Investigating Sensor Modality Informativeness and Stability for Behavioural Authentication

Andraž Krašovec[1,2], Veljko Pejović[2]

[1] European Commission, Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
[2] University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia

Abstract

Close friends can recognise each other from subtle behaviour cues: from the way one walks, the proficiency with which they complete certain tasks, posture, and other aspects. With an ever-increasing number of sensors built into everyday objects, computers should be capable of identifying users from their behaviour reflected in the sensor-sampled signals. While behaviour-based authentication systems have recently been demonstrated, their utility is still questionable. More specifically, whether and to which extent different sensing modalities capture long-term persistent user behavioural traits represents an open question. In this paper we tackle this issue by analysing the informativeness and temporal invariance of data collected for the purpose of user identification in an Internet-of-Things office-like environment. We discover that the most informative sensors are not necessarily reflecting the most stable behavioural traits, which may have consequences for the future development of sensor-based authentication systems.

Keywords

behavioural authentication, IoT environments, informativeness analysis

1. Introduction

The weaknesses of one-off authentication methods, such as passwords [1, 2], as well as the increasing privacy issues related to storing and using human biometric information [3], have paved the way for behavioural biometric-based authentication. Rather than collecting sensitive information (i.e. fingerprints, face images, etc.), such authentication harnesses inherent behavioural patterns of individuals and uses them for identification. Traditionally, these patterns could only be unpicked from data stemming from a user's interaction with an interface that would also serve as a sensor. Examples of early behavioural biometric-based solutions, thus, include systems for authentication based on keyboard typing patterns and touchscreen interactions [4, 5].

The emergence of Internet of Things (IoT) brought approximately 30 billion sensor-enabled devices to a wide range of environments [6]. It has recently been demonstrated that, once users are placed in an IoT-rich environment, human behavior reflected in sensor readings can be used for authentication [7]. Moving beyond single device sensing brings clear scalability and robustness benefits. For instance, authenticating users based on data coming from diverse sensors implies that different aspects of behaviour could be captured when relevant – e.g.

Human-Computer Interaction Slovenia 2021, November 11, 2021, Koper, Slovenia
andraz.krasovec@ec.europa.eu (A. Krašovec); veljko.pejovic@fri.uni-lj.si (V. Pejović)

keyboard-based authentication is only informative when a user is typing, but motion sensors could be used when a user is walking around a room instead; moreover, with an added dimension (i.e. sensing modality) the authentication space is broadened – e.g. two users who exhibit the exact same typing patterns could still be discerned if their walking patterns are different.

Although multimodal IoT-based authentication has been demonstrated, it is still not clear whether and to what extent different sensing modalities contribute to the end goal of user identification. Answering this question would sharpen the focus, so that sensors reflect very general kinds of behaviour could be discarded and more informative sensors would be deployed instead. Furthermore, behaviour-based authentication relies on (machine learning) models of individual behaviour that have to be constructed beforehand. The models enable the actual authentication only if the sensor data collected at the test time matches the data used for the model training. Large discrepancies among the test and the training datasets stemming from the variability of patterns registered by the sensors would render the system unusable. The question of whether the behaviour reflected in a certain sensor modality is "stable" or not is yet another open research question.

In this paper we tackle the above issues and study the informativeness and temporal persistence of IoT-based sensor data when it comes to user authentication. Our work is based on the analysis of a previously compiled dataset containing real-world sensor data collected while 20 users performed three different tasks two times each. The richness of the dataset (six different sensing modalities) enables us to answer the above research questions and bring the following specific contributions to the area:

- We calculate two metrics of feature informativeness brought by different IoT sensors and identify the most informative (combinations of) sensors for the task of behaviour-based authentication;
- We examine stability (temporal invariability) of different sensing modalities and identify sensor which can perform robust behaviour-based authentication across different usage sessions;
- We propose and discuss alternative sensing modalities that could address the deficiencies of the sensors used in the analysed dataset and potentially increase the reliability of behaviour-based authentication.

