Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 2nd, 2008

Biometric Authentication System using the Dichotomy Model

Swati Bharati, Rumeiz Haseem, Raheel Khan, Mark Ritzmann, Alex Wong
Seidenberg School of CSIS, Pace University, White Plains, NY, 10606, USA
{sb79667w, rh14126n, rk05851w, aw80162n}@pace.edu , marksritzmann@yahoo.com

Abstract

In this paper we explore the feasibility of the use of a dichotomy model in biometric authentication systems. The dichotomy model allows a measure of discrimination that is statistically inferable. The dichotomy methodology transforms a many-class problem into a two-class problem using a distance measure between pairs of samples from the same class and between pairs of samples from two different classes. We transform multi-class feature vector data from three existing biometric systems (keystroke, stylometry, and mouse movement) into dichotomy-model data. Finally, we perform two-class dichotomy classification experiments on these data to demonstrate the feasibility of the model and report results using standard authentication system performance statistics.

1. Introduction

In this paper we deal with the task of establishing the distinctiveness of an individual in a population based on a set of biometric measurements that are unique to an individual. While the most common biometrics includes the physiological biometrics, such as face, iris, fingerprint and voiceprint, we will focus on the less common behavioral biometrics of mouse movements, stylometry, and keystroke patterns. The increasing use of biometric authentication systems has been most recently seen in corporate and public security systems, consumer electronics and point of sale (POS) applications. In addition to security, the driving force behind biometric verification has been its convenience.

In biometrics, there are two important models that can be used for establishing the individuality of a person : identification (polychotomy, one-of-many decision) and authentication or verification (dichotomy, binary decision). In identification applications, a user is identified from within a population of, say, n users (one-of-n response) and is usually considered to be the more difficult of the two problems. In contrast, authentication involves the process whereby an individual is verified as being the person he/she claims to be or not. Therefore, in authentication applications a user is either accepted or rejected, that is, the output is a binary response, yes or no. The work will focus on the biometric authentication of an individual using the dichotomy model. It has been argued that the authentication (verification) problem is more suitable than the identification model for establishing the individuality of a person, specifically in cases where the number of classes is too large to be completely observed [5], for example, the population of an entire nation.

In order to establish the distinctiveness of an individual, i.e., to authenticate that individual, we start by transforming the many-class problem into a dichotomy. This is done by using a distance measure between two samples of the same class and between samples of two different classes. The usefulness of this model lies in that it allows for the inferential classification of individuals without having to sample all the classes. Also, it provides reliability when performing classification of all classes based on the data obtained from a small sample of classes representing the whole population. Using this model, all sample pairs are categorized as either the same class or different class. Given two biometric data samples, the distance between the two samples is first computed. This distance measure is used as data to be classified as positive (intra-variation, within person or identity) or negative (inter-variation, between different people or non-identity) [5].

The main purpose of this paper is to apply the dichotomy model to the feature vector data obtained from the existing keystroke, stylometry, and mouse movement biometric systems, and statistically evaluate the results. Also, the converted dichotomy-model data will be supplied to another research team for analysis by data mining techniques [6]. The paper is organized as follows. Section 2 identifies the related areas of work on biometric identification and authentication with or without using the dichotomy model. Section 3 discusses the dichotomy model, which is at the heart of the current work. Section 4 describes the format of the feature vector data supplied by the front-end systems and that of the converted dichotomy data. Section 5 describes the biometric authentication system implemented in Java. Section 6 describes the experiments and the statistical results obtained. Finally, section 7 draws conclusions and suggests possible areas of future work.

2. Related Research

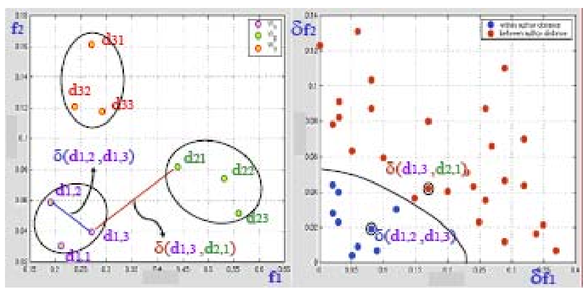
This paper utilizes the feature vector data obtained from the three existing biometric systems, namely, the Keystroke [3], the Stylometry [4] and the Mouse Movement [1] biometric systems, developed in previous years at Pace University. All three of these systems utilize the identification model in classification of the feature data. This work introduces the authentication aspect of classification, thereby extending the application of these existing systems. Implementation of the dichotomy model in biometric authentication systems is relatively new, although it has great potential and has proven to be a powerful technique in biometric authentication, as demonstrated in the work of Seung-Seok Choi *et al.* [5] and Sungsoo Yoon *et al.* [9] on iris authentication.

