Towards Multiple User Active Authentication in Mobile Devices

Pramuditha Perera and Vishal M. Patel
Department of Electrical and Computer Engineering
Rutgers, The State University of New Jersey
94 Brett Road, Piscataway, NJ 08854
pramuditha.perera@rutgers.edu, vishal.m.patel@rutgers.edu

***Abstract*— Traditionally, practical authentication systems have considered only a single enrolled subject for verification. However, with the advent of mobile devices this paradigm has changed since a mobile device may be accessed by more than a single enrolled user. In this context, verification of multiple enrolled users has a practical importance. We address the issue of performance degradation associated with multiple user authentication as compared to single user authentication. We interpret this problem in an open-set framework and introduce the notion of *probability of negativity* to alleviate the effect of multiple users in authentication. We further introduce a simple fusion scheme with the existing authentication methods to increase the intruder detection accuracy. Effectiveness of the proposed method is demonstrated using three publicly available face and touch gesture-based mobile active authentication datasets.**

I. Introduction

Active Authentication (AA) has emerged as an alternative technology to traditional onetime mobile authentication mechanisms in recent years [1]. Instead of querying for a PIN number or a password, AA thrives at authenticating a user based on his/her physiological and behavioral mannerisms. Such mannerisms are often distinctive from person to person and maintains sustenance over a period of time. Furthermore, it is very difficult to mimic such mannerisms. Therefore, in principle AA systems are difficult to spoof or hack. As a result, they provide better protection against intruders. Recent works in AA have shown to achieve relatively higher accuracy rates for single user authentication [1]. However, how these results are generalized for mobile devices with multiple users have not been studied in the literature (see Figure 1). Nevertheless, the need to have mobile systems with multiple users is starting to gather attention in the industry [2], [3].

From a security perspective, the main role of AA is to detect intrusions and to prevent intruders from accessing the device. Majority of work done in AA focuses only on detecting the presence of the owner (a single user) of the device. By doing so, it is implied that the whole world except for the owner of the device is a potential intruder. In a real world scenario, this is not always the case. In practice, the owner of the device may identify a group of trusted users whom he/she is comfortable with granting access to the mobile device. For instance, different family members may want to share a mobile device. In such a context, only persons outside the family would be considered to be intruders. Similarly, in hospitals or clinics, multiple medical professionals may want to share the same mobile device without switching between users. As in the earlier case, all members of the group should be identified as trusted users of the device. Therefore, in order for AA to be deployed in the real world, it should have the capability of authenticating a group of predefined set of users with higher accuracy. Nonetheless, as shown in Figure 2(b), it is usually experienced that recognition performance degrades greatly when number of users are increased in a system. Therefore, a more elaborated approach targeting multiple-user authentication is needed for mobile AA.

In this context, three types of stakeholders of a mobile device can be identified - the owner, trusted users and intruders. Function of AA is to grant access to both owner and trusted users while preventing intruders from accessing the device. It should be noted that at the time of enrollment, the device does not possess complete knowledge about all negative (intruder) classes in the world. It is provided with complete knowledge on positive classes (owner and trusted

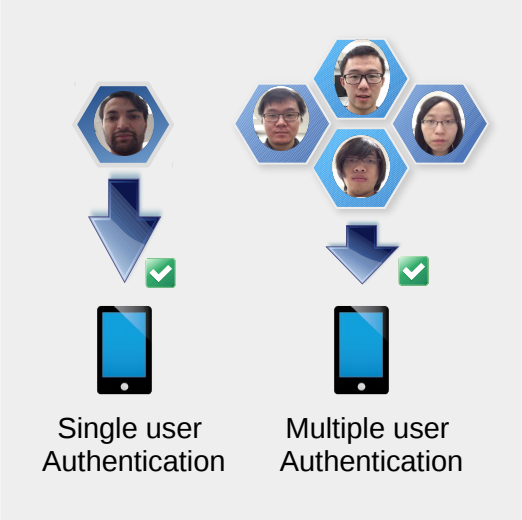


Fig. 1. Even though single-user authentication is widely studied, in reality, multiple subjects may access the same mobile device in tandem. In this paper, we study the multiple-user authentication problem.

This work was supported by US Office of Naval Research (ONR) Grant YIP N00014-16-1-3134.