The analysis performed in this paper points towards limited informativeness and a rather low stability of any single modality used for sensor-based authentication. Indeed, this has been hinted before, as multimodal continuous authentication has been shown to be far more accurate than any one-off authentication method relying on IoT sensors [7, 8]. Nevertheless, we believe that our work does not imply that opportunistic use of IoT sensors for authentication should be discarded. Instead, we believe that more sophisticated context-aware models as well as more diverse sensing modalities should be explored.

2. Related Work

Behavioural biometrics based authentication focuses on employing user's various behavioural patterns to validate one's identity. Modern approaches mostly rely on either sensor-rich

handheld and wearable devices, such as smartphones and smartwatches, or on exploiting sensor-equipped IoT environments.

Most common techniques on smart devices focus on either the user's interaction with the screen, which includes the analysis of navigation gestures, known as touch dynamics [5] or the keystroke dynamics [9] which study user's typing patterns, and even predate the modern computer [10]. Instead of requiring direct interaction from the user, it is sufficient for a device to be in a user's pocket, purse, or worn on a relevant body part for approaches reliant on the inertial measurement units (IMU) to function. This extends the applicability of authentication with smart devices, as, for example user can unlock her smartphone while running and listening to music, without the hassle of taking the phone out of her pocket. Most common authentication approach dependent on IMUs, is gait dynamics, which identifies users based on their walking patterns [11, 12].

Rather than burdening the user to never forget her authentication device, IoT environments-based authentication focuses on identifying the user without any concious actions from her, such as carrying a device, or even providing her fingerprint or remembering her password. Moving sensors from the user to the environment, IMUs still prove to be an effective modality for authentication [7]. Furthermore, wireless sensing technologies such as RFID [13] and short-mmwave radars [14] are closely tied to environmental IoT authentication.

The issue with most behavioural biometrics is the monomodality of their approaches, for example gait dynamics work well when a person is walking, while the moment she sits down, it is impossible to recognise her based on the way she walks. To overcome this limitation, multi modal approaches, that ensemble different sensing modalities into a more complete authentication system, started emerging. An overview of modality usage in biometric authentication by Ryu et al. [15] suggests most commonly used modalities in such systems include all of the above modalities, frequently tied to physiological biometrics, such as fingerprints and facial recognition. We follow the multi modal doctrine by collecting and publishing a dataset of different environmental sensors [7] which is further analysed in this paper.

3. Dataset

There are not many publicly available multi-modal IoT datasets that extend beyond data from smartphones and wearables. Therefore, we utilise a dataset we collected in our previous research [7]. It consists of data from fifteen different sensors, combined into five different sensor modalities. Twenty one participants completed three different everyday office tasks in an office-like environment. Additionally, we made the dataset publicly available [1].

We ask the participants to perform each of the devised tasks twice. The first task includes copying a body of text from a cardboard card to a PC and send it via email and it functions as a keyboard typing exercise, the second task requires the participants to look up for a weather forecast of a place of their choosing and suggest a couple tourist attractions to visit, based on the forecast, which should stimulate different PC usage patterns, and the third task sends the participants on a "treasure hunt", which includes navigating through a series of boxes including

[1] https://gitlab.fri.uni-lj.si/lrk/ca-iot

instructions on how to proceed and are placed around the environment. In the final box the instructions state to produce a graph from the data that are provided alongside the instructions.

The sensor we employ include; a) an accelerometer and b) a gyroscope, purposed to infer user's keyboard typing patterns, c) four force sensors, positioned on the corners of a plate that is placed under the mouse and keyboard, purposed to infer user's posture behind the computer, d) a PC monitor tool collecting the information about CPU, memory and network usage, purposed to infer user's computer usage patterns, and e) six infrared sensors, positioned around the environment, purposed to infer user's office navigation patterns. The complete dataset contains 115M raw datapoints over a span of fifteen hours and forty minutes.

