DYSAN: Dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks

Théo Jourdan[1], Antoine Boutet[1], Carole Frindel[1], Rosin Claude Ngueveu[2], and Sébastien Gambs[2]

[1]Univ. Lyon, INSA Lyon, Lyon, France. *{theo.jourdan,antoine.boutet,carole.frindel}@insa-lyon.fr*
[2]UQAM, Montreal, Quebec. *ngueveu.rosin-claude@courrier.uqam.ca*, *gambs.sebastien@uqam.ca*

Abstract

With the widespread adoption of the quantified self movement, an increasing number of users rely on mobile applications to monitor their physical activity through their smartphones. However, granting applications a direct access to sensor data expose users to privacy risks. In particular, motion sensor data are usually transmitted to analytics applications hosted on the cloud, which leverages on machine learning models to provide feedback on their activity status to users. In this setting, nothing prevents the service provider to infer private and sensitive information about a user such as health or demographic attributes. To address this issue, we propose DYSAN, a privacy-preserving framework to sanitize motion sensor data against unwanted sensitive inferences (*i.e.*, improving privacy) while limiting the loss of accuracy on the physical activity monitoring (*i.e.*, maintaining data utility). Our approach is inspired from the framework of Generative Adversarial Networks to sanitize the sensor data for the purpose of ensuring a good trade-off between utility and privacy. Experiments conducted on real datasets demonstrate that DYSAN can drastically limit the gender inference up to 41% (from 98% with raw data to 57% with sanitized data) while only reducing the accuracy of activity recognition by 3% (from 95% with raw data to 92% with sanitized data).

1 Introduction

The integration of motion sensors in smartphones and wearables has been accompanied by the growth of the quantified self movement [13]. For instance nowadays, users increasingly exploit these devices to monitor their physical activities. Usually, the motion sensor data are not analyzed directly on the device but are rather transmitted to analytics applications hosted on the cloud. These analytics applications leverage machine learning models to compute statistical indicators related to the status of users that are send back to them. While these analyses can bring many benefits from the health perspective [2, 10, 11], they can also lead to privacy breaches by exposing personal information regarding the individual concerned. Indeed, a large range of inferences can be done from motion sensor data including sensitive ones such as demographic and health-related attributes [3, 5, 6]. To address the issues raised by these scenarios, in this work we propose a solution sanitizing the motion sensor data in such a way that it hides sensitive attributes while preserving the activity information contained in the data. To achieve this objective, we design DYSAN, inspired from the framework of Generative Adversarial Networks (GANs) [9] to sanitize the sensor data. More precisely, by learning in a competitive manner several networks, DYSAN is able to build models sanitizing motion data to prevent inferences on a specified sensitive attribute while maintaining a high level of activity recognition. Furthermore, our approach aims at addressing the heterogeneous aspect of sensor data, which is inherent to the way each user moves, to the characteristics of the device used for data

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

collection and to the evolution of activity during the day. Thus, as one sanitizing model cannot provide the best utility and privacy trade-off for all users over time, DYSAN builds a set of diverse sanitizing models by exploring different combination of hyperparameters balancing loss functions of activity recognition, sensitive inference and data distortion terms. By doing so, DYSAN is able to dynamically select the model which provides the best trade-off over time according to the incoming sensor data. The evaluation of DYSAN on real datasets, in which the *gender* is considered as the sensitive information to hide, demonstrates that DYSAN can drastically limit the gender inference up to 41% while only inducing a drop of 3% on the accuracy of activity recognition. Our dynamic sanitization method overcomes several shortcomings of the state-of-the-art approaches, namely the use of the same sanitization model for all users over time, which may lead to a poor privacy-utility trade-off for atypical users.

2 DYSAN: Dynamic Sanitizer

To avoid an unwanted exploitation of the motion sensor data, these data are sanitized by DYSAN before being transmitted to the mobile application. This sanitizing process removes the correlations with the sensitive attribute in the sensor data while preserving the information necessary to detect the activity performed by a user. In addition, DYSAN also aims at limiting the distortion between the raw and sanitized data to preserve the utility for other analytical tasks. Finally, the resulting sanitized data are sent to an analytics application hosted on the cloud, exploiting machine learning models to classify the users activity and compute statistics related to their physical activity. Before exploiting DYSAN, multiple sanitizers corresponding to various utility and privacy trade-offs are built during the training. These models are then deployed on the smartphone. During the online phase, DYSAN selects the best sanitizer for the associated user. Both the training and the online phases are summarized in Figure 1.