This research effort deals with two evolving areas in biometrics – continuous monitoring and multi-modal verification. Both of these areas are emerging in importance in the deployment of operationally viable biometric systems. Continuous verification and multi-modal verification are important for several reasons. First, there is a growing case that one-time verification (usually done during sign-in) is neither robust nor secure enough. Several weaknesses are being pointed out – weaknesses that can be addressed and largely solved by continuous verification systems. When considering a continuous system, keystroke recognition becomes an outstanding option with significant inherent benefits. Second, there is a growing case that multi-mode systems outperform single mode systems. Again, a keystroke recognition system has several important inherent benefits in this area in that it lends itself to elegant co-existence with other biometrics, and especially with the related ones of mouse movement and stylometry investigated in this study.

Once a user is verified at login, session “hijacking” can occur [8]. Biometrics, and specifically keystroke dynamics, can be the perfect answer to this challenge – they are generally passive, they do not require additional devices, and they occur in the natural course of ones work. In order to further increase the security level of a system, continuous verification can be combined with the employment of more that one biometric mode, resulting in multimodal fusion [7].

3. Dichotomy Model Transformation

In this section we show how to transform such a large polychotomy problem into a simple dichotomy problem, in which a pattern is placed into one of only two categories. In the many class problem or polychotomizer [9], consider the case where the number of classes is small and it is possible to observe many instances of each class. In such a situation, the individuality of the classes can be determined statistically by clustering the instances into classes and inferring the separation to the entire population. This is a reliable and common method in establishing the individuality when a substantial number of instances for each class are observable. Now consider the many-class problem where the number of classes is too large to be observed, for example the population of a whole country. Most of the pattern recognition problems encountered today fall under this category. As the number of classes is enormously large and almost infinite, the solution to this problem seems inconceivable. To overcome this problem, we use the dichotomy model that can handle this type of many-class problem.



(a) Feature space (Polychotomy) (b) Feature Distance space (Dichotomy)

Figure 1. Transformation from (a) Feature domain to (b) Feature distance domain

To explain the dichotomy transformation process, take an example of three people $\{P_1,P_2,P_3\}$ where each person supplies three biometric samples. Figure 1 (from [9]) plots these biometric sample data for these three people in the feature space, exemplifying the polychotomy model. To perform the transformation of this feature space into distance vector space for real valued features, the process is a follows: we first take the vector distances of each of the three features between samples of the *same* person and categorize it as a *intra-person distance*. We use the symbol $x_{\oplus}$ to denote the intra person distance. Similarly, the *inter-person distances* are calculated by measuring the distances between the biometric samples of two *different* people. This inter-person distance is denoted by the symbol $x_{\varnothing}$. Let d_{ij} represent the feature vector of the i^{th} person’s j^{th}

biometric sample, then $x_\oplus$ and $x_\varnothing$ are calculated as follows :

$x_\oplus = |d_{ij} - d_{ik}|$ where i=1 to n, and j,k=1 to m, j≠k (1)
$x_\varnothing = |d_{ij} - d_{kl}|$ where i,k=1 to n, i≠k and j,l=1 to m (2)

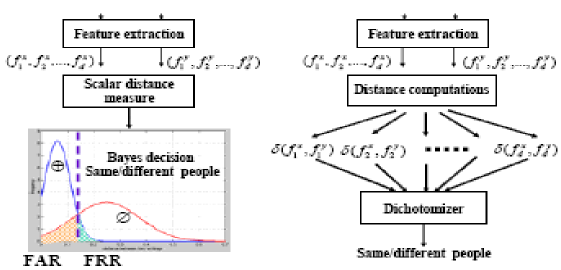
where n is the number of people, and m is the number of biometric samples per person. Note that the result of the dichotomy transform is a vector of the distances, not a scalar value.

Figure 1 (b) shows the original feature space transformed to a feature distance space. For example, an intra-person distance, W (within), and an inter-person distance, B (between), shown in Figure 1(a), corresponds to the points W and B in the feature distance plot in Figure 1(b), respectively. From this we can easily see that there are only two categories, namely, the intra-person (same) distance and the inter-person (different) distance, in the feature distance space.