Figure 1: A participant performing the second task. Most of the environmental bed is displayed in this image – the accelerometer and gyroscope attached to the top of the keyboard, force sensors under the corners of the white pressure plate, and the PC that was used for the user study, including the PC monitor components. The only modality missing in this photo, but still present in the environment, are the infrared sensors, which are placed around the experimental environment.

With the dataset acquired, we first apply some basic filtering techniques to exclude participants who either failed to follow the instructions, or we experienced technical difficulties with the data collection process. We retain fifteen users that successfully completed all six (2×3) tasks and do not have any missing sensor data. We process the remaining data by calculating time- and in case of accelerometer, gyroscope and force sensors also frequency domain features. A complete list of generated features is presented in [7].

In the following sections we utilise either the complete dataset (Section 4), or separate it on

task level, i.e. taking both sessions of a given task (Section 5).

4. Sensor Modality Informativeness

To better understand the inner workings of a multi-modal, environmental IoT authentication system, we calculate and analyse the informativeness of our generated feature set. That will enable us to better understand how to devise such systems in the future, especially in terms of selecting best performing sensing modalities, and replacing ones that do not provide sufficient user inference power. We estimate the informativeness of each feature based on two different informativeness metrics; relative mutual information and gini importance.

4.1. Relative Mutual Information

Our first informativeness evaluation metric is the Relative Mutual Information as defined in [5]. It separately calculates the mutual information between each observed feature and the user target variable, scaled by the entropy of the users:

$$\frac{H(U) - H(U|F)}{H(U)}$$

where $H(U)$ is the user entropy and $H(U|F)$ is the conditional user entropy of a given feature.We plot thirteen best scored features in Figure 2. Each bar represents a feature score. The key follows a simple pattern: *xy_zzz*, where *x* equals to a given sensor (a – accelerometer, g – gyroscope, f – force sensor, m – memory usage), *y* (optional) equals to an axis (x, y, z) for accelerometer and gyroscope, or a specific force sensor (a, b, c, d), while *zzz* equals to the generated feature (me – mean, min – minimum, max – maximum, mcr – mean crossing rate, sd – standard deviation).

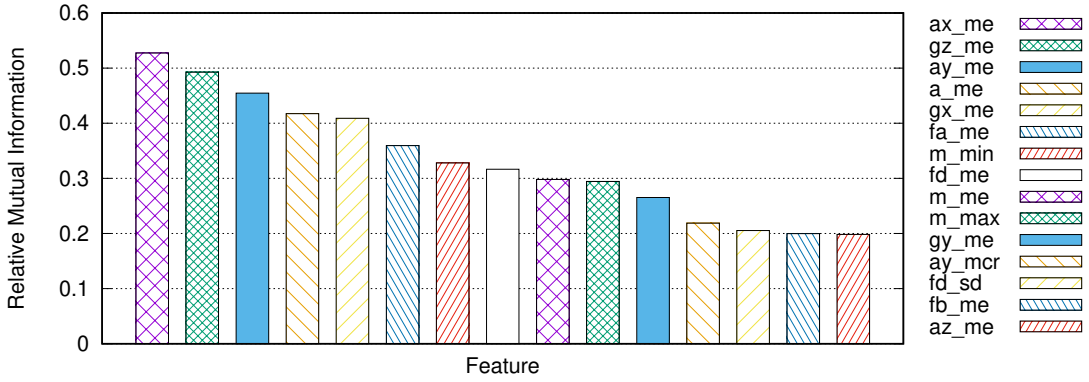

Figure 2: Relative mutual information of fifteen best performing features. Most of high scoring features are from the accelerometer, gyroscope, and force sensors, and belong to the time-domain feature space.

We observe a prevalence of time domain features, with the mean values of different senors being the most informative feature in most occurrences. Regarding the sensor modalities,

accelerometer provides best scores, followed by the gyroscope, force sensors, and memory usage. A similar trend extends beyond the displayed scores, as most frequency domain features, as well as data gathered from the infrared sensor are found as the least informative aspects of the gathered dataset.