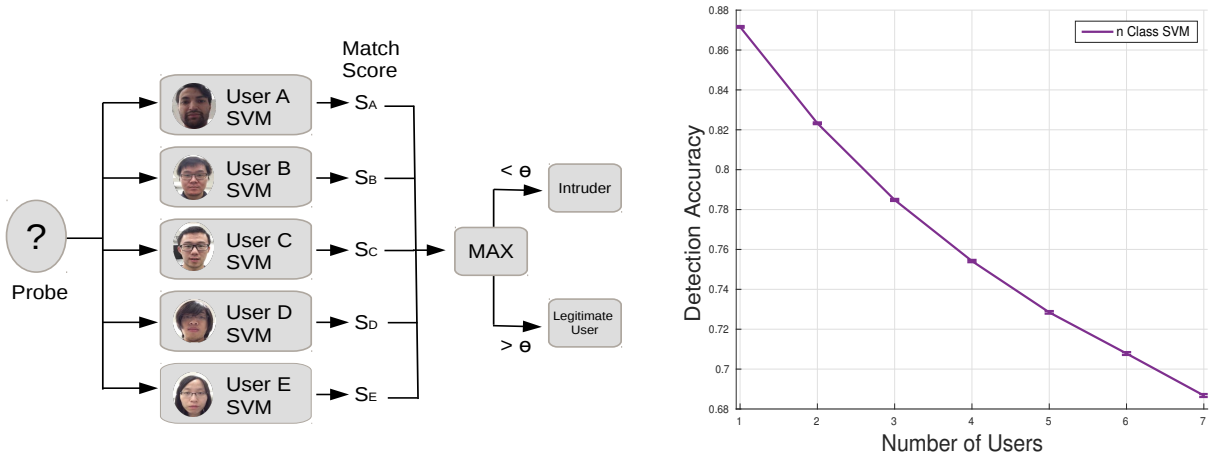


Fig. 2. Conventional n-class SVM scheme (a) A set of parallel SVM classifiers are trained for each enrolled user. Each probe is compared against each SVM and the maximum obtained score is considered. Verification is performed by thresholding the maximal obtained score. (b) Drop of performance in n-class SVM when number of users to be authenticate increases. The drop is nearly 20% when number of users are increased to seven. This experiment was conducted using the UMDAA-01 face data [4].

users) and an under-sampled knowledge of negative classes. Despite this limitation, the AA system should be able to reject any non-positive class at the time of testing. With this interpretation, AA can be formulated as an open-set detection problem as illustrated in Figure 3.

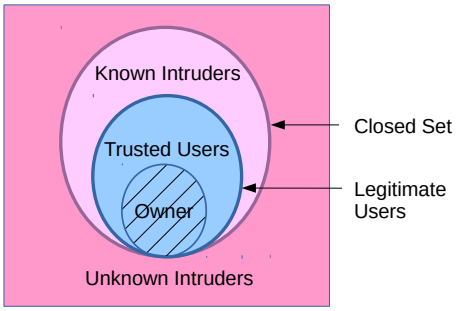


Fig. 3. The open-set problem in AA. At training time, information only about the closed set is available. At test time, negative samples outside the closed set may appear. The detector should reject all intruders (denoted in pink) while accepting all legitimate users (denoted in blue).

An open-set detector is a detector designed to detect the presence of one or more positive classes with an under-sampled knowledge of the world of the negative classes [5],[6],[7]. In other words, during the training phase, the detector has information on all positive classes and some negative classes. But, during testing, negative classes that were not present during training could occur. The goal of an open-set detector is to use training knowledge to reject both known and unknown negative classes. In AA, owner and the trusted users of the device are considered as positive classes. With this open-set interpretation, the goal would be to reject subjects outside this positive set.

In this work, we propose an initial direction to solve the multi-user AA problem. It should be noted that in practice it is possible to grant different levels of access in the device for owner and each trusted user. However, this is not the focus of our work. In the presented method, owner and trusted users would be given full access to the device without any distinction. Therefore, in what follows, we use the term legitimate users to denote both the owner and trusted users of the device.

Recognition in AA systems has been treated as a direct application of the standard recognition algorithms. Based on different sensor measurements, features and classifiers, various approaches have been previously proposed to tackle the AA problem in mobile devices. Systems based on swipe patterns [8], [9],[10], gait [11], [12] and device movement patterns [13] and behavioral patterns [14] have been previously developed for this purpose. Due to the availability of high quality front-facing cameras in modern mobile devices, face-based continuous authentication has become a possibility [15], [16], [17], [4],[18],[19]. The recently proposed facial attributes-based AA method in [20] has proven to yield the state of the art results in AA. Even though features and sensors have varied, in most of these applications, Support Vector Machines (SVM) have been used as the classifier due to its wide use, speed and better performance.

Most, if not all, previous works in AA have focused on developing verification systems for single user authentication. This is largely the case in generic recognition system design as well. Extensions for multiple-user recognition has been done by simply performing a series of single user tests in parallel (see Figure 2 (a)). Multiple user verification has been discussed in the speaker verification literature to an extent [21],[22]. However, discussion there is limited to segmenting different users. Upon segmentation, each user is verified parallely. Even though this extension is treated as trivial, as we will show in Section II, it leads to a significant performance degradation in the detection accuracy. Although specifically not aimed at AA, Cohort Analysis has been proposed in the literature [23] with the aim of normalizing inter-class

distributions to reduce this performance dip. Here, classes which are most similar to the class under consideration are used to develop a non-match distribution. Based on the non-match distribution, an inter-class normalization is performed by evaluating the log likelihood ratio of the probe. Due to the performed normalization, cohort analysis yields better performances when a fixed threshold is used over naive multiple-binary classifier schemes. However, this method does not address the primary contributor of the performance dip - issue of false alarm propagation (see Section II for more details).

