Computers & Security 120 (2022) 102780

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

CLICKA: Collecting and leveraging identity cues with keystroke dynamics

Oliver Buckley[a,*], Duncan Hodges[b], Jonathan Windle[a], Sally Earl[a]

[a] School of Computing Sciences, University of East Anglia, United Kingdom
[b] Centre for Electronic Warfare, Information and Cyber, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, SN6 8LA, United Kingdom

ARTICLE INFO

Article history:
Received 27 September 2021
Revised 16 May 2022
Accepted 27 May 2022
Available online 29 May 2022

Keywords:
Biometrics
Keystroke dynamics
Identification
Behavioural biometrics
Security
Identity

ABSTRACT

The way in which IT systems are usually secured is through the use of username and password pairs. However, these credentials are all too easily lost, stolen or compromised. The use of behavioural biometrics can be used to supplement these credentials to provide a greater level of assurance in the identity of an authenticated user. However, user behaviours can also be used to ascertain other identifiable information about an individual. In this paper we build upon the notion of keystroke dynamics (the analysis of typing behaviours) to infer an anonymous user's name and predict their native language. This work found that there is a discernible difference in the ranking of bigrams (based on their timing) contained within the name of a user and those that are not. As a result we propose that individuals will reliably type information they are familiar with in a discernibly different way. In our study we found that it should be possible to identify approximately a third of the bigrams forming an anonymous users name purely from how (not what) they type.

1. Introduction

The traditional approach to securing a computer, device or service will typically rely on the use of a username and password pair. However, this approach is far from perfect and suffers from a number of obvious challenges. Users are required to create credentials that are both memorable to the user but also not easily guessed or inferred by a malicious third-party. These two criteria are incongruous and as a result the instances of compromised accounts are all too common. The involvement of a human in the process means that it will often be difficult to create security credentials that conform to security guidelines for generating strong passwords Gehringer (2002).

Once a user has been successfully authenticated there are typically no further challenges to their identity, which leads to the question of how much confidence can we have that the authenticated user is who their credentials claim them to be? Behavioural biometrics Yampolskiy and Govindaraju (2008) is an area that can be used to supplement traditional security methods with a continuous identification approach. Keystroke dynamics Monrose and Rubin (2000) captures a users keystrokes as a means of confirming the identity of the current user. Traditional methods of identification (such as a password) relies on what the user has typed, in this instance whether the password entered matches the stored password. Keystroke dynamics instead focuses on the way in which the user types, focusing on the timings of various key presses and releases. These timings can be used to build a unique identifier, a behavioural fingerprint of sorts, for an individual user.

While keystroke dynamics can be used to increase the confidence in the identity of an authenticated user, they can also provide personally information the individual. For example, keystroke dynamics have been used to determine a range of soft biometric traits including: gender, handedness or typing style Idrus et al. (2014). The work in this paper looks to extend the notion of keystroke dynamics and soft biometric traits to predict the name of an anonymous user. Typically, keystroke dynamics is used to increase the confidence in the identity of an authenticated user. In contrast to this our research looks to examine an individual's typing behaviours to predict the name of an anonymous user.

In this work we hypothesise that an individual will type specific combinations of characters (n-grams) discernibly quicker than others, based on the user's own personal experiences. We expect those combinations that are more familiar and more commonly used to be typed more automatically and quickly than others. This hypothesis is based on Fitts and Posner's

* Corresponding author.
E-mail address: o.buckley@uea.ac.uk (O. Buckley).

https://doi.org/10.1016/j.cose.2022.102780

Fitts and Posner (1967) three-stage model of motor learning, where a skill becomes more automatic as familiarity increases. For example, we posit that a user will type n-grams associated with their own name in a different manner to those not present in their name. A user's name can be considered a static identifier, in that it is rarely updated and is constant across a range of uses and platforms. However, information such as passwords change regularly and should be different across each use. This is the main reason for targeting a user's name in this work, to act as a proof of concept for the underpinning idea.

The research presented in this paper aims to understand the identity information that could be revealed when simply typing. This information could be used to increase confidence in the identity of the person using a device, provide information about an anonymous and potentially malicious internet user or to highlight personal information leakage to users.

The remainder of the paper will be structured as follows: Section 2 will provide an overview of the related material, Section 3 will detail the approach used, Section 4 will outline the analysis and results. This will initially identify the statistical properties of the data collected and identify whether there are differences between the typing patterns associated with names and emails. The paper will then consider machine learning approaches to exploiting the statistical variation, this will be followed by a short discussion of the practicalities of reconstructing full names from the gathered data. Finally Section 6 will discuss the conclusions and future direction of this work.