4.2. Gini Importance

The other metric we utilise to evaluate the informativeness of generated features is the normalised gini importance. It is calculated by building a random forest machine learning model and averaging each feature's contribution to the decrease of impurity over trees, generated by the model. We choose the information gain as the decrease of impurity criteria, with the other available criteria – gini index, yielding similar results. The importance scores are normalised, meaning that the scores of all features sum up to one. We note that the gini importance score tends to favor either numerical features, or categorical features with high number of values [16]. All of our generated features are numerical, hence considered equally by this method.

Top fifteen gini importance scores are displayed in Figure 3. We reuse the key to note features in the plotted graph. Similarly to relative mutual information, features from the accelerometer and gyroscope are ranked highest, followed by the force sensors and memory usage patterns. Furthermore we observe the dominance of time domain features.

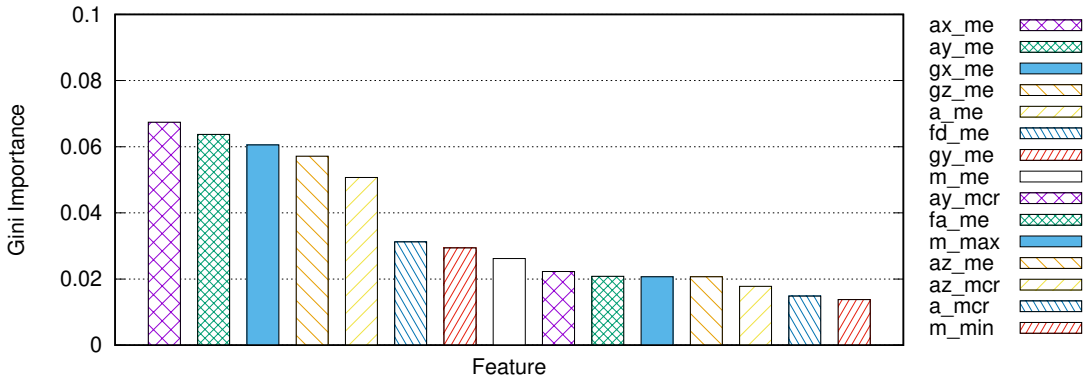

Figure 3: Fifteen most informative features according to the gini importance score. Similarly to relative mutual information, the accelerometer and gyroscope features produce the best scores, again belonging to the time-domain feature space.

5. User Behaviour Invariance

One major hurdle in behaviour based authentication are the varying user behaviour patterns that may be effected by the user's mood, cognitive load, and comfort level. Our work so far is no exception and the drop of accuracy in user inference when taking into account multiple sessions is significant. For instance, in [7] we achieve a 99% identification accuracy when the users are tracked through a single session, whereas when introducing multiple sessions our accuracy

drops to 70%. Therefore we investigate the potential of different sensor combinations not only in terms of raw accuracy, but also their ability to capture context-independent behavioural patterns that remain consistent throughout variations of user's behaviour.

5.1. Linear Discriminant Analysis Projection

We base our analysis on a two-dimensional linear discriminant analysis (LDA) projection, which is a supervised dimensionality reduction technique. To capture potential inter-sensor dependencies, we take all possible combinations of sensor pairs, as well as all possible combinations of user triplets. With the five different sensor modalities, and fifteen different users, we end up with $\binom{5}{2} * \binom{15}{3} = 4550$ different projections.