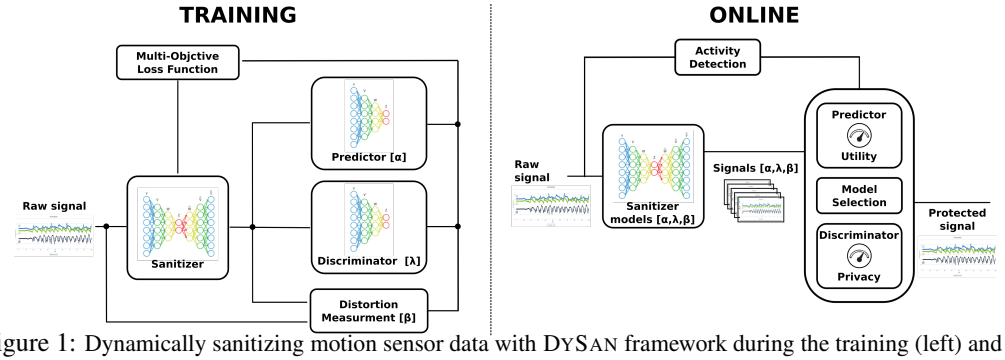


Figure 1: Dynamically sanitizing motion sensor data with DYSAN framework during the training (left) and the online (right) phases. Training phase allows to build different models that are distinguished by their parameters and online phase allows to choose among these models the most adapted to the final user.

Training Phase: DYSAN is composed of multiple building blocks that we detail hereafter: 1) a sanitizer which is an autoencoder that modifies the raw data to remove information correlated with the sensitive attribute while maintaining useful information for activity detection , 2) a discriminator which is a convolutional neural network that guides the sanitizer through the process of removing information related to the sensitive attribute, 3) a predictor which a convolution neural network that aims at helping the sanitizer in preserving as much information as possible with respect to the activity recognition task. , 4) a distortion measurement metric which is the last constraint on the sanitizer is the minimization of data distortion between the raw and sanitized data. During the training phase, we build a sanitizer for each set of possible values for the hyperparameters α and λ to explore the domain of the multi-objective loss function. This exploration will allow DYSAN to select the best model for each user during the online phase. The training procedure is summarized in Algorithm 1 (Appendix B). These neural networks compete against each other with opposing objectives until an equilibrium is reached. More precisely, the sanitizer is trained to fool the discriminator and maintained a high activity detection quantified with the predictor while limiting the data distortion.

Online Phase: Once deployed on the smartphone, DYSAN is composed of four components as depicted in Figure 1: the sanitizer, the discriminator, the predictor and an activity detection component. Specifically, DYSAN knows all the sanitizer, predictor and discriminator models built during the

training phase. This set of models correspond to the different possible utility and privacy trade-offs (*i.e.*, set of values explored for the α and λ hyperparameters). To find the best sanitizer over time, DYSAN evaluates the utility and the privacy of all models to select the best one. This evaluation requires to know the actual activity performed by the user and the sensitive attribute. While the sensitive attribute can be given by the user, the motion sensor data are not labeled with the activities as it is rather the objective of the activity recognition task to perform this inference. We use the activity detection component (see Figure 1) to annotate some motion sensor data with their activities on the smartphone. More precisely, we ask the user to follow a specific calibration process at the installation of DYSAN. During this process, the user is asked to perform a series of different activities for short periods to learn a specific supervised classifier to detect his activities. As the quantity of data available to train this classifier is limited, we rely on the use of random forests that are adapted to this context [4]. This random forest (RF) classifier is then used to label the raw data in order to evaluate the utility of all sanitizers. This evaluation is performed on a regular basis (*e.g.*, each period of p windows) and we compute the average accuracy over this period. By following this process, DYSAN is able to identify over time the sanitizer providing the best utility and privacy trade-off defined as a measure combining the accuracy of the activity recognition and the inference of the sensitive attribute.

