

User Authentication Incorporating
Feature Level Data Fusion of
Multiple Biometric Characteristics

Mark Abernethy

This thesis is presented for the degree of

Doctor of Philosophy

Murdoch University, January 2011.

Declaration

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

.....

Mark Abernethy

Abstract

This PhD research project developed and evaluated innovative approaches to computer system user authentication, using biometric characteristics. It involved experiments with a significant number of participants and development of new approaches to biometric data representation and analysis.

The initial authentication procedure, that we all perform when we log onto a computer system, is considered to be the first line of protection for computer systems. The password is the most common verification token used in initial authentication procedures. Unfortunately, passwords are subject to numerous attack vectors (loss, theft, guessing or cracking), and as a result unauthorised persons may gain access to the verification token and be incorrectly authenticated. This has led to password-based authentication procedures being responsible for a large proportion of computer network security breaches.

In recent years, the use of biometrics has been increasingly researched as an alternative to passwords in the initial authentication procedure. Biometrics concerns the physical traits and behavioural characteristics that make each individual unique. Biometric authentication involves the use of biometric technologies in authentication systems, with the aim to provide accurate verification (based on biometric characteristics).

Research has demonstrated that uni-modal biometric authentication (that is, authentication based on a single biometric characteristic) makes it difficult for an impostor to impersonate a legitimate user. More recent research is finding that multi-modal biometric authentication (that is, authentication based on the combination of multiple biometric characteristics) can make it even more difficult for an impostor to impersonate a legitimate user. Thus multi-modal biometrics claims improved accuracy and robustness.

Multi-modal biometrics requires consideration of various aspects of data integration, known to the field of data fusion. Multi-modal biometric research has, until recently, focused on the fusion of data (from multiple sources) at the decision level or the confidence score level. It has been proposed that fusion of data at the feature level will produce more accurate and reliable verification.

However, fusion of data at the feature level is a more difficult task than fusion at the other two levels. For decision level fusion, ‘accept’ or ‘reject’ results from the different data sources are fused. For confidence score level fusion, confidence scores (typically in the continuous interval $[0, 1]$) from the different data sources are fused. That is, for the aforementioned levels, the data from the multiple sources are of the same nature. Feature level fusion combines feature vectors, where the data from the different sources are most likely to consist of different units of measurement.

Data fusion literature formally specifies that data may be combined according to three paradigms: competitive, complementary, and cooperative. Competitive data fusion assesses data from all available sources, and bases classification upon the ‘best’ source. Complementary data fusion combines all available data from all sources, and bases classification upon this combined data. Cooperative data fusion involves the selection of the best features of each individual data source, and then combines the selected features prior to classification.

The objectives of the current study were to investigate the use of two individual biometric characteristics (keystroke dynamics and fingerprint recognition). For keystroke dynamics, feature selection was employed to reduce the variability associated with data from this characteristic. For fingerprint recognition, a new method was developed to represent fingerprint features. This was done to assist classification by Artificial Neural Networks, and to meet the requirement to facilitate fusion with the keystroke dynamics data at the feature level.

Whilst feature level data fusion was the primary objective, investigation of the two individual characteristics was conducted to enable comparison of results with the data fusion results. For the data fusion investigation, the complementary and cooperative paradigms were adopted, with the cooperative approach involving four stages.

The feature selection method chosen to filter keystroke dynamics data was based on normality statistics, and returned results comparable to many other research efforts. The fingerprint feature representation method developed for this experiment demonstrated an innovative and effective technique, which could be applicable in a uni-modal or a multi-modal context.

As the new fingerprint representation method resulted in a standard length feature vector for each fingerprint, data alignment and subsequent feature level data fusion was efficiently and practicably facilitated.

The experiment recruited 90 participants to provide typing and fingerprint samples. Of these, 140 keystroke dynamics samples and 140 fingerprint samples (from each participant) were utilised for the first two phases of the experiment. Phase three of the experiment involved the fusion of the samples from the first two phases, and thus there were 140 combined samples. These quantities provided 100 samples for false negative testing and 10,500 samples for false positive testing (for each participant for each phase of the experiment). These figures are similar or better than virtually all previous research studies in this field.