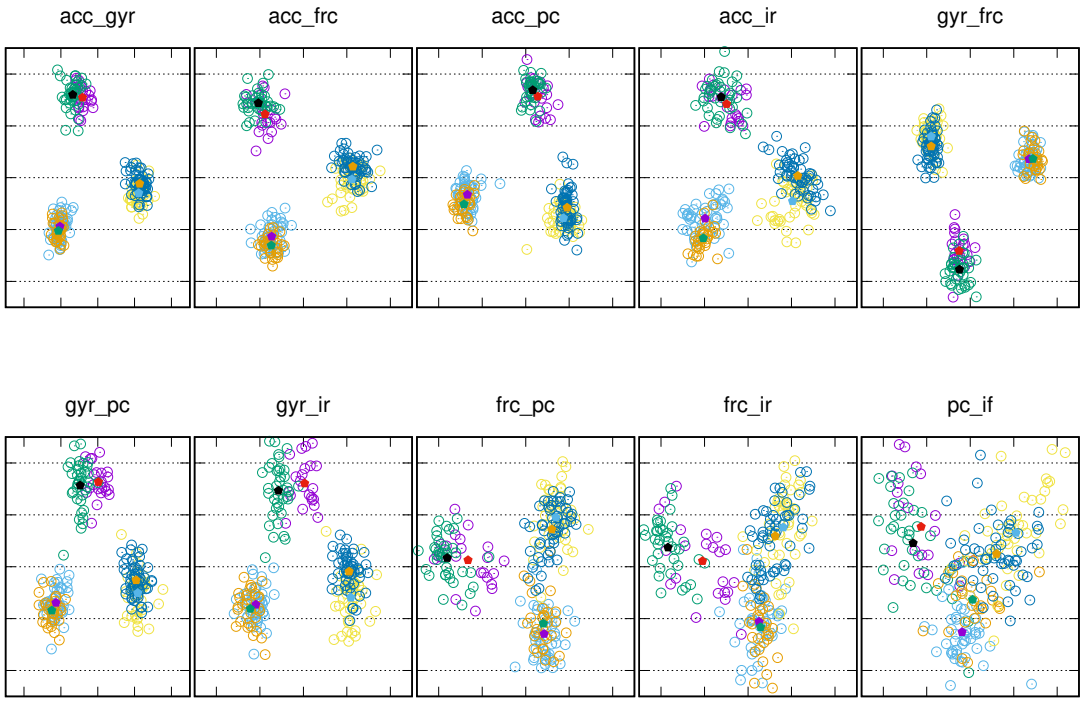

Figure 4: 2D LDA projection of all sensor modality pairs of three random users. Each subfigure title represents the sensor modality pair utilised, while different colours of points depict different user sessions. Pentagons relate to clusters' centroids.

In Figure 4 we display an example of LDA projections for three random users and all different sensor combinations. In each subfigure, circles represent projected LDA points, while pentagons are their centroids. Each colour equals to a single user session, where the following colour pairs correspond to the same user: (green, purple), (brown, light blue), and (yellow, dark blue). The titles represent two sensors that were taken into account in the specific projection, divided by an underscore (acc – accelerometer, gyr – gyroscope, frc – force sensors, pc – pc monitor, and ir

– infrared sensors). All subfigures are in the same scale, hence we are able to directly compare the clustering capability of each sensor modality pair.

Even to the naked eye, the ability to successfully cluster data of different users, varies with different sensor modality pairs. With either accelerometer or gyroscope included, the users clusters are clearly separable one from another, while in the absence of these two modalities, we observe clusters starting to blend.

5.2. Silhouette Scores

The convenience of visual representation of data with a 2D LDA enables us to quickly estimate result trends. However, to empirically evaluate clustering capabilities of different sensor modality pairs, we utilise the silhouette score, which is defined for each point a as:

$$S(a) = (y - x)/max(x, y)$$

where x is the mean intra-cluster distance to a given point, and y the mean distance to the points of the nearest cluster our given point is not a part of. Each individual value ranges from -1 to 1, with the latter representing the best possible value, while being closer to the former signifies a mis-clustered point. Mean of all silhouette scores represents a mean silhouette score of a sample i.e. data of three users and a sensor modality pair. To obtain the per-sensor modality pair score, we average all mean silhouette scores of a given sensor modality pair.

In Figure 5 we observe the final silhouette scores. In line with our assumptions from visually observing the LDA clusters, the accelerometer and gyroscope sensor modalities yield good results, with the combination of the two being the best overall. More surprisingly, the force sensor modality, which was scored much lower in informativeness scores (Section 4), is almost on a par with their silhouette scores. On the other side of the spectrum, including the infrared, and to a lesser extent the pc monitor sensor modalities, lowers the silhouette score, as those modalities do not contribute much to the user inference.