On the other hand, solving open-set recognition problems have gained a lot of attention in the recent literature. The need for open-set considerations in recognition has been motivated in [5] based on SVM classifiers. Traditional SVM classifier produces a positive half space based on positive and negative samples during the training phase. However, it is possible for negative classes unseen during the training phase to occupy regions of the obtained positive half space. In this case, these classes would contribute to false positives. In order to battle this issue, in [5] 1-vs-set machine was introduced where an additional plane was introduced to restrict the positive space in the classifier. In [7], matched distribution of the classifier was modeled using a Weibull distribution to arrive at a probability of inclusion score to increase the classifier performance. This was further improved in W-SVM algorithm [6], where both matched and non-matched distributions were modeled using the Extreme Value distributions accompanied by a Compact Abating Probability (CAP) model.

In this paper, we address the multi-user AA problem as an open-set recognition problem. We introduce an algorithm based on *probability of negativity* which yields superior performances in multiple-user recognition compared to the standard single user algorithms. Furthermore, we demonstrate that decision level fusion of prevailing open-set recognition methods with the proposed method further improves the recognition accuracy.

The paper is structured as follows. In Section II we introduce issues with multiple-user authentication. In Section III probability of negativity-based algorithm is introduced and a fusion mechanism is presented. Experimental results of the proposed method on the standard mobile AA datasets are presented in Section IV. Finally, Section V concludes the paper with a brief summary.

II. Issues with Multiple-user Authentication

Let us consider the case where l number of users are enrolled in the system. Let λ denote the set of legitimate users and set ν denote negative classes encountered during training. We denote the set of people not encountered during training stage (unknown unknowns) by the set ω. With this notation, $\Omega = \lambda \cup \nu \cup \omega$, where Ω is the total population.

At the first glance, the solution to mult-user AA seems to be trivial. With the above formulation, it is possible to train l 1-vs-all classifiers, $c_1, c_2, \ldots, c_l$, to determine the presence of each user as illustrated in Figure 2(a). Then, when a probe x is presented, presence of at least one of the legitimate users can be queried simply by considering whether the maximum classifier score $max(c_1(x), c_2(x), \ldots, c_l(x))$ is greater than some threshold θ. Here, $c_i(x)$ is the score output of the i^{th} classifier for the presented probe. If not, then an intrusion can be declared.

However, as illustrated in Figure 2(b), such an arrangement would cause the detection rate to deteriorate considerably as the number of enrolled users are increased. As can be seen from Figure 2(b), the detection accuracy decreases to 68% from 87% when the number of legitimate users have been increased to seven from one. It is needless to say that such a dip in performance would be detrimental to the functionality of an AA system. This experiment was conducted using the UMDAA-01 face data [4].

It is interesting to investigate the reasons for this performance degradation. This is mainly due to two contributory reasons - detection error propagation and distribution imbalance.

A. Detection Error Propagation

In general, the detection accuracy of multi-user authentication depends on the detection accuracy of each individual classifier. There are two cases that could result in a wrong detection. The first case is when a probe belongs to a legitimate user and when the corresponding classifier produces a false negative. If the probe belongs to the i^{th} class, then the probability of such an error is $1 - \alpha_i$ where α_i is the true positive rate of the i^{th} class. The second case is when a probe belongs to an intruder and when at least one of the classifiers falsely detects it as a positive. If the false positive rate for the i^{th} class is β_i, then the probability of such an event is $1 - (1 - \beta_1) \times (1 - \beta_2) \times \cdots \times (1 - \beta_l)$ [21]. For example, if the false positive rate of each individual classifier is 0.05, then there is a probability of 0.30 for a multi-user classifier with seven users to miss classify a negative (intruder) as a positive.

Out of the two considered cases, the latter is the more dominant contributory factor. We refer to this phenomena as *false alarm propagation* henceforth. This phenomena will be more evident if one or more classifiers produce matched and non-matched distributions with significant overlaps. If the effect of true positives are ignored, then the probability of the latter can be reduced simply by introducing Bonferroni correction [24]. However, application of such a solution will cause the true positive rate to reduce.

B. Distribution Imbalance

On the other hand, the trivial solution introduced above uses a fixed threshold to determine the presence of an intruder. In reality, it would be difficult to determine a common threshold with the same effectiveness for all classifiers unless the score distributions produced by each classifier is homogeneous. Score distributions produced by two SVM classifiers of two different users are shown in Figure 4. As evident from Figure 4, the spread and overlap of matched

and non-matched distributions are greatly different for the two classifiers for real cases.

In this paper, we seek a solution that would compensate both of these issues.

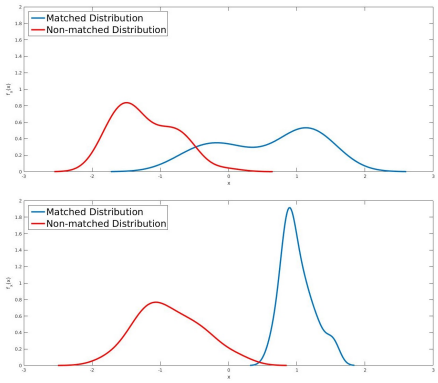


Fig. 4. Distribution Imbalance of SVM classifier scores. Shown here are matched (Blue) and non-matched (Red) distributions obtained for two SVM classifiers. Degree of overlap and range of distributions vary greatly across classifiers.