2. Related work

Biometrics are unique and measurable characteristics that can be used to identify and describe an individual and will typically fall into two broad categories: physiological and behavioural Delac and Grgic (2004). Physiological traits are related to distinguishing characteristics of the body of an individual, for example, fingerprints, eyes (both iris and retinal images), and vein recognition. Conversely, behavioural biometrics relate to the innate traits and behaviours displayed by individuals. Examples of behavioural biometrics include: keystroke dynamics (where a user's typing patterns are analysed), mouse dynamics (where a user's mouse movements are captured and analysed) and gait analysis.

Keystroke dynamics can be defined as the analysis of the way that an individual interacts with a keyboard, based on the timings of individual keystrokes Monrose and Rubin (2000). The typing patterns that are displayed by an individual can be uniquely identifiable in the same way as a person's signature or handwriting Douhou and Magnus (2009); Dvorak et al. (1936).

The study of keystroke dynamics has been an active area of interest since the early 1990s Bleha et al. (1990) and is usually deployed in one of two ways. The first application analyses keystroke dynamics when an individual is typing fixed text. For example, this approach is typically used as a means of password hardening Rudrapal et al. (2014). This approach to keystroke dynamics is typically looking to augment existing authentication methods. As well as the user providing something that they know (e.g. the password) the system will also analyse the way in which they type, based on an enrollment period involving the password being typed multiple times. This approach makes use of keystroke dynamics to confirm the identity of the user based on the credentials that they have supplied as well as the innate traits that they have displayed.

The second application of keystroke dynamics is the analysis of free text. These applications are broadly divided into two categories: those that require specific software to be installed on a client machine Pinto et al. (2014); Rybnik et al. (2008) and those that use a remote web-based method of data collection analysis. Messerman et al. Messerman et al. (2011) provide one such example of a remote, web-based collection and analysis approach to keystroke dynamics. This method of analysis is particularly attractive as a it provides a low-cost, remote means of identification and analysis without the requirement for specialised hardware or specific software to be installed. Instead data collection is accomplished by embedding code into any website that you have access to. This approach provides a means of continuous identification, where the user's identity is verified whenever they are actively typing. While it is expected that an individual will have identifiable typing behaviours, these behaviours are also liable to change. This could be due to increased familiarity with a keyboard or perhaps something more temporary, such as an injury or a change in mood Epp et al. (2011).

Typically, there are a limited number of features that can be extracted from keystroke data: dwell time, flight time and the timings of various substrings of characters Bleha et al. (1990). The dwell time is recorded with only a single keystroke and provides a measurement of the elapsed time between pressing and releasing a single key. The flight time measures the amount of time between releasing one key and pressing the next. The flight time requires the input of at least two keys and for experienced users this is often a negative value, as there is normally an overlap between the original key being released and the next key being pressed. Finally, an n-gram refers to the time taken to a type combination of n characters where the time is recorded between pressing the first key and releasing the *n-th* key Bergadano et al. (2002). In addition to these timing values, other features such as mistake ratio and middle time can also be gathered for analysis Giot et al. (2011).

One of the most common uses of keystroke dynamics is to confirm the identity of an authenticated individual Messerman et al. (2011). This method is essentially a pattern-matching process where the current typing activities are compared to a stored user profile. This means that the user has their current typing behaviours analysed to determine whether they match the stored behaviours of the authenticated user. Essentially this process is confirming whether or not the authenticated user is who their credentials claim them to be.

Keystroke dynamics are not only used in user identification and have also been used to successfully identify a range of *soft biometric traits*, as introduced by Jain et al. Jain et al. (2004). Soft biometric traits can be thought of as characteristics that provide some information about the individual but that cannot be used to uniquely distinguish between two individuals. Recent research has expanded this into other applications, such as detecting deception Monaro et al. (2017). These methods have also been used to better understand the writing process by using a keystroke data in conjunction with a range of complementary techniques. For example Leijten and Van Waes Leijten and Van Waes (2013) combine keystroke data with think aloud methods and eye tracking technologies to better understand the cognitive state of the author. Another notable area of research is the use of keystroke data to understand the stress levels of a user. Kolakowska Kołakowska (2016) uses keystroke data in isolation to understand the stress levels of programmers, and Vizer et al. Vizer et al. (2009) combines keystroke data with linguistic features to detect stress levels. The research that we present in this paper builds on this idea of soft biometric traits by using keystroke dynamics to determine identity data about an anonymous individual.