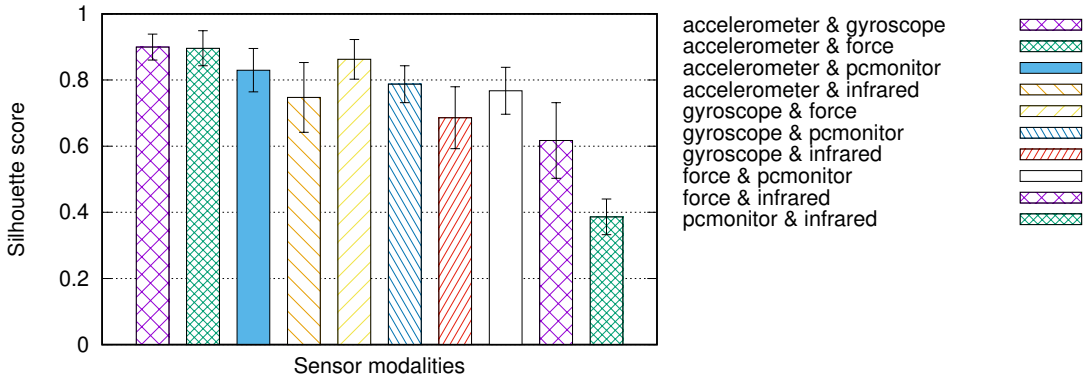

Figure 5: Mean silhouette scores of sensor modality pairs analysis. As expected the accelerometer and gyroscope sensor modalities perform the best, with the force sensors following closely behind. Error bars represent the standard deviation of the mean silhouette score.

5.3. Centroid Distances

Finally, our main goal of this section is to asses the consistency of different sensor modalities in inter-session user behaviour. We evaluate this by measuring the centroid distance of the session clusters that represent the same user performing the same task twice (two sessions). These are the distances between neighbouring pentagons in Figure 4. Each projection is standardized by removing the mean and scaling to standard deviation $y = \frac{x-\mu}{\sigma}$ before calculating the centroid distances.

Once more the accelerometer and gyroscope perform well. However, the best obtained (shortest) average centroid distance is achieved by pairing the acceleromenter with the force sensors, while the pairing with the gyroscope is not far behind. This result matches our findings with the mean silhouette score analysis and exhibits that while force sensors are not best suited for recognising users (as shown in Section 4), they have merit while it comes to inferring user identity accross sessions. We elaborate on this phenomenon in the Discussion section.

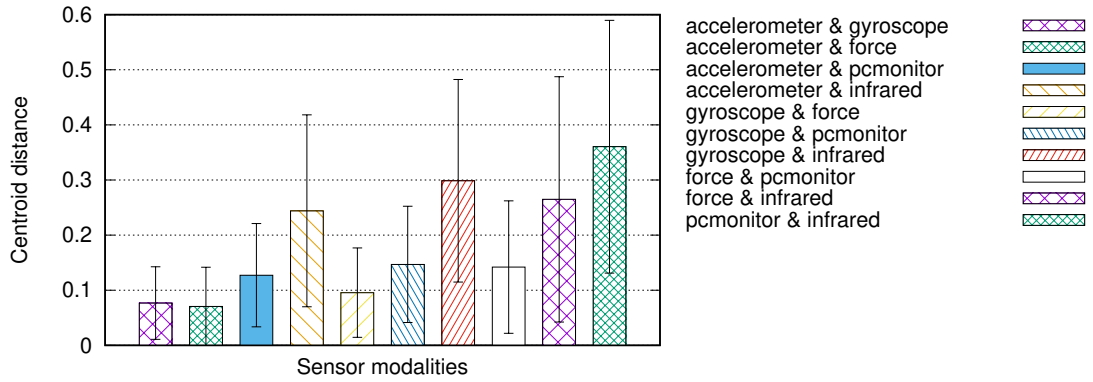

Figure 6: Centroid distances of 2D LDA clusters between two sessions of the same user for different sensor modality pairs. Beside the accelerometer and gyroscope, force sensors modality display great consistency between sessions. Error bars represent the standard deviation.