The sizes of the inter- and intra-person distances were derived in [9]. If n people provide m biometric samples each, the number of intra-person and inter-person distance classes, respectively, are:

$$n_\oplus = \frac{m \times (m-1) \times n}{2} \text{ and } n_\varnothing = m \times m \times \frac{n \times (n-1)}{2}.$$

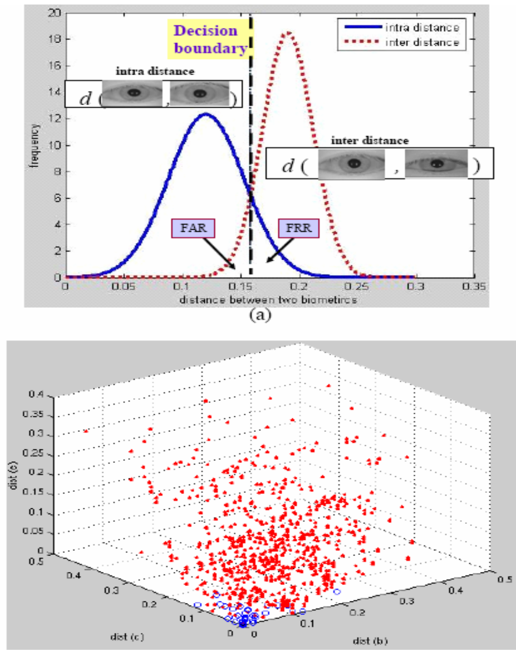
In the dichotomy model we formally state the problem as follows [9]: “given two randomly selected biometric samples, the problem is to determine whether or not those two samples are from the same person.” Figure 2 (from [9]) depicts the process of dichotomy transformation. Here let us assume that f_{ji} is the i^{th} feature of the j^{th} biometric data. First, features are obtained from both biometric data, say x and y: $\{f_1^x, f_2^x, \ldots, f_d^x\}$ and $\{f_1^y, f_2^y, \ldots, f_d^y\}$. Then each feature distance is computed: $\{\partial(f_1^x, f_1^y), \partial(f_2^x, f_2^y), \ldots, \partial(f_d^x, f_d^y)\}$, where ∂ is the absolute difference between two real values. The dichotomizer, as shown in Figure 2, takes the feature distance vector as the input and outputs the result as a binary response, i.e., “same person” or “different person”.



(a) Parametric Verification (b) Dichotomy Transformation

Figure 2. dichotomy transformation process

As shown in Figure 3(a), (from [9]), the two output distributions represent the relationship between two classes. Type I error (FRR) occurs when the same person’s biometric data are identified as coming from two different people and type II error (FAR) occurs when the biometric data provided by two different people are classified as coming from the same person [2]. The FRR is the probability of error that we classify two biometric data as different people even though they belong to the same person. On the other hand, FAR is the left-side area of the negative distribution, i.e, the probability of error that we classify two biometric data as coming from the same person even though they belong to two different people.



(b)

Figure 3. (a) Type I and II errors (b) 3-D Space Distribution.

Note that the intra-person distance distribution, as can be seen by observing Figure 3(b) (from [9]), is clustered toward the origin, while the inter-person distance distribution tends to be scattered and away from the origin. Therefore, the fact that the intra-person distance tends to be smaller is utilized when designing the dichotomizer in establishing the decision boundary between the intra and inter-person distances.

In the dichotomy model, the objective lies in the validation of the individuality of biometric data statistically, but not in the detection of the differences of specific instances. On the contrary, we are attempting to infer the individuality of the entire population based on the individuality of a sample of n people, where n is much less than the population. By definition, inferential statistics measures the reliability of individuality of the entire population based on information obtained from a sample drawn from the population. Therefore, we can justify the use of the dichotomy model as a reliable inferential statistical approach.