3 Evaluation

In this section, we evaluate the capacity of an analytics application to infer the gender of the user and its activity from the sanitized data provided by DYSAN and sent by the mobile application. For the evaluation, the models are trained on MotionSense dataset and tested on MobiAct dataset in order to be in a context of transfer learning with new users. We compare DYSAN against baseline

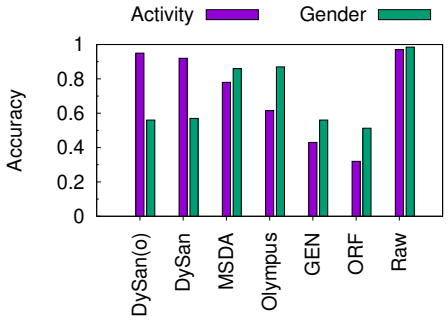


Figure 2: DYSAN provides the best privacy protection compared to state-of-the-art approaches at the cost of a slightly smaller accuracy in term of activity detection.

approaches (Figure 2). The baselines such as MSDA [8], Olympus [12] and GEN [7] also relies on an adversarial approach to optimize the utility and privacy trade-off but none of them takes into account the specificity of the users with a dynamic approach. ORF [4] is based on a Random Forest approach that extract the relevant features for both privacy and utility in order to select those that prevent the inference of the sensitive attribute. Two versions of DYSAN are given to represent, DYSAN where the annotations of the activities are known and the online version, DYSAN(o), where the activities are not given but inferred from the random forest (RF) classifier. The first version has been added for a more fair comparison to state-of-the-art that does not evaluate models as we suggest. Indeed, to dynamically select the sanitizer model, DYSAN needs to estimate the model providing the best utility and privacy trade-off with respect to the considered parameters. To achieve this, DYSAN relies on a calibration process to build a RF classifier on the raw data used as a reference to predict the current activity performed by the user. This RF classifier provides an average accuracy of 94% on the activity recognition for MobiAct dataset. While these accuracies are high, an activity wrongly predicted by this classifier leads to a selection of the sanitizer model that does not correspond to the best utility and privacy trade-off. As depicted on Figure 2, results for MobiAct show that DYSAN and DYSAN(o) outperform other approaches by limiting the gender inference to 55% and 54% while only reducing the accuracy of activity recognition by 2% and 5% compared to using the raw data, respectively. Although GEN and ORF also limit significantly the gender inference, the accuracy of the activity detection is drastically impacted (43% and 32%, respectively). By dynamically selecting the best sanitizer model for each window of raw data, DYSAN(o) makes the gender inference close to a random guess while preserving an accurate activity detection.

Dynamic selection of sanitizing model: During the training phase, DYSAN computes the sanitizer models corresponding to all possible utility and privacy trade-off by exploring the range of values for the hyperparameters α and λ. We evaluate here the benefit to dynamically adapt the sanitizing model according to the incoming data of each user compared to two static baseline approaches. Firstly, we compute the accuracy for both the gender inference and the activity recognition when the sanitizer model is fixed for all the users. This case represents the behaviors of all comparative baselines where the considered model is the one providing the best performance (*i.e.*, the utility and privacy trade-off) on average for all the users. Secondly, we consider a personalized solution where the sanitizer model is personalized for each user. In this case, the sanitizing model is the one which provides the smallest accuracy in term of gender inference and the best accuracy in term of activity recognition according to the whole models set for a specific user. This solution provides a sanitizer model personalization but the selected model is static and does not change according to the evolution of the incoming data (and the associated changes in term of performed activity).

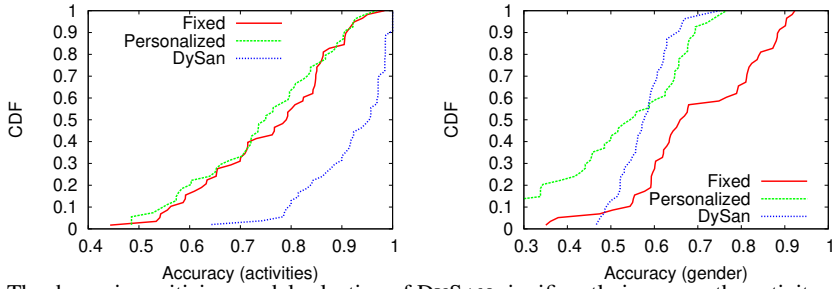


Figure 3: The dynamic sanitizing model selection of DYSAN significantly improves the activity recognition in case of transfer learning (*i.e.*, MobiAct dataset). In the same time, by dynamically adapting the sanitizing model for each user according to the incoming data, DYSAN greatly improved the protection against gender inference (the distribution of the gender accuracy is more centered around 0.5 which corresponds to a random guess).

