B-I00 MINECO), BioGuard (Ayudas Fundacin BBVA a Equipos de Investigacin Cientfica 2017). A. Acien and R. Tolosana are supported by a FPI and postdoc fellowship from the Spanish MINECO.

References

[1] A. Acien, J. V. Monaco, M. Morales, Aythami, R. Vera-Rodriguez, and J. Fierrez. TypeNet: Scaling up Keystroke Biometrics. *arXiv:2004.03627*, Feb. 2020.

[2] A. Acien, A. Morales, J. Fierrez, R. Vera-Rodriguez, and I. Bartolome. Be-CAPTCHA: Detecting human behavior in smartphone interaction using multiple inbuilt sensors. In *AAAI Workshop on Artificial for Cyber Security (AICS)*, Feb. 2020.

[3] T. Agrawal, R. Gupta, and S. Narayanan. Multimodal detection of fake social media use through a fusion of classification and pairwise ranking systems. In *Proc. European Signal Processing Conference*, pages 1045–1049, 2017.

[4] M. L. Ali, K. Thakur, C. C. Tappert, and M. Qiu. Keystroke biometric user verification using Hidden Markov Model. In *Proc. of IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)*, pages 204–209, 2016.

[5] A. Alsultan and K. Warwick. Keystroke dynamics authentication: A survey of free-text. *International Journal of Computer Science Issues*, 10:1–10, 01 2013.

[6] F. Bergadano, D. Gunetti, and C. Picardi. User authentication through keystroke dynamics. *ACM Transactions on Information Forensics and Security*, 5(4):367397, 2002.

[7] J. Bringer, H. Chabanne, and A. Patey. Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends. *IEEE Signal Processing Magazine*, 30(2):42–56, 2013.

[8] V. Dhakal, A. M. Feit, P. O. Kristensson, and A. Oulasvirta. Observations on typing from 136 million keystrokes. In *Proc. ACM Conf. on Human Factors in Computing Systems (CHI)*, 2018.

[9] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho. Multiple classifiers in biometrics. part 2: Trends and challenges. *Information Fusion*, 44:103–112, November 2018.

[10] Z. Gorvett. No, drinking water doesn't kill coronavirus. *BBC*, Mar. 2020.

[11] D. Gunetti and C. Picardi. Keystroke analysis of free text. *ACM Transactions on Information Forensics and Security*, 8(3):312347, 2005.

[12] A. Hern. Twitter to remove harmful fake news about coronavirus. *The Guardian*, Mar. 2020.

[13] E. J. Kindt. Privacy and data protection issues of biometric applications. *Springer*, 2013.

[14] B. Marr. Coronavirus Fake News: How Facebook, Twitter, And Instagram Are Tackling The Problem. *Forbes*, Mar. 2020.

[15] J. V. Monaco. Robust keystroke biometric anomaly detection. *arXiv:1606.09075*, June 2016.

[16] J. V. Monaco and C. C. Tappert. The partially observable Hidden Markov Model and its application to keystroke dynamics. *Pattern Recognition*, 76:449–462, 2018.

[17] F. Monrose and A. Rubin. Authentication via keystroke dynamics. In *Proc. of ACM Conference on Computer and Communications Security (CCS)*, page 4856, 1997.

[18] A. Morales, J. Fierrez, et al. Keystroke Biometrics Ongoing Competition. *IEEE Access*, 4:7736–7746, 2016.

[19] A. Morales, J. Fierrez, and J. Ortega-Garcia. Towards predicting good users for biometric recognition based on keystroke dynamics. In *Proc. of European Conference on Computer Vision Workshops, LNCS-8926*, pages 711–724, 2014.

[20] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy*, pages 111–125, 2008.

[21] P. Campisi (Ed.). Security and privacy in biometrics. *Springer*, 2013.

[22] G. Pennycook and D. Rand. The right way to fight fake news. *New York Times*, Mar. 2020.

[23] M. Perry. Bug 1517: Reduce precision of time for javascript, 2015. https://gitweb.torproject.org/user/mikeperry/tor-browser.git/commit/?h=bug1517.

[24] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36, 2017.

[25] T. Sim and R. Janakiraman. Are digraphs good for free-text keystroke dynamics? In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[26] S. Suwajanakorn, S. Seitz, and I. Kemelmacher-Shlizerman. Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics*, 36(4):1–13, 2017.

[27] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection. *arXiv:2001.00179*, 2020.

[28] M. Wehner. France had to tell citizens that cocaine wont cure coronavirus. *New York Post*, Mar. 2020.