4. The Feature Data

Feature extraction is performed by the existing front-end systems, namely the Keystroke [3], the Stylometry [4] and the Mouse Movement [1] biometric systems. The normalized feature vector file provided by these three systems are given in the following pre-defined format, with fields in a record comma delimited and items in a field slash delimited :

- The first record contains the name and description of the file
- The second record contains the number of pattern instances
- The remaining number-of-pattern-instances records contain the following fields:
    - ID data (e.g. name/gender/date-of-birth or age)
    - person's application-related information (e.g., handedness for mouse of keystroke biometric, nationality for speech accents, etc.)
    - equipment-related information (e.g., mouse type, keyboard type, etc.)
    - task performed (e.g., copy task, free-text email task, etc.)
    - number of feature attributes / measurements
    - sequence of feature attributes / measurements

Examples of typical feature vector files are :

Mouse movement biometric data example

Mouse movement biometric data example created September 2007
4
MaryJones/F/26, left-handed, Dell mouse, fixed 10-button sequence/used right hand, 2, 0.13668, 0.53375
MaryJones/F/26, left-handed, Dell mouse, fixed 10-button sequence/used right hand, 2, 0.14378, 0.56275
JohnSmith/M/27, right-handed, optical mouse, random 10-button sequence/used right hand, 2, 0.53628, 0.43865
JohnSmith/M/27, right-handed, optical mouse, random 10-button sequence/used right hand, 2, 0.43628, 0.53865

Stylometry biometric data example

Stylometry biometric data example created September 2007
2
MaryJones/F/26, bachelors degree, Dell laptop, structured email task, 2, 0.13668, 0.53375
ChrisHill/F/02-04-1983, PhD degree, Dell desktop, free email task, 2, 0.39734, 0.92862

Keystroke biometric data example

Keystroke biometric data example created September 2007
4
JohnSmith/M/06-01-1980, right-handed, Dell laptop, email task, 2, 0.53628, 0.43865
JohnSmith/M/06-01-1980, right-handed, Dell laptop, email task, 2, 0.43628, 0.53865
JohnSmith/M/04-21-1982, left-handed, Dell desktop, copy task, 2, 0.88321, 0.43464
JohnSmith/M/04-21-1982, left-handed, Dell desktop, copy task, 2, 0.78721, 0.33262

Although, as seen from the examples, each system provides different combinations of information and the actual files hold a large number of data records, the format of the data has been standardized to facilitate the dichotomy transformation. An example of the transformed dichotomy data is given in Figure 7.

5. The Biometric Authentication System

A generic biometric authentication system was developed in Java to implement the dichotomy model. The system consists mainly of two parts: a simple standalone GUI and dichotomy transformation utilities.

5.1 The GUI

The GUI provides the end-user with a simple interface to facilitate the authentication process. The user can choose the training and testing data files to conduct an experiment, and then click on "Apply Dichotomy Model" to perform the dichotomy transformation (Figure 4).

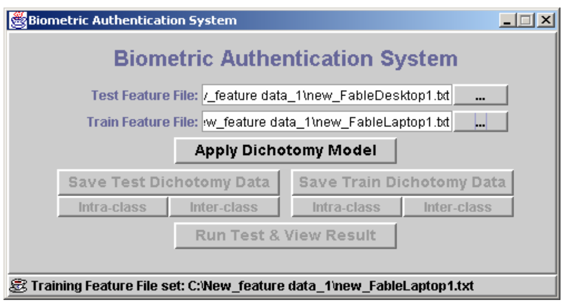


Figure 4. The GUI of the Biometric Authentication System

The user can choose the maximum number of inter or intra class samples to create, thereby allowing for a reduced set of randomly selected inter class samples being used for experimentation, as the number of inter class samples tends to be very high (Figure 5).

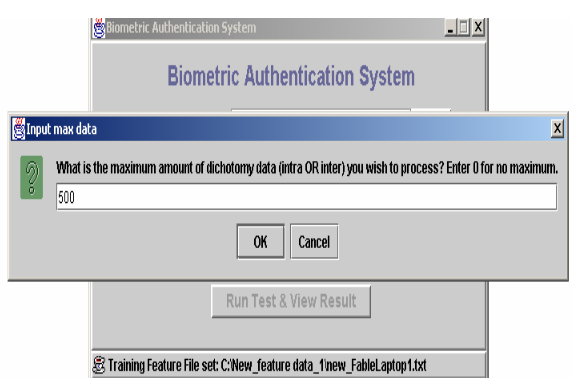


Figure 5. Choose Maximum Size of Intra – Inter Class Data

The user can then save the testing and training dichotomy data in individual files either as a combination of intra and inter class samples or as separate intra and inter class files (Figure 6). A sample dichotomy data file is shown in Figure 7.