We compare these both static solutions against DYSAN where the considered sanitizer model for each user changes according to the incoming data in order to maximize the utility and privacy trade-off over time. Figures 3 depicts on MobiAct dataset the cumulative distribution (*i.e.*, CDF) of the accuracy of the activity recognition and the gender inference respectively, when a fixed, a personalized, and a dynamic sanitizing model is considered. Firstly, results show that the accuracy in both classification tasks is highly heterogeneous over the population of users. This high heterogeneity reflects the fact that a static model is not well adapted for all users or for all activity performed by the user which motivates our dynamic approach. Specifically, results show that dynamically adapting the sanitizing model significantly improves the activity recognition compared to using a static model in case of transfer learning (*i.e.*, MobiAct dataset). For the gender inference, the objective of the sanitizer is to provide an accuracy around 0.5 which corresponds to a random guess for all users. However, results depicted in Figure 3 clearly shows that a fixed model for all users fails to protect against gender inference. Indeed, the distribution reports a wide range of accuracy over the users where it is possible to infer the gender with 80% of confidence for 60% of the users. Adopting a personalized sanitizer model for each user decreases the accuracy of the gender prediction compared to a fixed model for all users but the distribution of the accuracy is still large (from 0.3 to 0.8). By dynamically adapting the sanitizing model according to the incoming data, DYSAN greatly improves the protection against gender inference compared to using a fixed model with a sharper distribution centered around 0.5.

4 Conclusion

We presented DYSAN, a privacy-preserving framework which sanitizes motion sensor data in order to prevent unwanted inference of sensitive information. At the same time, DYSAN preserves as much as possible the useful information for activity recognition and other estimators of physical activity monitoring. Results show that DYSAN drastically reduces the risk of gender inference without impacting the ability to detect the activity or to monitor the number of steps. We also show that the dynamic sanitizing model selection of DYSAN successfully adapts the protection to each user over time according to the evolution of the incoming data. Lastly, we compared our approach with existing approaches and demonstrated that DYSAN provides better control over privacy-utility trade-off.

References

[1] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the crowd: The privacy bounds of human mobility. Nature, 3, 2013.

[2] G. D. Fulk and E. Sazonov. Using sensors to measure activity in people with stroke. Topics in Stroke Rehabilitation, 18(6):746–757, 2011.

[3] J. Han, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang. Accomplice: Location inference using accelerometers on smartphones. In COMSNETS, pages 1–9, 2012.

[4] T. Jourdan, A. Boutet, and C. Frindel. Toward privacy in iot mobile devices for activity recognition. In MobiQuitous, pages 155–165, 2018.

[5] J. L. Kröger, P. Raschke, and T. R. Bhuiyan. Privacy implications of accelerometer data: a review of possible inferences. In ICCSP, pages 81–87, 2019.

[6] S.-W. Lee and K. Mase. Activity and location recognition using wearable sensors. IEEE pervasive computing, 1(3):24–32, 2002.

[7] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi. Protecting sensory data against sensitive inferences. In W-P2DS'18, pages 2:1–2:6. ACM, 2018.

[8] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi. Mobile sensor data anonymization. IoTDI, 2019.

[9] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng. Recent progress on generative adversarial networks (gans): A survey. IEEE Access, 7:36322–36333, 2019.

[10] E. Park, H.-J. Chang, and H. S. Nam. Use of machine learning classifiers and sensor data to detect neurological deficit in stroke patients. J Med Internet Res, 19(4):e120, 2017.

[11] J. Qi, P. Yang, D. Fan, and Z. Deng. A survey of physical activity monitoring and assessment using internet of things technology. In CIT/IUCC/DASC/PICOM, pages 2353–2358, 2015.

[12] N. Raval, A. Machanavajjhala, and J. Pan. Olympus: Sensor privacy through utility aware obfuscation. Proceedings on Privacy Enhancing Technologies, 2019(1), 2019.

[13] H. Yang, J. Yu, H. Zo, and M. Choi. User acceptance of wearable devices: An extended perspective of perceived value. Telematics and Informatics, 33(2):256–269, 2016.