The research in this paper evolves these nascent research areas to exploit keystroke dynamics to infer identity data about an unknown and anonymous user. This work is a fusion of the previously discussed methods, which could be used to not just confirm the identity of a user but also infer key identity information about the user, such as name or email address. The aim of this research is a novel and innovative idea that looks to expand the existing areas of research in keystroke dynamics.

3. Method

3.1. Data collection

In order to capture keystroke data from a range of users a remote web-based data collection framework was required. The data collection frame used a simple web form with JavaScript key listeners attached to each of the text fields within the form.

Initially participants were required to provide some biographic details including:

- Forename
- Surname
- Age
- Handedness
- Gender

The participants were then required to copy a dynamically generated paragraph of text three times. The text is generated based on the participant's name. The forename and surname are segmented into two-character substrings (bigrams). The use of longer n-grams (e.g. trigrams or beyond) might potentially offer more information about a user's name. However, this would also decrease the frequency of their appearance in a user's name. This reduced frequency could limit the effectiveness of the machine learning model. The user's name is segmented into bigrams as below:

Participants name: **John Smith** Bigrams: **jo, oh, hn, sm, mi, it, th**

The generated bigrams are then used to create a paragraph of text that is specific to each participant. This approach ensures the data collection captures the participant typing the elements of their name, although not necessarily in the correct order. Without this approach it is difficult to guarantee a user would type all of the elements of their name.

When completing the typing tasks, the website makes use of JavaScript keyboard listeners to ensure that whenever a key is pressed the software will capture:

- The key that was pressed
- The timestamp of when the key was pressed
- The timestamp of when the key was released

It is important to note that the users were prevented from pasting text into any of the boxes to ensure that keystroke patterns were collected.

The study gained ethical approval through University of East Anglia Research Ethics Committee and all data was held securely following an approved Data Management Plan.

Participants were recruited across traditional social media channels using snowball sampling Goodman (1961) and also publicising the study through participant recruitment channels of Reddit and other websites. This wider reach attempted to reach as broad a range of user as possible. Participants were required to use a physical keyboard in order to participate. The study recruited 84 participants, all of whom were over the age of 18.

3.2. Inferring a User's name

3.2.1. Data preparation

Once the data had been collected a ranking of the 2-letter n-grams (bigrams) was created for each of the participants. This meant segmenting the data into bigrams and calculating the flight time, which is the time between releasing the first key and pressing the next.

Initial analysis was carried out to determine which of the typing measures (flight, dwell or a combination of the two) best enabled the inference of a participant's name. A Kolmogorov-Smirnoff two- tailed test was used to compare the similarity of two distributions, alongside a p-value (the probability that the event occurred by chance). This was used to compare the cumulative distribution function (CDF) for bigrams that were in a user's name and those not found in a user's name. Table 1 provides a comparison of these p-values at different levels of significance, which in turn highlights that flight time was the most effective measure.

The bigrams were ranked from the fastest to the slowest for each participant. Rankings are used rather than raw timings in order to normalise any variation between participants and their typing speed, experience or any physical characteristics (e.g. hand size). The research focuses on the relative speed between different bigrams rather than the absolute speed. This is based on the hypothesis that an individual user will type particular n-grams faster than others based on their familiarity. For example, if a user has the n-gram 'iv' in their name then it is hypothesised that this n-gram will rank more highly than for a user without the n-gram in their name.

The final stage of data preparation was to annotate the data to indicate whether or not a bigram appeared in the participants name or email address. This resulted in a CSV file for each typing activity with the following fields:

- Bigram (e.g. AB, CD, ER etc.)
- Ranking (e.g. fastest through to slowest)
- Does the bigram appear in the participant's email address?
- Does the bigram appear in the participant's name?

3.3. Inferring native language

In addition to developing a model to predict the name of an anonymous user this research aims to determine the native language of an individual based on their typing patterns. The data collection process is the same as that used when predicting name, where participants are required to provide demographic data and complete a number of typing tasks. During these tasks the software will again capture the same metrics (key pressed, time pressed and time released).

This experiment used Prolific as a tool for recruitment as it allows target recruitment of specific demographics, in this case native language. Participants were recruited based on their native language, with five languages, which all use the Roman alphabet. The aim was to recruit an even distribution of participants across all five of the selected languages, with 100 participants in each group. Ultimately, we were able to collect 492 usable typing samples for this experiment owing to some data being corrupt or incomplete. The breakdown of participants collected is as follows:

- English - 92
- French - 100
- Spanish - 100
- Italian - 100
- German - 100

3.3.1. Data preparation

The data preparation is a similar process, with the typing data broken down into bigrams for each participant. Again, the bigrams are ranked based on their flight time, which is the time between the first key being released and the second being pressed.