The results of the three phases of the experiment were calculated as the two performance variables, the false rejection rate (FRR)—measuring the false negatives—and the false acceptance rate (FAR)—measuring the false positives.

The keystroke dynamics investigation returned an average FAR of 0.02766095 and an average FRR of 0.0862, which were at least comparable with other research in the field.

The fingerprint recognition investigation returned an average FAR of 0.0 and an average FRR of 0.0022, which were as good as (or better than) other research in the field.

The feature level data fusion adopting the complementary approach returned an average FAR of 0.0 and an average FRR of 0.0004. Feature level data fusion adopting the cooperative approach returned respective average FAR and FRR results of 0.00000381 and 0.0004 for stage 1, 0.0 and 0.0006 for stage 2, 0.0 and 0.001 for stage 3, and 0.0 and 0.001 for stage 4.

The research demonstrated that uni-modal biometric authentication systems provide an accurate alternative to traditional password-based authentication methods. Additionally, the keystroke dynamics investigation demonstrated that filtering ‘noisy’ data from raw data improved accuracy for this biometric characteristic (though other filtering methods than that used in this research may improve accuracy further). Also, the newly developed fingerprint representation method demonstrated excellent results, and indicated that its use for future research (in representing two dimensional data for classification by Artificial Neural Networks) could be advantageous.

The data fusion investigation demonstrated that multi-modal biometric authentication systems provide additional accuracy improvement (as well as a perceived robustness) compared to uni-modal biometric authentication systems. Feature level data fusion demonstrated improved accuracy compared with confidence score level and decision level data fusion methods. The new fingerprint representation method (which provided an innovative technique for representing data from any two dimensional data source) facilitated feature level data fusion with keystroke dynamic data, and the results validate the importance of using feature rich data.

Contents

1	Introduction	1
1.1	Context Of The Study	1
1.2	Motivation And Objectives For This Research	6
1.2.1	Motivation For The Study	6
1.2.2	Objectives of the Study	10
1.2.3	Research Questions	11
1.3	Significance Of The Research	12
1.4	Scope Of The Research	14
1.5	Experimental Method And The Rationale For Its Selection	16
1.6	Outline Of This Dissertation	17
1.7	Conclusion	18
2	Background	19
2.1	Introduction	19
2.2	Biometrics	20
2.2.1	Overview of Biometrics	20
2.2.2	Biometric Authentication Systems	22
2.2.3	Biometric Performance Variables and System Errors	27
2.2.4	Biometric Characteristics	30
2.2.4.1	Deoxyribonucleic Acid (DNA)	32
2.2.4.2	Facial Recognition	33
2.2.4.3	Iris Pattern Recognition	34
2.2.4.4	Retinal Pattern Recognition	35
2.2.4.5	Speaker Recognition	36

2.2.4.6	Fingerprint Recognition	37
2.2.4.7	Palmprint Recognition	37
2.2.4.8	Hand Geometry	38
2.2.4.9	Keystroke Dynamics	39
2.2.4.10	Signature Recognition	39
2.2.4.11	Gait Recognition	40
2.2.4.12	Body Odor Recognition	40
2.2.4.13	More Detailed Discussion	41
2.3	Data Fusion And Multi-Modal Biometrics	41
2.3.1	Data Fusion	41
2.3.1.1	Paradigms of Data Fusion	44
2.3.1.2	Formal Levels of Fusion	48
2.3.1.3	Data Alignment	49
2.3.2	Multi-Modal Biometrics	50
2.3.2.1	Levels of Fusion In Multi-Modal Biometrics	52
2.3.2.2	Review of Multi-Modal Biometrics Research	61
2.4	Pattern Recognition And Artificial Neural Networks	78
2.4.1	Pattern Recognition	79
2.4.1.1	Classification Schemes	81
2.4.2	Artificial Neural Networks	84
2.4.2.1	Imitating The Biological Model	85
2.4.2.2	ANN Architectures	94
2.4.3	The Multi-Layer Perceptron As A Pattern Classifier	113
2.5	Conclusion	116
3	Keystroke Dynamics	119
3.1	Introduction	119
3.2	Overview of Keystroke Dynamics	119
3.3	Metrics	121
3.4	Keystroke Dynamics Related Research	124