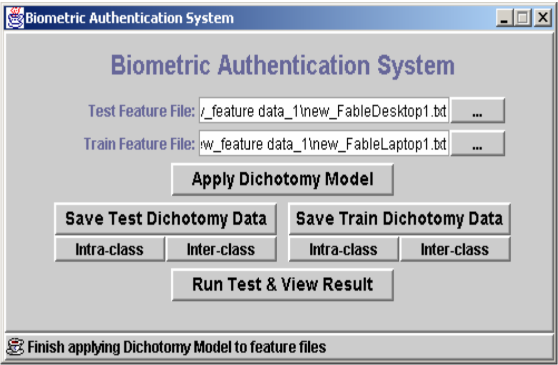


Figure 6. Save Dichotomy Data, Run Test and View Result

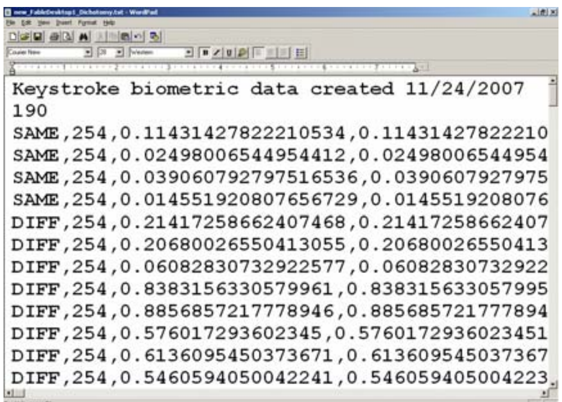

Keystroke biometric data created 11/24/2007
190
SAME,254,0.11431427822210534,0.1143142782221O
SAME,254,0.02498006544954412,0.02498006544954
SAME,254,0.03906079279751653 6,0.0390607927975
SAME,254,0.01455192080765672 9,0.0145519208076
DIFF,254,0.21417258662407468,0.21417258662407
DIFF,254,0.20680026550413055,0.20680026550413
DIFF,254,0.06082830732922577,0.06082830732922
DIFF,254,0.8383156330579961,0.838315633057995
DIFF,254,0.8856857217778946,0.885685721777894
DIFF,254,0.576017293602345,0.5760172936023451
DIFF,254,0.6136095450373671,0.613609545037367
DIFF,254,0.5460594050042241,0.546059405004223

Figure 7. Dichotomy data file

Example experimental results obtained from running the tests are shown in Figure 8, and they can be saved to a .html file for future perusal.

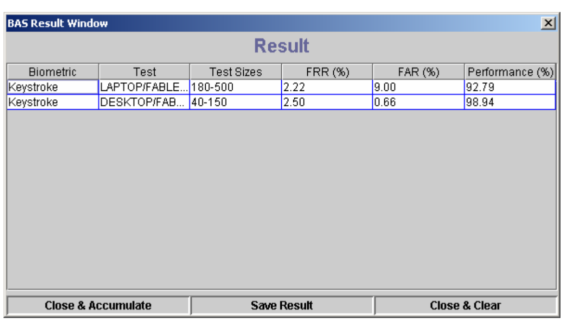


Figure 8. View Results

The system maintains a log file, which records all system activities. This log file can be viewed by clicking on the status bar of the main window, and. cleared by clicking on the “Clear Log” option at the bottom of the screen (Figure 9).

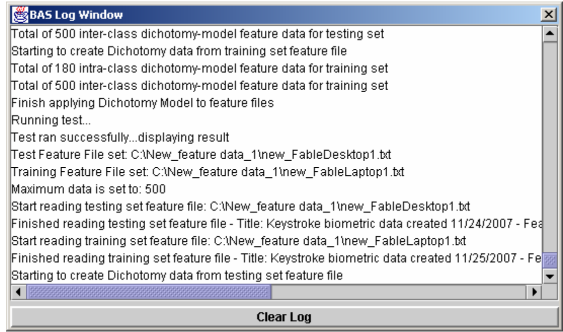


Figure 9. The log file maintained by the system, displaying all completed actions.

5.2 The Dichotomy Transformation Utilities

The Dichotomy utilities include the functionalities required for the dichotomy transformation and the verification of the individuals. We use equation (1) and equation (2) to perform the dichotomy data conversion of the original feature vector data. The verification is done using the nearest neighbor algorithm with Euclidean distance.

6. Experimental Results

This section describes the experiments performed and the results obtained. All data obtained from the three biometric systems (mouse movement, stylometry, and keystroke) were divided into training and testing sets as required by the individual experiments and then the dichotomy conversions were performed; thereafter the experiments were conducted, to obtain the results which indicated the error rates and the overall system performance.