The hypothesis behind this process and preparation is similar in that each of the five languages have bigrams that appear more frequently than others. For example, 'th' is very common in the English language and as such participants that are used to using English will have a greater familiarity with these bigrams. As such the aim of this experiment is to leverage these common linguistic traits to predict an anonymous user's native language.

Table 1
A comparison of p-values at different levels of significance for CDF of bigrams that appear in a user's names and those that do not.

	Portion of bigrams with statistically different rankings between those who have that bigram in their name and those who do not		
Typing metric	0.01 significance	0.05 significance	0.10 significance
Flight time	4.7%	22.1%	33.5%
Dwell time	12.1%	22.7%	32.8%
Dwell and flight time	5.6%	23.6%	33.1%

Table 2
A summary of results for different machine learning algorithms.

Classifier	In Name (%)	False Positives (%)	Accuracy (%)	Balanced Accuracy (%)
XGBoost (1)	45.30	8.33	88.74	68.25
XGBoost (2)	64.87	21.97	76.15	70.83
Decision Tree	27.91	4.17	91.97	61.75
K-NN	17.80	4.60	90.99	56.46
Naive Bayes	12.18	8.98	86.31	51.34
SVM	53.57	19.78	77.73	66.34
AdaBoost	27.40	3.84	92.28	61.67
Random Forest	10.11	0.49	94.69	54.80

4. Analysis and results

4.1. Inferring a User's name

A stratified data resampling technique was adopted, with the data being resample 30 times. During the resampling the data is split into training and test splits at a ratio of 80% training data, which is used to build a model, and 20% test data, which is used to validate the model that have been produced.

While resampling the data is also standardised using only the training data, it should be noted that the data in this experiment is imbalanced with only approximately 5% of the labels in both the test and training dataset being labelled as true. This means that the number of bigrams that are in the typing data and that also appear in a user's name is very low. The data shows far more examples of bigrams that are not in the user's name, than examples of those that occur in a user's name.

A range of machine learning algorithms, and to ensure a fair comparison each algorithm was trained and evaluated on 30 resamples, where the resamples were the same across all algorithms. Hyperparameter optimisation is crucial to the performance of each algorithm and as such the hyperparameters were tuned for each of the algorithms that were used. A cross-validation grid search technique was used to choose the optimal parameters based on the training data alone. The parameter grid for important values (e.g. k = 1,3,7) was searched for the optimal value. This ensured that the test data provided no influence on the selection of parameters.

Table 2 highlights the performance of the different algorithms that have been applied to predicting the bigrams in a user's name. It can be seen that some algorithms offer an excellent accuracy, however, owing to the sparsity of the training data this figure cannot be used in isolation when determining the best performing algorithm. For example, a decision tree appears to be a very accurate classifier with an accuracy of nearly 92%. However, in this instance the algorithm was very good at predicting those bigrams that did not appear in the user's name.

A more representative measure of success is the algorithm's balanced accuracy. This provides a measure of accuracy in terms of bigrams that do appear in a name and those that do not. XGBoost produced the highest balanced accuracy and in name accuracy. The hyperparameters were crucial to this performance and the two XGBoost results highlighted in Table 1 describe some interesting trends. XGBoost(1) focused on reducing false positives, whereas XGBoost(2) was tuned to maximise accuracy, at the expense of introducing false positives. Despite only being separated by a minimal margin in balanced accuracy, there is a notable difference in both accuracy and false positives predicted. This is important as in future work this underpinning approach could be used in a predictive system, and the sensitivity of false positives may be of particular importance.

Using the balanced accuracy, the XGBoost algorithm offers the best performance with 70% of an anonymous user's name based on the observation of their typing behaviours alone.

4.1.1. Predicting a User's name

In order to fully understand how this technique could be applied in a real-world situation we developed a number of tests. These tests were designed to illustrate the capability of this approach in predicting the names of anonymous users. The method for predicting names made use of the US census data from 2000 Bureau (2010).

The US census data was used to generate a list of the most common names. Once our technique had predicted a set of bigrams that could form part of a user's name a list of potential names containing these bigrams was drawn from the census data. The popularity of each of these names within the census enabled the options to be ranked in terms of popularity. In essence we can exploit the knowledge that not all names are equally likely to guide the prediction. This does lead to a number of interesting conclusions, it was noted that one of the most common surnames in the US census is the name Hernandez, this is not the case in the UK. This implies the algorithm can be further steered by knowledge of the potential user bases.