Appendices

A Neural Network Architecture

We provide in this section details about the underlying neural networks of DYSAN.

A.1 Discriminator Net

1. Input (125,6)
2. Conv1D (64, kernel_size=6, stride=1, activation=ReLU)
3. AvgPool1D(kernel_size=2, stride=2)
4. BatchNorm1D(100, eps=1e-05, momentum=0.1)
5. Dropout(p=0.5)
6. Dense(64, activation=ReLU)
7. Dense(2, activation=softmax)

A.2 Predictor Net

1. Input (125,6)
2. Conv1D (100, kernel_size=6, stride=1, activation=ReLU)
3. AvgPool1D(kernel_size=2, stride=2)
4. BatchNorm1D(100, eps=1e-05, momentum=0.1)
5. Conv1D(100, kernel_size=5, stride=1, activation=ReLU)
6. AvgPool1d(kernel_size=2, stride=2)
7. Conv1D(160, kernel_size=5, stride=1, activation=ReLU)
8. AvgPool1d(kernel_size=2, stride=2)
9. Conv1D(160, kernel_size=5, stride=1, activation=ReLU)
10. AvgPool1d(kernel_size=2, stride=2)
11. Dropout(p=0.5)
12. Dense(64, activation=ReLU)
13. Dense(4, activation=softmax)

A.3 Sanitizer Net

1. Input (125,6)
2. Conv1D (64, kernel_size=6, stride=1,)
3. Conv1D (128, kernel_size=5, stride=1)
4. Dense(128)
5. Dense(64, activation=LeakyReLU(0.01))
6. Dense(64)
7. Dense(128)
8. Deconv1D (128, kernel_size=5, stride=1)
9. Deconv1D (64, kernel_size=5, stride=1, activation=softmax)

	Mean	Std	Skewness	Kurtosis	Energy
Raw	0.81	0.47	1.65	4.81	139.06
DySan	0.68 (-15.9%)	0.77 (+62.9%)	0.40 (-75.7%)	1.28 (-73.5%)	230.87 (+66.0%)
GEN	0.28 (-65.4%)	0.12 (-74.7%)	0.51 (-69.2%)	0.08 (-98.3%)	12.11 (-91.3%)
Olympus	5.40 (+566.4%)	2.52 (+433.1%)	0.61 (-62.8%)	0.29 (-94.0%)	4631.47 (+3230.5%)
MSDA	0.54 (-33.5%)	0.24 (-49.9%)	0.41 (-75.2%)	-0.11 (-102.2%)	51.87 (-62.7%)

Figure 4: Similarities metric between the raw data and the different baselines. Mean, standard deviation (std), skewness, kurtosis, energy are given in percentage of relative error.

B Algorithm of the Training Phase

Algorithm 1 DYSAN training algorithm

Input: X, λ, α, max_epoch, $batch\_size$, K_{pred}, K_{disc}.
Outputs: S_{an}, D_{isc}, P_{red}.
train(M, **trParams): Train the model M using trParams.
freeze(M): Freeze the model M parameters and avoid modifications.
{Initialisation}
S_{an}, D_{isc}, P_{red}, $X_d = \mathsf{shuffle}(X)$, $X_p = \mathsf{shuffle}(X)$
$\mathsf{Iterations} = \frac{|D|}{\mathsf{batch\_size}}$
{Training Procedure}
for $e = 1$ **to** max_epoch **do**
  for $i = 1$ **to** Iterations **do**
    Sample batch B of size batch_size from X
    train(S_{an}, B, $J^{S_{an}}$, α, λ, freeze(P_{red}), freeze(D_{isc}))
    for $k = 1$ **to** K_{pred} **do**
      Sample batch B of size batch_size from X_p
      train(P_{red}, B, $Loss_{Activities}$, freeze(S_{an}))
    end for
    for $k = 1$ **to** K_{disc} **do**
      Sample batch B of size batch_size from X_d
      train(D_{isc}, B, $Loss_{Sensitive}$, freeze(S_{an}))
    end for
  end for
end for