6. Discussion

Numerous alternatives to password-based authentication have been proposed by the research community over the last two decades [3, 4, 5, 13]. Occasionally impressive results cited in the literature, however, have recently been questioned on the grounds of very modest sample sizes and limited insights about why an approach may work or not [17]. Testing a novel authentication approach is extremely laborious, thus, small sample sizes are likely to persist for a while. Nevertheless, detailed examination of the authentication algorithm's inner workings may shed the light on its potential to provide scalable and robust authentication over a long period of time.

In this paper we analysed a behaviour-based authentication technique that harnesses IoT sensors placed in an office-like environment. The mutlitude of sensor modalities used in a

corresponding machine learning model poses a question of how and to what extent does a particular modality contribute to user identification. For this, we conducted an informativeness analysis. We find out that the relative mutual information, commonly used in behaviour biometrics-based authentication, does not clearly identify the most informative sensors. Instead, gini importance metric tends to clearly point out to sensors that are "closer" to a user, in our case accelerometer and gyroscope placed on a keyboard, as sensors that are likely to enable discernability among individuals.

Human behaviour tends to vary over time. Typing speed, for instance, might be modulated by a person's knowledge of the typed text, tiredness, emotions, and other factors. Whether the particular aspect of human behaviour captured by a sensor is going to persist in the same manner over time is a question we aimed to answer in the second part of the paper. Here we first presented the LDA projections to visually compare users' behaviour across the same tasks re-executed at different points in time. Then, we calculated the silhouette scores to assess whether a particular modality is likely to be informative over a longer period of time. Surprisingly, we find that a sensor that was not itself highly informative for user authentication – the force sensor – introduces stability not observed with some more informative sensors, such as the gyroscope. We further confirm this by calculating the centroid distances among clusters formed upon data stemming from different combinations of sensors. Centroids corresponding to the same person should not "move" across sessions, if the sensor-reflected behaviour is persistent. We find that including the force sensor in the mix indeed leads to smaller centroid displacements.

Why is behaviour reflect in the force sensor persistent? In the setup we analysed the force sensors were placed under the front panel of a desk the users were working at. The sensors, thus, likely capture the behaviour related to a user's general posture at a desk – the way one leans on the desk and the way arms are held during the typing. We postulate that unlike some other modalities, such as the typing speed, the force sensor reflects behaviour that does not change much between two sessions in the described experiments. While people may type faster or slower depending on whether they are familiar with a task or not (and indeed, we noticed that in the second repetition of the same task they often tend to type faster), they are unlikely to change their posture due to their perception of the task.

Our identification of a rather crude force sensor as a pillar of stability in behaviour-based authentication calls for the reconsideration of sensor types used for identification in IoT environments. Recently, wireless sensing has demonstrated impressive results: millimeter wave radars have been used to construct detailed 3D meshes of human users [18] and the same radars have been used for gesture recognition [19]. Guided by the above results and the fact that the sensing range can be tuned according to the distances and the details of interest, in our future work we aim to include wireless sensing in order to capture general postures of users and harness these for authentication.

7. Conclusion

In this paper we analysed real-world sensor data collected from an office-like IoT environment where 20 users conducted three different tasks in two sessions each. We focused on identifying the most informative sensors for behaviour-based authentication. We show that while sensors

that the users interact the most carry the highest identification potential, the behaviour they reflect is not necessarily the most stable. Instead, we observe that the sensors reflecting more intrinsic properties of users, such as their posture, tend to exhibit a higher temporal invariance. Consequently, our work stresses the need for additional sensors to be included, should mutlimodal IoT sensing become a viable authentication method in the future.