III. Probability of Negativity

In this section, we introduce our framework to battle the aforementioned contributory factors. If the observations of each individual are independent, then it is possible to train l independent classifiers, $c_1, c_2, \ldots, c_l$, based on the enrolled data. For each classifier c_i, let $n(.)$ and $m(.)$ be the non-matched and matched distributions, respectively.

Then, for a given probe x, $n(x)$ and $m(x)$ represent the probability of the probe been from a known negative class and from a known positive class, respectively. We introduce a quantity *probability of negativity* N, which measures how likely a given probe is to be from a known negative class. Quantity N based on the i^{th} classifier is define as

$$N_i(x) = P(x \in \nu | f_i) = \frac{n(x)}{n(x) + m(x)}. \tag{1}$$

Taking the open-set rejection arguments into consideration, the above definition can be extended to obtain the probability of been from a negative class as

$$N_i(x) = \begin{cases} 1, & \text{if } x < floor(n(.)) \\ 1, & \text{if } x > ceil(m(.)) \\ 0, & \text{if } ceil(n(.)) > x > ceil(m(.)) \\ \frac{n(x)}{n(x)+m(x)}, & \text{Otherwise,} \end{cases} \tag{2}$$

where $ceil(t)$ and $floor(t)$ denote the upper and lower bounds of the distribution t, respectively. This extension simply takes into account whether the considered probe falls outside the bounds of the known probability distributions; i.e. whether it belongs to an open set class. Since open set classes are inherently negative in this context, probability of been from a negative class is set equal to 1.

Based on the independent assumption, the joint probability of a probe from been a negative class based on all classifiers can be evaluated as

$$N_{\nu,\lambda}(x) = P(x \text{ being a intruder} | c_1, c_2, c_3, \ldots, c_l) \tag{3}$$

$$= N_1 \times N_2 \times \cdots \times N_l. \tag{4}$$

Since, a probe belonging to a negative class and belonging to a positive class are collectively exhaustive, given a probe, the event of the probe belonging to a positive class can be identified using the following rule

$$P_n(x) = \begin{cases} 1\ , & \text{if } 1 - N_{\nu,\lambda} > \delta \\ 0, & \text{Otherwise.} \end{cases} \tag{5}$$

This approach takes into account match and non-match distributions and their range as well as overlaps when deciding whether there is sufficient evidence to suggest that a given probe is from a known positive distribution. A probe will be recognized as a known positive only if there is a strong evidence to its claim. An intuitive explanation why P_n SVM introduces an overall gain in terms of the detection accuracy is explained in Figure 5 through an illustrative example. It should be noted that by definition N_i is invariant to different degrees of overlaps between matched and non-matched distributions. Therefore, it acts as a reasonably robust statistic across different classifiers.

A. Decision Level Fusion

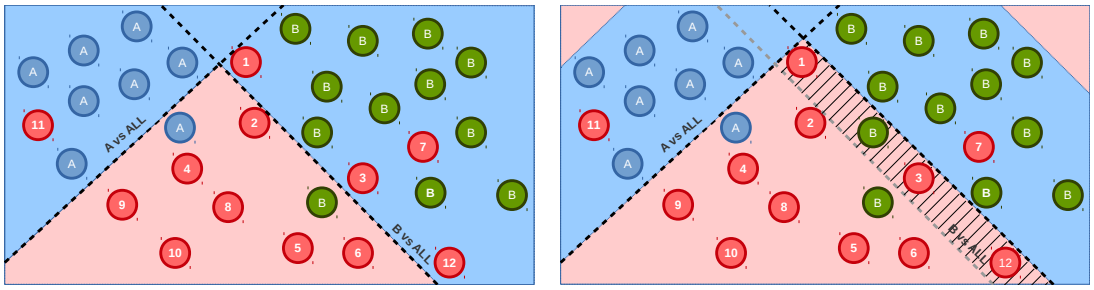
One of the limitations of the proposed probability of negativity-based mechanism is its crude definition of open set boundaries. According to the proposed framework above, bounds of open-set region are simply defined using the extreme samples of matched and non-matched distributions of the classifier. However, this hard definition of open-set may not generalize well for all cases. For example, it is possible for the sampled maximum matched value $ceil(m(.))$ to be extremely high due to an anomaly. In this case, proposed method would erroneously identify many negative test probes as positive occurrences, thereby generating more false positives. In order to alleviate this issue, we propose a decision level fusion of the standard SVM decisions with probability of negativity decision using the binary AND operator [25],

$$MUSVM(x) = SVM(x) \text{ AND } P_n(x). \tag{6}$$

In other words, probe x is declared to be from a positive class only when both SVM and P_n classifiers classify it as a positive. In addition, we consider W-SVM, which is specifically designed to reject open-set samples for decision level fusion process as follows

$$MUWSVM(x) = WSVM(x) \text{ AND } P_n(x). \tag{7}$$

It should be noted that any decision making system that incorporates open-set rejection can be used for this purpose instead. As shown in the results section, when WSVM yields better results than SVM, MUWSVM demonstrates significantly better performances compared to MUSVM. Similarly, if an alternative decision scheme with a higher accuracy is used for fusion, then the corresponding fusion scheme is expected to yield comparatively better results.