3.4.1	Static Verification	125
3.4.2	Dynamic Verification	146
3.5	Summary of Keystroke Dynamics	150
3.6	Conclusion	154
4	Fingerprint Recognition	155
4.1	Introduction	155
4.2	Overview of Fingerprint Recognition	156
4.2.1	The Uniqueness of Fingerprint	160
4.3	Fingerprint Features	162
4.3.1	Global Features	162
4.3.2	Local Features	164
4.4	Automated Fingerprint Identification Systems	167
4.4.1	Fingerprint Acquisition	169
4.4.1.1	Off-Line Fingerprint Acquisition	169
4.4.1.2	Latent Fingerprints	170
4.4.1.3	Live-Scan Fingerprint Acquisition	173
4.4.2	Fingerprint Representation	175
4.4.3	Pre-processing	177
4.4.4	Feature Extraction	179
4.4.5	Fingerprint Classification	183
4.4.5.1	Feature Extraction For Classification	187
4.4.5.2	Classification Techniques	188
4.4.6	Fingerprint Verification	190
4.4.6.1	Feature Extraction For Verification	192
4.4.6.2	Verification Techniques	192
4.5	Minutiae-based Matching Related Research	196
4.6	Summary Of Minutiae-Based Matching Techniques	218
4.6.1	Approach Adopted By The Reviewed Research Efforts	219
4.6.2	Approach Adopted In The Current Experiment	222
4.6.3	Rationale For The Adopted Approach	223

4.7	Conclusion	224
5	Experimental Methods	225
5.1	Introduction	225
5.2	Experimental Overview	226
5.3	Participants	227
5.4	Keystroke Dynamics	229
5.4.1	Software	229
5.4.2	Data Collection	232
5.4.3	Metrics	234
5.4.4	Pre-processing	235
5.4.4.1	Keystroke Dynamics Feature Selection	239
5.4.5	Final Analysis Procedure	247
5.4.5.1	Training Phase	248
5.4.5.2	Testing Phase	251
5.5	Fingerprint Recognition	252
5.5.1	Software	252
5.5.2	Data Collection	255
5.5.3	Fingerprint Feature Extraction	257
5.5.4	Local Feature Registration	258
5.5.4.1	Model Feature Set	261
5.5.4.2	Scene Feature Set Alignment	262
5.5.5	Feature Selection	277
5.5.6	Final Analysis Procedure	283
5.5.6.1	Training Phase	283
5.5.6.2	Testing Phase	285
5.6	Feature Level Data Fusion	286
5.6.1	Introduction	286
5.6.2	Complementary Data Fusion Approach	287
5.6.2.1	Complementary Fusion of Feature Data	288
5.6.2.2	Final Analysis Procedure	290

5.6.3	Cooperative Data Fusion Approach	291
5.6.3.1	Selection of Feature Metrics	295
5.6.3.2	Cooperative Fusion of Feature Data	305
5.6.3.3	Final Analysis Procedure	307
5.7	Experimental Validity	309
5.7.1	Internal Validity	310
5.7.2	External Validity	313
5.8	Conclusion	314
6	Research Results And Analysis Method	317
6.1	Overview	317
6.2	Classification of Authentication Outcomes	317
6.2.1	Classification Measurement	317
6.2.2	Receiver Operating Characteristics (ROC) Graphs	323
6.2.2.1	ROC space	323
6.2.2.2	Area Under The ROC Curve	328
6.2.2.3	Optimal Operating Point	330
6.3	Applying ROC In This Study	333
6.3.1	Introduction	333
6.3.2	Calculation of ROC Operating Points	335
6.3.3	Calculation of The Area Under The ROC Curve	337
6.3.4	Calculation of Decision Threshold	338
6.4	Results	343
6.4.1	Keystroke Dynamics (Phase 1)	345
6.4.2	Fingerprint Recognition (Phase 2)	351
6.4.3	Data Fusion (Phase 3)	353
6.4.3.1	Complementary Data Fusion	353
6.4.3.2	Cooperative Data Fusion	355
6.5	Conclusion	363