References

[1] R. Morris, K. Thompson, Password Security: A Case History, Communications of the ACM 22 (1979) 594–597. doi:10.1145/359168.359172.

[2] V. Zimmermann, N. Gerber, The password is dead, long live the password – A laboratory study on user perceptions of authentication schemes, International Journal of Human Computer Studies 133 (2020) 26–44. URL: https://doi.org/10.1016/j.ijhcs.2019.08.006. doi:10.1016/j.ijhcs.2019.08.006.

[3] D. F. Smith, A. Wiliem, B. C. Lovell, Face recognition on consumer devices: Reflections on replay attacks, IEEE Transactions on Information Forensics and Security 10 (2015) 736–745. doi:10.1109/TIFS.2015.2398819.

[4] S. Roy, D. Sinha, U. Roy, User authentication: keystroke dynamics with soft biometric features, Internet of Things (IoT): Technologies, Applications, Challenges and Solutions (2017) 99.

[5] M. Frank, R. Biedert, E. Ma, I. Martinovic, D. Song, Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication, IEEE Transactions on Information Forensics and Security (2013). doi:10.1109/TIFS.2012.2225048.

[6] I. Analytics, State of IoT 2021, 2021. URL: https://iot-analytics.com/number-connected-iot-devices/.

[7] A. Krašovec, D. Pellarini, D. Geneiatakis, G. Baldini, V. Pejović, Not Quite Yourself Today: Behaviour-Based Continuous Authentication in IoT Environments, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4 (2020).

[8] M. Khamis, M. Hassib, E. Von Zezschwitz, A. Bulling, F. Alt, GazeTouchPIN: Protecting sensitive data on mobile devices using secure multimodal authentication, ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction 2017-Janua (2017) 446–450. doi:10.1145/3136755.3136809.

[9] J. Kim, P. Kang, Freely typed keystroke dynamics-based user authentication for mobile devices based on heterogeneous features, Pattern Recognition 108 (2020). doi:10.1016/j.patcog.2020.107556.

[10] D. Buschek, A. De Luca, F. Alt, Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices, Conference on Human Factors in Computing Systems - Proceedings 2015-April (2015) 1393–1402. doi:10.1145/2702123.2702252.

[11] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, M. Savvides, Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics, 2012 IEEE 5th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2012 (2012) 8–15. doi:10.1109/BTAS.2012.6374552.

[12] F. Sun, C. Mao, X. Fan, Y. Li, Accelerometer-Based Speed-Adaptive Gait Authentication

Method for Wearable IoT Devices, IEEE Internet of Things Journal 6 (2019) 820–830. doi:10.1109/JIOT.2018.2860592.

[13] C. Feng, J. I. E. Xiong, L. Chang, F. Wang, J. U. Wang, D. Fang, RF-Identity : Non-Intrusive Person Identification Based on Commodity RFID Devices 5 (2021) 1–23.

[14] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, A. Markham, Human tracking and identification through a millimeter wave radar, Ad Hoc Networks 116 (2021) 102475. URL: https://doi.org/10.1016/j.adhoc.2021.102475. doi:10.1016/j.adhoc.2021.102475.

[15] R. Ryu, S. Yeom, S. H. Kim, D. Herbert, Continuous Multimodal Biometric Authentication Schemes: A Systematic Review, IEEE Access 9 (2021) 34541–34557. doi:10.1109/ACCESS.2021.3061589.

[16] S. Nembrini, I. R. König, M. N. Wright, The revival of the Gini importance?, Bioinformatics 34 (2018) 3711–3718. doi:10.1093/bioinformatics/bty373.

[17] S. Sugrim, C. Liu, J. Lindqvist, Recruit until it fails: Exploring performance limits for identification systems, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3 (2019) 1–26.

[18] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, L. Su, mmmesh: towards 3d real-time dynamic human mesh construction using millimeter-wave, in: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, 2021, pp. 269–282.

[19] S. Palipana, D. Salami, L. A. Leiva, S. Sigg, Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5 (2021) 1–27.