6.1 Mouse Movement Experiments

Biometric data were obtained from the mouse movement front-end system [1]. The data consisted of 205 samples from 10 subjects: 5 subjects contributed 30 samples each, 4 subjects contributed 10 samples each, and 1 subject contributed 15 samples.

The training set comprised of 115 samples from 5 subjects, where three subjects contributed 30 samples, one subject contributed 15 samples, and one subject contributed 10 samples. The testing set consisted of the remaining 90 samples: three subjects contributed 10 samples each, and two subjects contributed 30 samples each. The number of intra and inter-class data samples generated was 1455 and 5100, respectively, for the training set and 1005 and 3000, respectively, for the testing set. Two tests were conducted, for the first trained on 5 subjects and tested on the remaining 5 subjects, and the second reversed training and testing. Table 1 shows the results obtained.

Intra-Inter class Sizes		FRR (%)	FAR (%)	Performance (%)
Train	**Test**			
1455-5100	1005-3000	64.37	22.83	66.74
1005-3000	1455-5100	58.48	23.64	68.61

Table 1. Mouse movement test results. Subjects clicked on a fixed sequence of 25 buttons.

A second set of tests were performed by creating the training set from 50 samples captured from all 5 subjects, each contributing 10 samples. The testing set was formed from another 50 samples, which were captured at an approximate three-week interval and contributed by the same 5 subjects, 10 samples each. The tests were performed twice, with the second test reversing the test and training data. The results, shown in Table 2, indicate an improvement of performance.

Intra-Inter class Sizes		FRR (%)	FAR (%)	Performance (%)
Train	**Test**			
225-1000	225-1000	72.00	16.30	73.46
225-1000	225-1000	78.66	16.30	72.24

Table 2. Mouse movement test results for tests on all subjects but training set and testing set samples captured at a three-week interval.

6.2 Stylometry Experiments

Stylometry data samples were received from the corresponding front-end project [4]. The data consisted of 120 samples, 10 samples from each of the 12 subjects. These data were divided into two equal sets of 60 samples from 6 subjects for training and testing. Therefore, the tests were conducted on different subjects. After dichotomy-model transformation, each set had 270 intra-class and 1500 inter-class samples. Two tests were conducted for these sets, with the second test reversing the training and testing data. The results of the tests are shown in Table 3. The performance of the system on these sets of data can be observed to be better in comparison to the mouse movement data shown in the previous section. However, the FRR in these tests are very high, which could be attributed to the training set not having sufficient amount of data, whereby a generalization to new users was not effectively obtainable.

Intra-Inter class Sizes		FRR (%)	FAR (%)	Performance (%)
Train	**Test**			
270-1500	270-1500	91.11	10.80	76.94
270-1500	270-1500	73.70	24.86	67.68

Table 3. Stylometry test results.

Another set of tests was conducted, this time on the same subjects but different conditions, by dividing the number of samples from each subject in half. The training set consisted of 60 samples, 5 samples each from all ten subjects. The testing set consisted of the remaining 60 samples, 5 samples from each of the ten subjects. The results are shown in Table 4. The performance increased in comparison to the previous test.

Intra-Inter class Sizes		FRR (%)	FAR (%)	Performance (%)
Train	**Test**			
120-1650	120-1650	93.33	5.27	88.75
120-1650	120-1650	85.83	10.12	84.74

Table 4. Stylometry test results: train and test on all subjects by dividing the samples.

6.3 Keystroke Experiments

In the following subsections, we describe the three different types of experiments performed on the keystroke biometric data and the results derived. A large quantity of keystroke biometric data was obtained from the corresponding front-end project [3]. The keystroke system has been developed and extended over several years and a large amount of data accrued over this period. This allowed us to perform the most extensive set of authentication experiments on different sets of these data. The data is separated as “old” data obtained over the last few years, and “new” data obtained over recent weeks. The “old” data is divided into 4 sets, each comprised of 180 samples, contributed by 36 subjects, 5 samples each, and each set varying in the type of keyboard and/or mode of entry used (Desktop/Copy, Desktop/Free, Laptop/Copy, and Laptop/Free). The “new” data is spread over 12 sets and is time dependent with each group of 4 sets collected at different times.