(a) (b)

Fig. 5. Illustration of why P_n works better in multi-user authentication. Here, letters correspond to positive classes and numbers correspond to negative classes. Shown in (a) is a 2-class SVM setup. Note that SVM classifier A is bound to yield a very high accuracy as compared to SVM classifier B. However, when both A and B are enrolled, subjects 1,3,7, 11 and 12 result in false accepts. Evaluation of P_n reveals that shaded region in (b) is not conclusive enough to be treated as a positive region. Hence decision boundary is further shifted as shown in (b). As a result, 1,3 and 12 are now correctly classified as negative. In the process one subject of the B class is misclassified. However, there is an overall gain in the detection accuracy.

IV. Experimental Results

In order to study the performance of multi-user authentication across different methods, experiments were conducted using three publicly available AA datasets. In each dataset, a common protocol was followed for testing. For a single trial where n number of legitimate users were considered, seven users were randomly selected to simulate legitimate users. Out of the remaining users, 10 users were randomly selected to form the negative class for training (known intruders). The remaining users were considered as unknown intruders. Testing was performed using legitimate users and unknown intruders. For a given n, this process was conducted using all possible ${}^{7}_{n}C$ combinations. In our tests, 50 such trials were conducted in order to arrive at a generalized result.

In our experiments, the performance of eight methods were compared for each dataset.

1) n class SVM : Implemented using LibSVM as described in Section II.
2) PISVM [7] : Used the published code for [7][1].
3) WSVM [6]: Used the published code for [6][2].
4) Calibrated SVM
5) P_n SVM : Implemented using LibSVM as described in equation (5).
6) MUSVM: n-class SVM fused with P_n SVM
7) MUWSVM: WSVM fused with P_n SVM
8) MU-Calibrated SVM: Calibrated SVM fused with P_n SVM

When a tunable parameter was present in a method, the best suited parameter was first sought using a validation set prior to the tests. In calibrated SVM, the matched and non-matched distributions of SVM scores were taken into consideration to perform calibration. Here, a threshold δ was selected in anticipation to automatically pick the best operating point based on the distributions obtained in training to maximize the difference between true positives and false positives as

$$\delta = \arg\max_{x} |F_m(x) - F_n(x)|, \tag{8}$$

[1]https://github.com/ljain2/libsvm-openset
[2]https://github.com/ljain2/libsvm-openset

where $F_m(.)$ and $F_n(.)$ are the cumulative distribution functions of matched and non-matched distributions, respectively. Detection accuracy is used to measure the performance of different methods. It is defined as

$$\text{Detection Accuracy} = \frac{TP + TN}{2}, \tag{9}$$

where TP and TN are the true positive rate and true negative rate, respectively. In all tests, equal number of positive and negative test samples were used in order to prevent accuracy paradox.

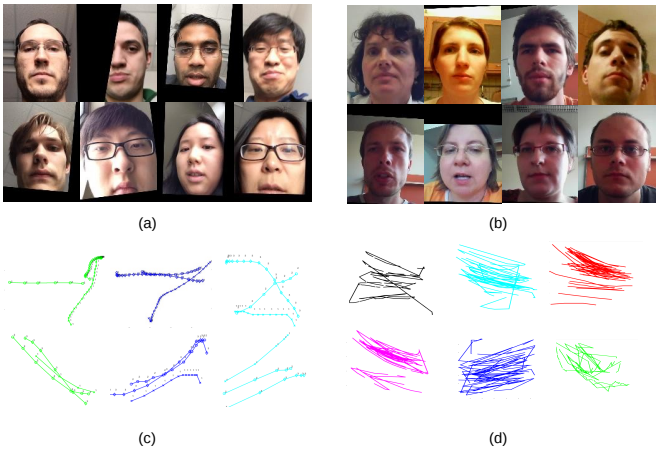


Fig. 6. Sample data from (a) UMDAA-01 face dataset, (b) MOBIO face dataset, and (c) UMDAA-01 touch gesture dataset (d) Touchalytics touch gesture dataset.

A. UMDAA-01 Dataset

The UMDAA-01 dataset [4] consists of sensor observations of 50 individuals taken from an iPhone 5 device across three sessions performing five tasks including an enrollment task. Both face information as well as touch gestures are simultaneously captured in this dataset. Sample detected face images and touch gestures from this dataset are shown in Figure 6(a) and (c), respectively. As suggested in [4], enrollment data was used as gallery and the data

from the other sessions was used as probes.

Results on the Face Data: Face images of each individual were detected using the Viola-Jones face detector [26]. Detected face images were cropped, aligned and normalized. The method presented in [20] was used to obtain the attribute features of each detected face image. Similarity of two images was evaluated using an SVM with an RBF kernel using the attribute features.