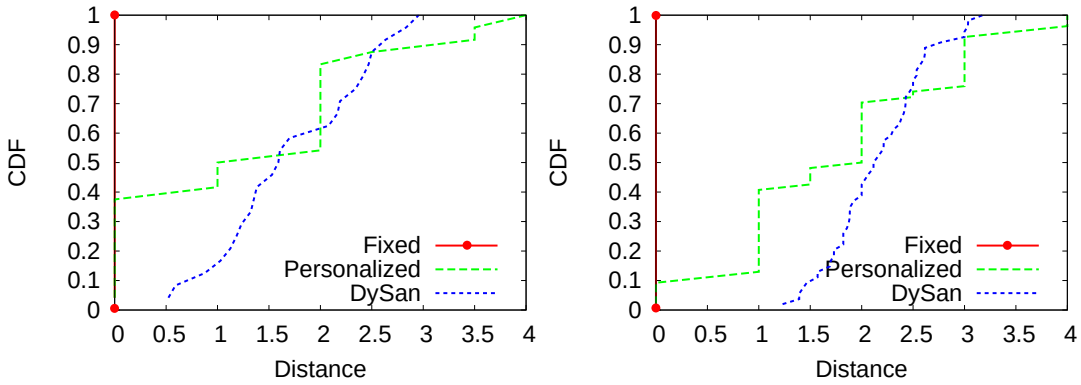
C Sanitized Data Distortion

Table 4 gives complementary results concerning the similarity analysis of the data sanitized between the different baselines, with simple quantitative measures. Here the raw measures plus the percentage relative error are given for each baselines. Even if those metrics gives few information about the shapes of the signals, we can still observe that Olympus, the only baselines that does not take into account the distortion of the data during training, is the one that have his measures very far from the raw data. For example the standard deviation is almost fives times higher than the original data showing a very noisy signal.

D Dynamic sanitizing mode selection

We evaluate the variation of the sanitizer model selection of DYSAN compared to static approaches using either one model fixed for all users or one personalized model for each user. To achieved that, we measure the distance between the hyperparameters α and λ corresponding to the best privacy and utility trade-off on average for all users (*i.e.*, the model fixed for all users) and the model selected for each user (*i.e.*, a personalized model) or according to the incoming data (*i.e.*, the model dynamically selected by DYSAN). Figure 5 reports the distribution of this distance for both datasets. Results

show that almost 40% of the users of MotionSense dataset have a personalized sanitized model which corresponds to the model providing the best trade-off on average for all users. In addition, for both datasets, results show a large variability in term of distance over all users highlighting the necessity to provide a variety of models to adapt the sanitization.



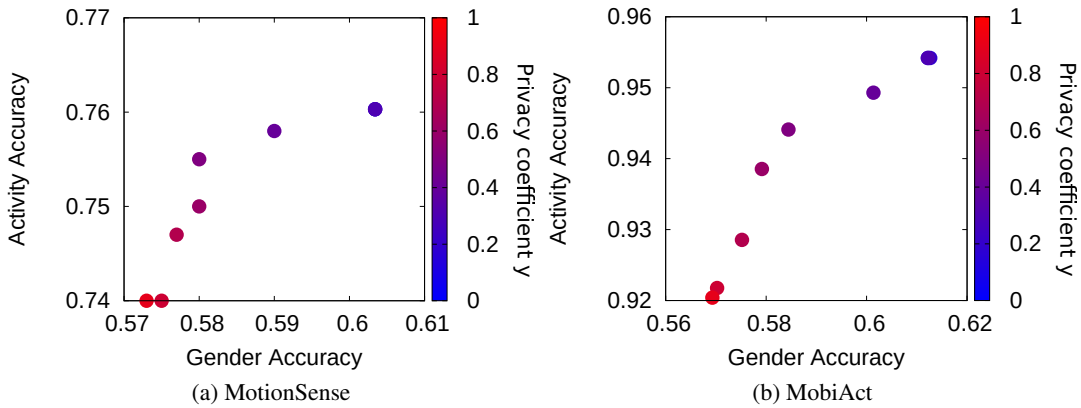
(a) MotionSense (b) MobiAct

Figure 5: DYSAN provides a large variability in terms of distance over all users highlighting the necessity to provide a variety of models to adapt the sanitization.

To complete this analysis, we also counted the number of different models used by DYSAN for each user. Figure 6 depicted for both datasets the distribution of the percentage of all possible sanitized models (36 in our experiment) selected by DYSAN for each user. Results show a large range of number of different models selected ranging from 20% to 50%. This result show that DYSAN successfully adapts the sanitization according to the evolution of the incoming data.