6.3.1 Different Subjects Same Conditions

The first sets of experiments were performed on the “old” keystroke data. For this, each of the 4 sets was equally divided into two to create training and testing sets, each set consisting of 18 different subjects, but the conditions remained the same. For example for the first test, the training set consisted of 90 samples with 18 subjects contributing 5 samples each and the testing set consisted of the remaining 90 samples contributed by the 18 other subjects, where all subjects performed on Desktop/Copy conditions. Table 5 summarizes the results obtained by performing the test on the dichotomy-converted data. The intra and inter class sizes were 180 and 3825, respectively and the tests in this case were run on all the dichotomy data without reducing the size of the inter class data. In the third test a smaller intra-inter class size is observed due to the original data being a few records short. The results show the system performance in nearly all cases above 95%.

Conditions	Intra-Inter Class Sizes		FRR (%)	FAR (%)	Performance (%)
	Train	Test			
Desktop/ Copy	180-3825	180-3825	11.11	6.01	93.75
Laptop/ Copy	180-3825	180-3825	7.77	4.36	95.48
Desktop/ Free	171-3570	176-3740	28.40	1.39	97.39
Laptop/ Free	180-3825	180-3825	15.55	3.73	95.73

Table 5. Test results for "old" keystroke biometric data on different subjects and same conditions using all intra and inter-class data samples.

The same set of tests were performed a second time, only in this case using a reduced set of randomly selected 500 inter-class data samples. The results are shown in Table 6. As can be observed, the system performance decreased in comparison to the previous tests, but the FRR was reduced and the FAR increased.

Conditions	Intra-Inter Class Sizes		FRR (%)	FAR (%)	Performance (%)
	Train	Test			
Desktop/ Copy	180-500	180-500	10.00	13.40	87.50
Laptop/ Copy	180-500	180-500	1.66	10.20	92.05
Desktop/ Free	171-500	176-500	18.75	5.00	91.42
Laptop/ Free	180-500	180-500	9.44	10.80	89.55

Table 6. Test results for "old" keystroke biometric data on different subjects and same conditions using only 500 inter-class data samples.

6.3.2 Same Subjects Different Conditions

The next sets of experiments were also performed on the "old" keystroke data. In this case, the experiments were conducted on the same subjects but for different conditions. For example, in the first experiment, one set of 180 samples was used as the training set and a second 180-sample set was used as the testing set. Therefore, the same 36 subjects were tested but on Desktop/Copy conditions as opposed to Desktop/Free conditions. The results of the experiments are shown in Table 7. The intra and inter class samples generated were 360 and 15750, respectively. However, the tests were performed on a reduced set of inter-class data samples, where a randomized set of 500 inter-class data samples was selected from the whole set and consequently used for the tests.

Conditions		Intra-Inter Class Sizes		FRR (%)	FAR (%)	Performance (%)
Train	Test	Train	Test			
Desktop/ Copy	Desktop/ Free	360-500	347-500	8.06	17.80	86.18
Desktop/ Free	Desktop/ Copy	347-500	360-500	3.33	13.00	91.04
Laptop/ Copy	Laptop/ Free	360-500	360-500	3.61	40.40	75.00
Laptop/ Free	Laptop/ Copy	360-500	360-500	5.83	3.40	95.58
Desktop/ Copy	Laptop/ Copy	360-500	360-500	5.27	6.80	93.83
Laptop/ Copy	Desktop/ Copy	360-500	360-500	4.72	18.00	87.55
Desktop/ Free	Laptop/ Free	347-500	360-500	3.05	38.80	76.16
Laptop/ Free	Desktop/ Free	360-500	347-500	8.93	6.80	92.32
Desktop/ Copy	Laptop/ Free	360-500	360-500	5.83	22.20	84.65
Laptop/ Free	Desktop/ Copy	360-500	360-500	5.27	8.79	92.67
Desktop/ Free	Laptop/ Copy	347-500	360-500	1.66	14.39	90.93
Laptop/ Copy	Desktop/ Free	360-500	347-500	3.17	27.60	82.40

Table 7. Test results for "old" keystroke data on same subjects and different conditions.

The results indicate a performance of an average of 87%, while the error rate remains low.

6.3.3 Longitudinal Authentication Same Subjects Different Conditions

The final set of experiments involved the "new" keystroke data obtained during the fall of 2007 at intervals of two and four weeks. With these data a set of longitudinal authentication experiments were conducted with training and testing performed on the data captured from the same subjects using the same type of equipment and same tasks, but at different data capture times. Table 8 shows the results of training on the baseline data set and testing on the set captured after an interval of two weeks. Similarly, Table 9 shows the results of training on the baseline data set and testing on the set captured after an interval of four weeks. System performance is above 90% for most cases, and shows essentially no degradation in performance from the shorter to the longer time interval between data collection used for training and testing.