To test the effect of the prediction we took the initial set of bigrams that were predicted and either injected random bigrams (to simulate the prediction providing a false positive) or randomly removing bigrams (to simulate the prediction missing a bigram). Using this altered set of bigrams it was then possible to see the effect of the prediction ranking. This experiment was performed over 100 times per participant and the results are shown in Fig. 1, with a red line representing the average performance.

This shows that on average the approach was more resilient to false-positives, as shown by the asymmetry in the average performance shown in Fig. 1. This underlines the importance of using XGBoost, tuned with different hyperparameters. The results detailed in Fig. 1 clearly illustrate that the introduction of false pos-

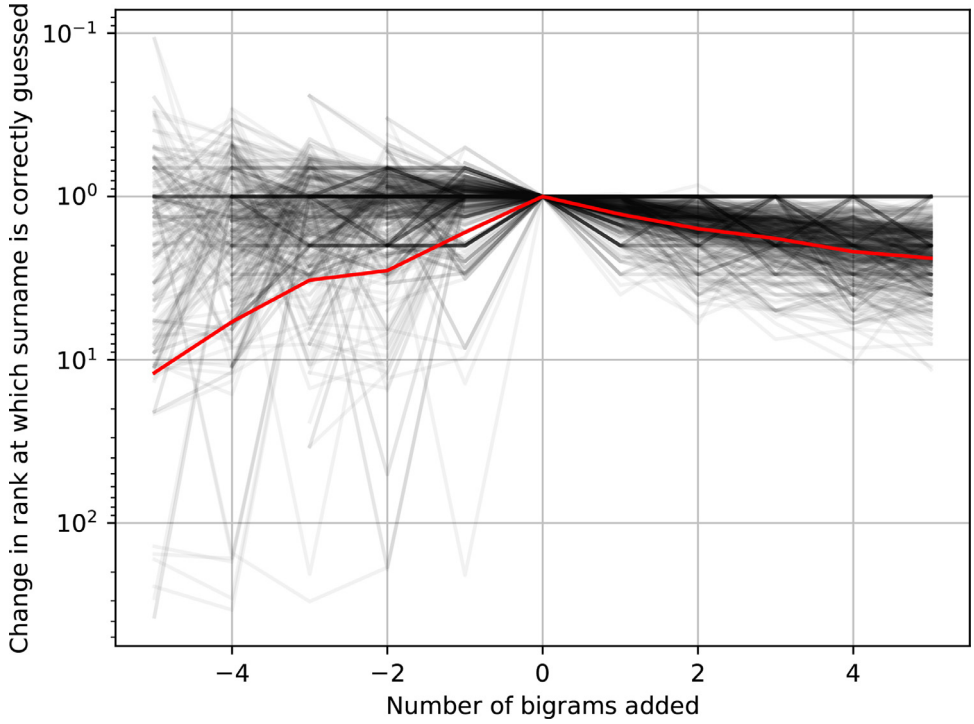

Fig. 1. The impact on ranking of false positives. The red line shows the average impact of adding or removing bigrams, with each of the black lines representing one of 10,000 tests. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

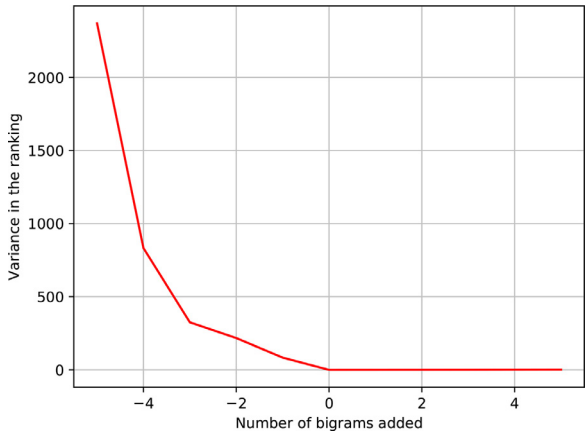

Fig. 2. The variance in ranking based on the number of bigrams added or removed.

itives does not significantly impact the prediction process. It is also noteworthy that the decay is also more predictable, with each false-positive experiment having a smaller variance than the equivalent 'miss' as shown in Fig. 2.

5. Inferring native language

As with the previous experiment to determine a participant's name, the timing between a key being released and the next key being pressed is recorded and assigned to the bigram value. There is a likelihood that these bigrams will be repeated, for example, the bigram 'en' will occur more than once in the typed passage of text. In this instance, where there are multiple instances of bigrams, the average (mean) time for every occurrence of that bigram is calculated.