E Utility and privacy trade-off selection for DYSAN

As described in Section 2, the best sanitizer model is selected according to the definition of the utility and privacy trade-off defined by weight coefficients x and y. Figure 7 depicts the evolution of the utility and privacy trade-off according to x and y for both datasets.



(a) MotionSense (b) MobiAct

Figure 7: The variation of the Privacy coefficient y from 0.1 to 0.9 implies a variation of the trade-off between Utility and Privacy. For both dataset, when y increase, the Privacy increase (the gender accuracy decrease) and the Utility decrease (the activity accuracy decrease)

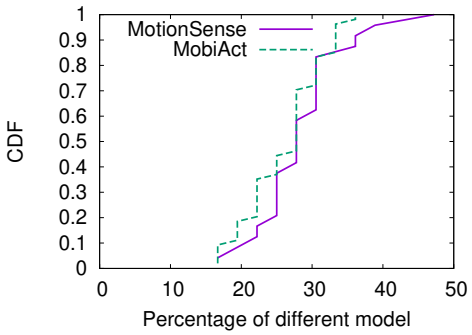


Figure 6: The data of each user is sanitized with a wide variety of models (from 20% to 50% of all the models) showing that DYSAN successfully adapts the sanitization according to the evolution of the incoming data.

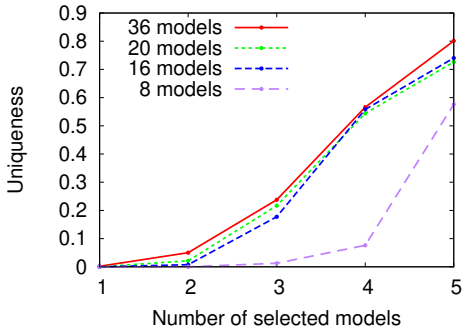


Figure 8: The uniqueness of the selected models remains low for fingerprints with less than 5 models, and depends on the number of available sanitizing models for the selection.

F Information leakage in model selection

As DYSAN dynamically selects the sanitizing model to use for each window of incoming data, the set of selected models could be leveraged to identify each user. Indeed, this set of sanitizing models chosen by a user could act as a unique fingerprint. To evaluate this potential information leakage, we quantify the uniqueness following the methodology presented in [1]. More precisely, the uniqueness for each user is estimated as the percentage of 100 random sets of p selected sanitizing models that are unique. Figure 8 reports for MobiAct dataset the distribution of the uniqueness with p (*i.e.*, the size of fingerprint) from 1 to 5 and with different number of sanitizing models available for the selection. As expected, results show that the larger the fingerprint, the more unique the behaviour of a user becomes. However, at least 5 models are needed to have a strong confidence (around 80% of uniqueness) when 36 sanitizing models are exploited. To reduce this uniqueness, a lower number of sanitizing models (i.e., through the hyperparameters values explored in the training phase) should be proposes. Indeed, less choice for model selection leads to have more users who share common models. Results show that exploiting less available sanitizing models reduces the uniquenes.

Reducing the number of sanitizing models by covering less hyperparameter values limits the achievable space for the utility and privacy trade-off. Consequently, a degradation of the accuracy for both the activity detection and the gender interference is observed. Table 1 presents the performances obtained with different number of sanitizing models available for the selection. Results show that from 36 to 20 sanitizing models, the accuracy in activity recognition decreases by only 3% and increase by 2% the gender inference.

Information leakage in model selection leading to user re-identification is only possible if the adversary is able to characterize each selected sanitizing model from the sanitized data. In this case, the adversary could maintain a fingerprint per user to conduct its re-identification attack. To evaluate this capability, we measure the level of distortion using the Dynamic Time Warping of the sanitized data for each sanitizing model. Over all sanitizing models, our results show a very low

standard deviation of the DTW. This low value indicates a small difference in terms of distortion when different sanitizing models are exploited, thus making it difficult for an adversary to identify the selected model from the sanitized data. This re-identification attack consequently seems difficult to achieve.

	Activity accuracy (%)	Gender accuracy (%)
36 models	92	57
20 models	89	59
16 models	88	63
8 models	86	66

Table 1: Reducing the number of sanitizing models available for the selection decreases the accuracy in activity recognition while increasing the accuracy in gender inference.