7 Discussion Of Results	367
7.1 Introduction	367
7.2 Discussion	367
7.2.1 Discussion Of Keystroke Dynamics Results	369
7.2.1.1 Summary of Keystroke Dynamics Results	387
7.2.2 Discussion Of Fingerprint Recognition Results	389
7.2.2.1 Summary of Fingerprint Recognition Results	401
7.2.3 Discussion Of Data Fusion Results	404
7.2.3.1 Complementary Data Fusion	404
7.2.3.2 Summary of Complementary Data Fusion Results	413
7.2.3.3 Cooperative Data Fusion	414
7.2.3.4 Summary of Cooperative Data Fusion Results	432
7.3 Conclusion	434
8 Conclusion	437
8.1 Research Purpose and Objectives	437
8.2 Main Contribution of the Research	441
8.3 Limitations of the Research	444
8.4 Implications and Practical Application of the Research	448
8.5 Future Research Directions	451
8.6 Final Remarks	453
Appendix A	455
A.1 Reported Security Breaches And Vulnerabilities	456
Appendix B	461
B.1 Keystroke Dynamics Metrics Selection Worked Example	461
Appendix C	467
C.1 Keystroke Dynamics Phase Software	467
C.1.1 Pre-processing	468
C.1.2 Experimental Procedure	469
C.2 Fingerprint Recognition Phase Software	472

C.2.1	Pre-processing	473
C.2.2	Experimental Procedure	474
C.3	Data Fusion Phase Software	476
C.3.1	Complementary Data Fusion Software	477
C.3.2	Cooperative Data Fusion Software	479
Appendix D		483
D.1	Keystroke Dynamics ROC Examples	483

List of Tables

2.1	Summary of Biometric Characteristics for Authentication Systems . .	30
2.2	Summary of Reviewed Literature Involving Multi-Modal Biometrics . .	62
3.1	Metric Calculation for a Two-Key Combination	124
3.2	Summary of Reviewed Literature Involving Static Verification	126
3.3	Summary of Reviewed Literature Involving Dynamic Verification . . .	147
4.1	FBI Latent Fingerprint Collection Procedures	171
4.2	Correlation of Early Fingerprint Classes	184
4.3	Proportion of Fingerprint Classes	186
4.4	Fingerprint Classes and Their Singular Points	189
4.5	Summary of Reviewed Literature Involving Minutiae-Based Matching	198
4.6	Performance Metrics Experiment by He et al., 2003	209
4.7	Performance Metrics Experiment by Tong et al., 2005	211
5.1	Priority Ratings	241
5.2	Indices Of Selected Metrics For Participant One	244
5.3	Indices Of Selected Metrics For Participant Three	244
5.4	Example of Global and Selected Metrics for a Participants Input File	246
5.5	Example Registration Tables	261
5.6	Local Area Alignment Coordinates	265
5.7	Boundary Limits For Candidate Transformation Factors	272
5.8	Global Adjustment Ranges	274
5.9	Example Output From '.tab' File	281
5.10	Participants with Unmatched Features After Selection	282

5.11	Approximate Relative Local Gain for Keystroke Dynamics	300
5.12	Average Local Gain Proportions	301
5.13	Number of Metrics Per Percentage	303
6.1	AUC Statistic Descriptions	328
6.3	Comparison Between AUC and TPMean for Keystroke Dynamics . .	340
6.4	Confidence Levels	341
6.5	Keystroke Dynamics Statistics for Threshold Calculation	345
6.6	Keystroke Dynamics Results	346
6.7	Fingerprint Recognition Statistics for Threshold Calculation	351
6.8	Fingerprint Recognition Results	352
6.9	Complementary Data Fusion Statistics for Threshold Calculation . .	353
6.10	Complementary Data Fusion Results	354
6.11	Cooperative Data Fusion (40%) Statistics for Threshold Calculation .	355
6.12	Cooperative Data Fusion (40%) Results	356
6.13	Cooperative Data Fusion (50%) Statistics for Threshold Calculation .	357
6.14	Cooperative Data Fusion (50%) Results	358
6.15	Cooperative Data Fusion (60%) Statistics for Threshold Calculation .	359
6.16	Cooperative Data Fusion (60%) Results	360
6.17	Cooperative Data Fusion (70%) Statistics for Threshold Calculation .	361
6.18	Cooperative Data Fusion (70%) Results	362
6.19	Summary Statistics of Experimental Results	364
7.1	Corresponding Table Numbers	368
7.2	Duplication of Keystroke Dynamics Statistics	370
7.3	Duplication of Keystroke Dynamics Results	371
7.4	Summary of Reviewed Papers Using Statistical Analysis Methods . .	375
7.5	Summary of Reviewed Papers Using Machine Learning Techniques . .	378
7.6	Summary of Reviewed Papers Using Artificial Neural Networks	382
7.7	Duplication of Fingerprint Recognition Statistics	390
7.8	Duplication of Fingerprint Recognition Results	391
7.9	Summary of Fingerprint Recognition Results For Reviewed Papers . .	393