When considering an individual's native language, the five most common bigrams for each language, according to Goldhan et al Goldhahn et al. (2012), were used but owing to an overlap in bigram popularity across the five languages this resulted in a total of 15 bigrams. These bigrams are:

- th
- he
- in
- er
- an
- de
- es
- en
- el
- la
- le
- et
- il
- ch
- ei

Each participant completed a typing activity, which captured the timings for a range of bigrams, not just those that were listed previously. Again, a participant's bigrams were then ranked, across all bigrams with these rankings forming hte input for the machine learning classification. Rankings are used, as opposed to raw timings to mitigate the differences in typing speed and experience across the entire cohort of participants.

The data was resampled 30 times and used an 80:20 split, where 80% of the data was used in training and 20% used to evaluate the model. The data was standardised using the Scikit-learn StandScaler, which is used to normalise the range of feature variables.

Random undersampling was used in some experiments where appropriate. This randomly removes training data from the majority class, for example, where there are 50 positive cases and 75 negative cases 25 negatives would be removed at random to ensure there are 50 in each class for training purposes. No sampling

Table 3
A summary of the results to classify a user's native language into English or other based on keystroke analysis.

Classifier	Accuracy	Balanced Accuracy	F1
SVM	0.72	0.71	0.72
Linear SVM	0.70	0.71	0.70
XGBoost	0.67	0.67	0.67
Naive Bayes	0.76	0.64	0.76
Decision Tree	0.71	0.59	0.71
KNN	0.70	0.55	0.70

is applied to the test data and is still unbalanced as removing test data will not improve the results and may ultimately be considered unreliable.

A number of machine learning classifiers have been used and for each of them hyperparameter tuning has been performed. A cross-validated grid search was used to find the parameters that offer the best results on the training data alone. These parameters are then used to train and evaluate each of the classifiers. A classifier is evaluated using the test data, which it has not seen before the evaluation and that is not used during hyperparameter optimisation. The process is repeated for 30 different data resamples, that is to say 30 different splits of training and test data. Each classifier uses the same data resamples in order to ensure a fair comparison between all classifiers.

5.0.2. Comparing english against all results

The first test performed was to develop a model for distinguishing between English and the other four languages used. This model looks to determine whether an individual's native language is English or something else. Essentially, this will provide a binary decision between whether English is their native language or not.

The best average balanced accuracy was 71%, which was achieved using the SVC classifier, as can be seen in Table 3, with a hyperparameter C value of 1,000,000 and gamma 1e-6. This provides a significantly better result than a random guess, and there is scope for further improving this result with a great volume of data and improved training model.

5.0.3. Native language prediction

The next stage was to develop a model to classify a user's native language as English, French, German, Spanish or Italian based on their typing behaviours.

The best average balanced accuracy was 45%, which was again achieved using the SVC classifier. The hyperparameter C value was 10 and a gamma value 0.01.

While the accuracy is not as good as the simple binary decision of English or another, 45% represents a better performance than a random guess.

6. Conclusions and future work

In this paper we have presented a unique method of inferring identity data about an unknown individual based solely on the analysis of their typing behaviours. This process uses a fairly sparse dataset, which simply includes the time that keys are pressed and released. The process presented then segments the data into two character substrings (bigrams), which are ranked based on the speed in which they are typed.

Our proposed method of predicting the bigrams contained within an anonymous user's name has proved to be successful, with a balanced accuracy of 71%. The ability to accurately predict more than half of the bigrams contained within a name offers all of the components required to reconstruct the name.

The XGBoost Chen and Guestrin (2016) algorithm has meant that we are able to predict with 70.83% balanced accuracy, the presence of a bigram within an anonymous user's name. This is a novel application of keystroke dynamics that evolves the current research to provide identifiable data, which is unwittingly leaked by users.

The biggest area for improvement for this result would be to collect an increased volume of data, when building a machine learning model in the first instance. A larger initial data set would provide a more robust and generalisable model, which could offer a greater level of accuracy. However, the 71% accuracy achieved with a relatively small initial cohort (84 participants) provides a validation of the hypothesis and highlights that this is indeed a promising approach.

When considering the native language of an anonymous user the experiment was able to work with a much larger cohort of 492 participants in total. The most promising result was when considering native language as a binary choice, that is to say between English or another language. The balanced accuracy of 70% shows a good level of accuracy when determining whether English is the user's native language.