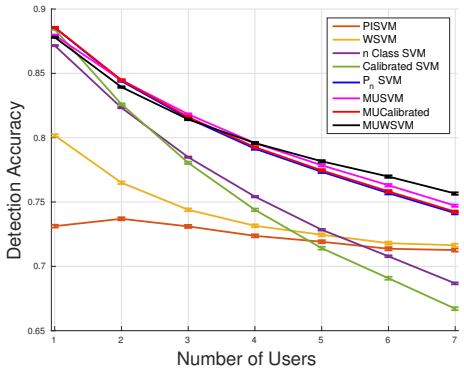


Fig. 7. Variation of detection accuracy with number of enrolled users for UMDAA-01 face dataset.

Figure 7 and Table I show the detection results corresponding to different methods as we vary the number of enrolled users. As can be seen from Figure 7, the performance of n-class SVM is on par with the other methods for a single user. But as the number of users increase, the performance degrades greatly. In comparison, PISVM and WSVM, which are specifically designed for open-set recognition, seem to yield comparatively poor performances when only a few users are present. However, both methods seem robust to the increase in the number of users. Comparatively, P_nSVM outperforms all before-mentioned methods. Compared to n-class SVM, improvement induced by P_nSVM is just 0.1% in the case of a single user. However, for seven users improvement increases up to 5%. As expected, the fusion of P_nSVM with the other forms of SVM produces the best results. In the UMDAA-01 dataset, improvement of MU-Calibrated SVM over n-class SVM is about 7% for the case of seven users.

Results on the Touch Data: From each swipe data, a 27-dimensional feature vector is extracted using the method described in [8]. A binary SVM with the RBF kernel was used to generate matching scores. Shown in Figure 8 is the variation of detection accuracy with different number of enrolled users for each method considered. Observations of Figure 8 are also summarized in Table II.

Similar to the face data, the performance is greatly degraded for n-class SVM when the number of enrolled users are increased. But, again P_nSVM outperforms n-class SVM. However, it appears for this dataset that both WSVM and PISVM are at least 5% worse than n-class SVM and the calibrated SVM for all considered cases. But surprisingly, WSVM fused with P_n SVM produces the best results for

Mean Accuracy	Number of Users						
	1	2	3	4	5	6	7
PISVM	0.73	0.74	0.73	0.72	0.72	0.71	0.71
WSVM	0.80	0.77	0.74	0.73	0.72	0.72	0.72
n-Class SVM	0.87	0.82	0.78	0.75	0.73	0.71	0.69
Calibrated SVM	0.88	0.83	0.78	0.74	0.71	0.69	0.67
P_n SVM	0.89	0.84	0.82	0.79	0.77	0.76	0.74
MUSVM	0.88	0.84	**0.82**	**0.80**	0.78	0.76	0.75
MU-Calibrated	**0.89**	**0.84**	0.82	0.79	0.77	0.76	0.74
MUWSVM	0.88	0.84	0.81	0.80	**0.78**	**0.77**	**0.76**

TABLE I

VARIATION OF DETECTION ACCURACY FOR DIFFERENT NUMBER OF USERS IN THE UMDAA-01 FACE DATASET.

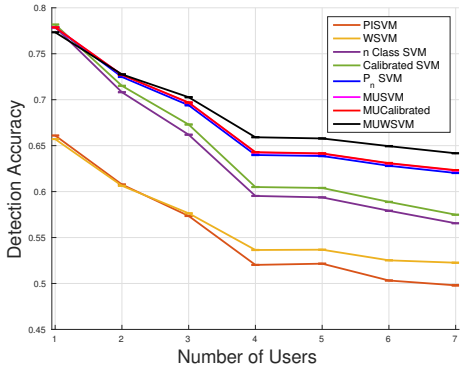


Fig. 8. Variation of detection accuracy with number of enrolled users for the UMDAA-01 touch dataset.

Mean Accuracy	Number of Users						
	1	2	3	4	5	6	7
PISVM	0.66	0.61	0.57	0.52	0.52	0.50	0.50
WSVM	0.66	0.61	0.58	0.54	0.54	0.52	0.52
n-Class SVM	**0.78**	0.71	0.66	0.60	0.59	0.58	0.57
Calibrated SVM	0.78	0.72	0.67	0.60	0.60	0.59	0.57
P_n SVM	0.78	0.72	0.69	0.64	0.64	0.63	0.62
MUSVM	0.78	0.73	0.70	0.64	0.64	0.63	0.62
MU-Calibrated	0.78	0.73	0.70	0.64	0.64	0.63	0.62
MUWSVM	0.77	**0.73**	**0.70**	**0.66**	**0.66**	**0.65**	**0.64**

TABLE II

VARIATION OF DETECTION ACCURACY FOR DIFFERENT NUMBER OF USERS IN UMDAA-01 TOUCH DATASET.

the majority of cases - in all cases where more than one user is present. This happens because open-set rejection is incorporated into P_nSVM by fusing it with WSVM.

B. Touchalytics Dataset

The Touchalitics dataset [8] contains touch data of 37 users collected across 7 tasks. Sample touch gestures from this data set are shown in Figure 6(d). 31-D touch gesture features are extracted using the method described in [8] from each gesture. A single class SVM with the RBF kernel was used to generate match and non-match scores. Detection accuracies for different number of users are shown in Figure 9 and Table III. The highest accuracy for a single user case was recorded by the calibrated-SVM. However, this figure drops by nearly 20% as the number of users are increased to seven. A similar trend can be seen from the n-class SVM curve. In majority of the considered cases, MUWSVM yields the best detection accuracies. When the number of users are seven, it is 12% better than n-class SVM. It should be noted that even though fusion results are superior, WSVM has not yielded very high accuracies in this dataset.