Condition	Intra-Inter Class Sizes		FRR (%)	FAR (%)	Performance (%)
	Train	Test			
Desktop/ Copy	40-150	40-150	2.50	4.66	95.78
Laptop/ Copy	40-150	40-150	2.50	10.00	91.57
Desktop/ Free	40-150	40-150	0.00	4.66	96.31
Laptop/ Free	40-150	40-150	0.00	10.00	92.10

Table 8. Longitudinal authentication test results on same subjects and conditions but at two-week data collection interval.

Condition	Intra-Inter class Sizes		FRR (%)	FAR (%)	Performance (%)
	Train	Test			
Desktop/ Copy	40-150	40-150	2.50	12.66	89.47
Laptop/ Copy	40-150	40-150	0.00	0.00	100.00
Desktop/ Free	40-150	40-150	2.50	1.33	98.42
Laptop/ Free	40-150	40-150	0.00	8.00	93.68

Table 9. Longitudinal authentication test results on same subjects and conditions but at four-week data collection interval.

In summary, for the keystroke biometric authentication experiments, the results show a consistency in the overall performance of the system and it maintains a low error rate for all of the experiments conducted under varying conditions.

7. Conclusions

In this paper, we endeavored to utilize the dichotomy model in the authentication of biometric data obtained from the Keystroke, the Stylometry and the Mouse Movement biometric systems. We made an effort in establishing that the dichotomy model is the preferred model over the polychotomy model when dealing with an enormous number of classes where the whole population is not available for sampling, that it is the statistically inferable approach.

When dealing with a small number of users (classes), as for the mouse movement and stylometry biometric data, the system performance ranged between 66% and 76% with high error rates. However, the results on the keystroke biometric data show that the dichotomy model may be a feasible solution to the authentication problem when a large number of classes are involved. The overall system performance was observed to be above 90% in most cases for the keystroke biometric dichotomy data with corresponding low error rates, FAR less than 15% in most cases and FRR almost always less than 10%.

Future work would entail in the comparative analysis of the dichotomy authentication results obtained in this paper with polychotomy authentication results obtained on the same keystroke biometric data. Also, it would be interesting to see whether the results for the mouse movement and stylometry data improved significantly as the sample sizes increased.

8. References

[1] Nkem Ajufor, Antony Amalraj, Rafael Diaz, Mohammed Islam, Michael Lampe, "Refinement of a Mouse Movement Biometric System," Proc. CSIS Research Day, Pace Univ., May 2008.

[2] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, “Guide to Biometrics,” Springer Professional Computing, ISBN 0-387-40089-3, 2003

[3] T. Buch, A. Cotoranu, E. Jeskey, F. Tihon and M. Villani "An Enhanced Keystroke Biometric System and Associated Studies," Proc. CSIS Research Day, Pace Univ., May 2008.

[4] K. Calix, M. Connors, D. Levy, H. Manzar, G. McCabe, and S. Westcott, "Stylometry for E-mail Author Identification and Authentication," Proc. CSIS Research Day, Pace Univ., May 2008.

[5] Seung-Seok Choi, Sungsoo Yoon, Sung-Hyuk Cha and Charles C. Tappert, “Use of histogram distances in Iris Authentication”. http://www.csis.pace.edu/~ctappert/srd2004/paper06.pdf

[6] Clara Eusebi, Cosmin Gilga, Deepa John, Andre Maisonave, "A Data Mining Study of Mouse Movement, Stylometry, and Keystroke Biometric Data," Proc. CSIS Research Day, Pace Univ., May 2008.

[7] A. Ross and A. Jain “Information Fusion in Biometrics” Pattern Recognition Letters, vol 24, no. 13, pp 2115-2125, 2003

[8] T. Sim, S. Zhang, R. Janafiraman, and S. Kumar, “Continuous Verification Using Multimodal Biometrics“ IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no.4, April 2007

[9] Sungsoo Yoon, Seung-Seok Choi, Sung-Hyuk Cha, Yillbyung Lee,Charles C. Tappert, “On the Individuality of the Iris Biometric”, ICGST-GVIP Journal, Volume (5), Issue (5), May 2005. http://www.icgst.com/gvip/v5/P1150511001.pdf