However, when trying to predict the difference between one of five languages (English, French, German, Spanish, Italian) the accuracy is greatly reduced (to approximately 45%). While this result is still better than a random guess it is less than ideal. This could be for a number of factors, for example, a number of languages chosen share common bigrams. Similarly, there are changes that could be made to the way that the data is modelled when considering which popular bigrams to include.

This work offers a number of potential areas for further development as part of future work, whether that is to further improve the results or investigate new directions.

The current approach focuses solely on the use of bigrams, when monitoring a user's typing behaviours. Future work would look to investigate the impact of longer substrings, e.g. trigrams or longer, to understand the optimum n-gram length. Additionally, we rely only on the letters that a user types, and ignore punctuation. In future iterations of this work we would look to understand the implications of additional characters and punctuation as well as the of capital letters.

Predicting whether a user was a native English speaker or not proved to deliver good results (70%), however, this did not translate when trying to predict the language as a choice of five possible classes. Future work could see optimisations to the data collection and processing methods.

We have focused entirely on languages that use the Roman alphabet, expanding the scope of this work to cover languages that utilise different alphabets, would provide an interesting challenge.

The techniques identified in this paper have potential uses in predicting passwords and other security credentials. For example, Wang et al. Wang et al. (2016) present TarGuess a framework encompassing guessing algorithms based on the data that is potentially available to attackers. Identifying commonly typed, high-ranking bigrams could provide a starting point for one such approach to password prediction. Similarly, this could be used to increase the security of a user's password, by enabling users to avoid credentials containing their more commonly type bigrams. For instance, work by Pal et al. Pal et al. (2019) developed a model to warn users against picking passwords that are more susceptible to attack (e.g. if they were included in a previous breach).

The research carried out to date focuses on physical keyboards. As technology and our use of it continues to evolve there is an increased reliance on smartphones and tablets. The majority of these personal devices use a virtual keyboard on a touchscreen and so the logical evolution of this work is to attempt to apply the same principles to touchscreens. This could also leverage new paradigms for typing such as swipe keyboards.

Declaration of Competing Interest and Funding Information

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This research was funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/V002775/1), which is funded in part by the UK Home Office and security and intelligence agencies (see the public grant decision here: https://gtr.ukri.org/projects?ref=ES%2FV002775%2F1). The funding arrangements required this paper to be reviewed to ensure that its contents did not violate the Official Secrets Act nor disclose sensitive, classified and/or personal information.

CRediT authorship contribution statement

Oliver Buckley: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Duncan Hodges:** Conceptualization, Methodology, Writing – original draft. **Jonathan Windle:** Software, Investigation, Data curation. **Sally Earl:** Investigation, Writing – original draft.