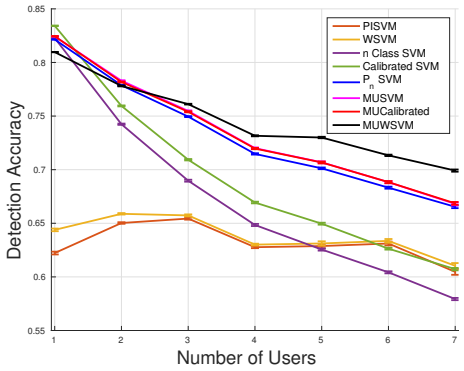


Fig. 9. Variation of detection accuracy with number of enrolled users for the Touchalytics touch dataset.

Mean Accuracy	Number of Users 1	2	3	4	5	6	7
PISVM	0.62	0.65	0.65	0.63	0.63	0.63	0.60
WSVM	0.64	0.66	0.66	0.63	0.63	0.63	0.61
n-Class SVM	0.82	0.74	0.69	0.65	0.63	0.60	0.58
Calibrated SVM	**0.83**	0.76	0.71	0.67	0.65	0.63	0.61
P_n SVM	0.82	0.78	0.75	0.71	0.70	0.68	0.67
MUSVM	0.82	**0.78**	0.75	0.72	0.71	0.69	0.67
MU-Calibrated	0.82	0.78	0.75	0.72	0.71	0.69	0.67
MUWSVM	0.81	0.78	**0.76**	**0.73**	**0.73**	**0.71**	**0.70**

TABLE III

VARIATION OF DETECTION ACCURACY FOR DIFFERENT NUMBER OF USERS IN THE TOUCHALYTICS TOUCH DATASET.

C. Mobio Dataset

The MOBIO dataset [16] contains videos of 152 subjects taken across two phases where each phase consists of six sessions each (See Figure 6(b)). Videos in this dataset are acquired using a standard 2008 Macbook laptop computer and a NOKIA N93i mobile phone. For tests on the MOBIO dataset, only videos captured using the mobile device were used. Facial attributes of each video frame were extracted using the method proposed in [20] and an SVM with an RBF kernel was used for pairwise matching.

Variation of the detection accuracy for different number of users for all considered methods are illustrated in Figure 10 and Table IV. Compared to the other considered datasets, MOBIO face data set produces higher accuracies because most of the images are captured in complete front views with a low degree of clipping. Therefore, detection accuracy of the one-user case is significantly high in this case (94.3% for n-class SVM) due to high separation between match and non-match distributions. However, as the number of classes increase to seven, same as in other cases, performance degrades by nearly 18%. Due to the low separation between distributions, improvement introduced by P_n SVM is not significant when the number of users increase. On the other hand, compared with n-Class SVM, open-set verification methods [7],[6] has performed relatively poorly as compared to the other tests. Therefore, in the WUSVM method, where P_n SVM is fused with WSVM, the improvement is not as significant as in the earlier tests. In this test, MUSVM where P_n is fused with n-Class SVM, produces the best results for all considered cases. When seven users are present, MUSVM performs 3% better than the conventional n-SVM method.

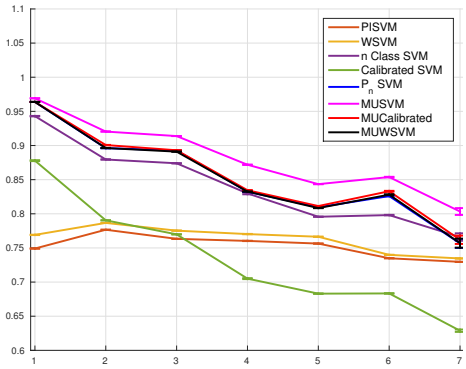


Fig. 10. Variation of detection accuracy with number of enrolled users for the Mobio face dataset.

In all considered experiments, there was a significant degradation of performance when the number of users was increased. In general, P_n SVM was able to rectify this effect to a certain degree. Open-set verification methods PISVM and WSVM did not have high detection accuracies compared to n-class SVM. But, since these methods have high capabilities in rejecting open-set probes, when P_n SVM was fused with these open-set methods, better detection results were obtained in general. The only exception was when WSM was poor at rejecting open-set samples in the MOBIO dataset. In this particular instance, MUSVM where P_n SVM was fused with n-Class SVM yielded the best results. In all considered cases, an improvement in the range of 3-12% was achieved by employing the proposed methods

Mean Accuracy	Number of Users						
	1	2	3	4	5	6	7
PISVM	0.75	0.78	0.76	0.76	0.76	0.73	0.73
WSVM	0.77	0.79	0.78	0.77	0.77	0.74	0.73
n-Class SVM	0.94	0.88	0.87	0.83	0.80	0.80	0.77
Calibrated SVM	0.87	0.79	0.77	0.71	0.68	0.68	0.63
P_n SVM	0.96	0.90	0.89	0.83	0.81	0.83	0.76
MUSVM	**0.97**	**0.92**	**0.91**	**0.87**	**0.84**	**0.85**	**0.80**
MU-Calibrated	0.96	0.90	0.89	0.83	0.81	0.83	0.76
MUWSVM	0.96	0.90	0.89	0.83	0.81	0.83	0.76

TABLE IV

VARIATION OF DETECTION ACCURACY FOR DIFFERENT NUMBER OF USERS IN THE MOBIO FACE DATASET.

when the number of phone users were selected to be seven.

However, it seems that the performance of the fusion scheme highly depends on the second mechanism used for fusing. When this secondary method lacks open-set rejection functionality, fusion results are seen to be modest. This can be rectified by simply using a more robust open-set rejection tool.

V. CONCLUSION

We introduced the problem of multiple user verification in mobile AA. It was shown that as the number of users increase, detection accuracies corresponding to the traditional methods degrade significantly. This problem was interpreted as an open-set problem and a solution based on the *probability of negativity* was proposed. In order to further improve the detection accuracy by facilitating better open-set rejection, a simple binary fusion scheme was introduced. The proposed method was tested on face and touch gesture-based AA datasets and it was shown that the proposed methods produce better performance in terms of detection accuracy as the number of users increase.

When carrying out this work, providing different functionalities for each legitimate user was not considered. All users were given the same access level instead. In the future, we will study how differential access levels would affect verification in the presence of multiple legitimate users.

REFERENCES

[1] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, July 2016.

[2] (2016) Multi-user for shared ios devices. [Online]. Available: https://www.mobileiron.com/en/datasheet/multi-user-shared-ios-devices-datasheet

[3] (2015) On multi-user access to your ios device. [Online]. Available: http://www.loopinsight.com/2015/01/19/on-multi-user-access-to-your-ios-device/

[4] M. E. Fathy, V. M. Patel, and R. Chellappa, "Face-based active authentication on mobile devices," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015.

[5] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult, "Towards open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 36, July 2013.

[6] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 36, November 2014.

[7] L. P. Jain, W. J. Scheirer, and T. E. Boult, "Multi-class open set recognition using probability of inclusion," in *The European Conference on Computer Vision (ECCV)*, September 2014.

[8] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136–148, Jan 2013.

[9] A. Serwadda, V. Phoha, and Z. Wang, "Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Sept 2013, pp. 1–8.

[10] R. Kumar and V. Phoha, "Continuous authentication of smartphone users by fusing typing, swiping and phone movement patterns," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Sept 2016.

[11] Y. Zhong and Y. Deng, "Sensor orientation invariant mobile gait biometrics," in *IEEE International Joint Conference on Biometrics*, Sept 2014, pp. 1–8.

[12] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides, "Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Sept 2012, pp. 8–15.

[13] Z. Sitova, J. Sedenka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani, "HMOG: A new biometric modality for continuous authentication of smartphone users," *CoRR*, vol. abs/1501.01199, 2015.

[14] R. G. L. Fridman, S. Weber and M. Kam, "Active authentication on mobile devices via stylometry, gps location, web browsing behavior, and application usage patterns," *IEEE Systems Journal*, 2015.

[15] A. Hadid, J. Heikkila, O. Silven, and M. Pietikainen, "Face and eye detection for person authentication in mobile phones," in *ACM/IEEE International Conference on Distributed Smart Cameras*, Sept 2007, pp. 101–108.

[16] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, D. Matrouf, J.-F. Bonastre, P. Tresadern, and T. Cootes, "Bi-modal person recognition on a mobile phone: Using mobile phone data," in *IEEE International Conference on Multimedia and Expo Workshops*, July 2012, pp. 635–640.

[17] D. Crouse, H. Han, D. Chandra, B. Barbello, and A. K. Jain, "Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data," in *International Conference on Biometrics*, 2015.

[18] U. Mahbub, S. Sakar, V. Patel, and R. Chellappa, "Active authentication for smartphones: A challenge data set and benchmark results," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, Sept 2016.

[19] P. Perera and V. M. Patel, "Quickest intrusion detection in mobile active user authentication," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2016.

[20] P. Samangouei, V. M. Patel, and R. Chellappa, "Attribute-based continuous user authentication on mobile devices," in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2015.

[21] A. Martin and M. Przybocki, "Speaker recognition in a multi-speaker environment," in *INTERSPEECH*, 2001.

[22] R. B. Dunn, D. A. Reynolds, and T. F. Quatieri, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 10, no. 1-3, pp. 93–112, 2000.

[23] G. Aggarwal, N. Ratha, R. Bolle, and R. Chellappa, "Multi-biometric cohort analysis for biometric fusion," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008.

[24] J. McDonald, *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, 2014.

[25] A. Hicklin, B. Ulery, and C. . Watson, "A brief introduction to biometric fusion," NIST, Tech. Rep., Sep 2006.

[26] P. Viola and M. J. Jones, "Robust real-time face detection," in *International journal of computer vision (IJCV)*, vol. 57, 2004, pp. 137–154.