References

Bergadano, F., Gunetti, D., Picardi, C., 2002. User authentication through keystroke dynamics. ACM Trans. Inf. Syst. Secur. (TISSEC) 5 (4), 367–397.
Bleha, S., Slivinsky, C., Hussien, B., 1990. Computer-access security systems using keystroke dynamics. IEEE Trans. Pattern. Anal. Mach. Intell. 12 (12), 1217–1222.
Bureau, U. S. C., 2010. Frequently occurring surnames from the 2010 census. https://www.census.gov/topics/population/genealogy/data/2010_surnames.html.
Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, pp. 785–794.
Delac, K., Grgic, M., 2004. A survey of biometric recognition methods. In: Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium. IEEE, pp. 184–193.
Douhou, S., Magnus, J.R., 2009. The reliability of user authentication through keystroke dynamics. Stat. Neerl. 63 (4), 432–449.
Dvorak, A., Merrick, N.L., Dealey, W.L., Ford, G.C., 1936. Typewriting behavior. New York: American Book Company 1 (6).
Epp, C., Lippold, M., Mandryk, R.L., 2011. Identifying emotional states using keystroke dynamics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 715–724.
Fitts, P.M., Posner, M.I., 1967. Human performance. Brooks/Cole.
Gehringer, E.F., 2002. Choosing passwords: Security and Human Factors. In: Technology and Society, 2002.(ISTAS'02). 2002 International Symposium on. IEEE, pp. 369–373.
Giot, R., El-Abed, M., Rosenberger, C., 2011. Keystroke Dynamics Overview. In: Biometrics. InTech, pp. 157–182.
Goldhahn, D., Eckart, T., Quasthoff, U., et al., 2012. Building large monolingual dictionaries at the leipzig corpora collection: From 100 to 200 languages. In: LREC, Vol. 29, pp. 31–43.
Goodman, L.A., 1961. Snowball sampling. Annal. Math. Stat. 32 (1), 148–170. http://www.jstor.org/stable/2237615.
Idrus, S.Z.S., Cherrier, E., Rosenberger, C., Bours, P., 2014. Soft biometrics for keystroke dynamics: profiling individuals while typing passwords. Comput. Secur. 45, 147–155.
Jain, A.K., Dass, S.C., Nandakumar, K., 2004. Soft Biometric Traits for Personal Recognition Systems. In: Biometric Authentication. Springer, pp. 731–738.
Kołakowska, A., 2016. Towards detecting programmers' stress on the basis of keystroke dynamics. In: 2016 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, pp. 1621–1626.
Leijten, M., Van Waes, L., 2013. Keystroke logging in writing research: using inputlog to analyze and visualize writing processes. Written Commun. 30 (3), 358–392.
Messerman, A., Mustafić, T., Camtepe, S.A., Albayrak, S., 2011. Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics. In: Biometrics (IJCB), 2011 International Joint Conference on. IEEE, pp. 1–8.
Monaro, M., Spolaor, R., Li, Q., Conti, M., Gamberini, L., Sartori, G., 2017. Type me the truth! detecting deceitful users via keystroke dynamics. In: Proceedings of the 12th International Conference on Availability, Reliability and Security, pp. 1–6.
Monrose, F., Rubin, A.D., 2000. Keystroke dynamics as a biometric for authentication. Future Generati. Comput. Syst. 16 (4), 351–359.
Pal, B., Daniel, T., Chatterjee, R., Ristenpart, T., 2019. Beyond credential stuffing: Password similarity models using neural networks. In: 2019 IEEE Symposium on Security and Privacy (SP). IEEE, pp. 417–434.
Pinto, P., Patrão, B., Santos, H., 2014. Free typed text using keystroke dynamics for continuous authentication. In: IFIP International Conference on Communications and Multimedia Security. Springer, pp. 33–45.
Rudrapal, D., Das, S., Debbarma, S., 2014. Improvisation of biometrics authentication and identification through keystrokes pattern analysis. In: International Conference on Distributed Computing and Internet Technology. Springer, pp. 287–292.
Rybnik, M., Tabedzki, M., Saeed, K., 2008. A keystroke dynamics based system for user identification. In: Computer Information Systems and Industrial Management Applications, 2008. CISIM'08. 7th. IEEE, pp. 225–230.
Vizer, L.M., Zhou, L., Sears, A., 2009. Automated stress detection using keystroke and linguistic features: an exploratory study. Int. J. Hum. Comput. Stud. 67 (10), 870–886.
Wang, D., Zhang, Z., Wang, P., Yan, J., Huang, X., 2016. Targeted online password guessing: An underestimated threat. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp. 1242–1254.
Yampolskiy, R.V., Govindaraju, V., 2008. Behavioural biometrics: a survey and classification. Int. J. Biom. 1 (1), 81–113.

Oliver Buckley holds a PhD degree in computer science from the University of Wales, Bangor, United Kingdom, specialising in soft tissue deformation simulation using haptics and visualisation. He is currently an Associate Professor in Cyber Security with the School of Computing Sciences, University of East Anglia. Prior to this he worked with Cranfield University, where he was a member of the Information Operations research group. His current research interests include cyber security, behavioural biometrics, with a current focus on keystroke dynamics, as well as the application of machine learning, and visualisation within the domain.

Jonathan Windle received a BSc degree in Computer Science from the University of East Anglia. During his undergraduate he undertook a Year in Industry placement at Boeing Defence UK as a software engineer. After completion of this degree he worked as a Research Associate focusing on user identification using keystroke dynamics. He is currently working towards a PhD at the University of East Anglia with research focusing on speech driven synthesis of character animation.

Duncan Hodges received his PhD in Electronics and Communications Engineering from the University of Bath, UK. He is currently a Senior Lecturer in Cyberspace Operations at Cranfield University and is based at the Defence Academy of the United Kingdom, Shrivenham. His research focuses on how our adversaries use digital technologies to achieve their goals.

Sally Earl received an LLB in Law from the University of East Anglia, and an MSc Computing Sciences, also from the University of East Anglia. After completion of her Masters Sally worked as a Research Associate, focusing on trust and privacy in chatbots.

Clicka: Collecting and leveraging identity cues with keystroke dynamics

Buckley, Oliver

2022-06-09

Buckley O, Hodges D, Windle J, Earl S. (2022) Clicka: Collecting and leveraging identity cues with keystroke dynamics, Computers and Security, Volume 120, September 2022, Article number 102780

https://doi.org/10.1016/j.cose.2022.102780