7.10	Duplication of Complementary Data Fusion Statistics	405
7.11	Duplication of Complementary Data Fusion Results	407
7.12	Summary of Reviewed Papers Using Complementary Data Fusion . . .	408
7.13	Duplication of Cooperative Data Fusion (Stage 1 – 40%) Statistics . .	416
7.14	Duplication of Cooperative Data Fusion (Stage 1 – 40%) Results . . .	417
7.15	Duplication of Cooperative Data Fusion (Stage 2 – 50%) Statistics . .	419
7.16	Duplication of Cooperative Data Fusion (Stage 2 – 50%) Results . . .	420
7.17	Duplication of Cooperative Data Fusion (Stage 3 – 60%) Statistics . .	422
7.18	Duplication of Cooperative Data Fusion (Stage 3 – 60%) Results . . .	424
7.19	Duplication of Cooperative Data Fusion (Stage 4 – 70%) Statistics . .	425
7.20	Duplication of Cooperative Data Fusion (Stage 4 – 70%) Results . . .	427
7.21	Summary of Reviewed Papers Using Cooperative Data Fusion	429
A.1	Reported Security Breaches (1988-2003)	456
A.2	Reported Vulnerabilities (1995-2008)	457
A.3	Number of Internet Users (December 1995-June 2002)	458
B.1	Coefficient Values For Each Metric	462
B.2	Sorted Coefficient Values And Associated Metric Numbers	463
B.3	Sorted Metrics With Rank Allocations	464
B.4	Accumulated Rank Score For Metrics	465
C.1	Keystroke Dynamics Directory Structure	468
C.2	Fingerprint Recognition Directory Structure	473
C.3	Complementary Data Fusion Directory Structure	477
C.4	Directory Structure	479

List of Figures

2.1	The Generic Biometric Authentication System	24
2.2	Complementary Data Fusion Paradigm	45
2.3	Competitive Data Fusion Paradigm	46
2.4	Cooperative Data Fusion Paradigm	47
2.5	Data Fusion Levels In Multi-Modal Biometrics	53
2.6	Feature Level Data Fusion	56
2.7	Confidence Score Level Data Fusion	58
2.8	Decision Level Data Fusion	60
2.9	Components of a Biological Neuron	85
2.10	Simple Non-linear Model of a Neuron	88
2.11	Illustrations of Step-wise, Piece-wise, And Sigmoid Functions	89
2.12	The Single Layer Perceptron	96
2.13	The Multi-Layer Perceptron	98
2.14	The Hopfield Neural Network	103
2.15	The Self-Organising Map (SOM)	106
2.16	Adaptive Resonance Theory (ART)	109
3.1	States of a Two-Key Combination	123
3.2	Keystroke Durations and Digraph Latencies for the digraph “th”	123
4.1	Fingerprint Impression Illustrating Ridges And Furrows	156
4.2	Fingerprint Impression Illustrating Core And Delta Points	163
4.3	Local Fingerprint Features Types	165
4.4	Local Features Illustrating Minutiae Positions	166

4.5	Captured Fingerprint And Its Orientation Field	179
4.6	Captured Fingerprint, Binary, and Thinned Representations	181
4.7	Fingerprint Classes defined by Henry	184
5.1	Graphical User Interface for Keystroke Dynamics Capture Program	230
5.2	Creation of Training and Testing Files For ANN Processing	249
5.3	Graphical User Interface for Fingerprint Feature Capture Program	253
5.4	Alignment Example	260
5.5	Local Area Alignment Example	264
5.6	Filter Model	296
5.7	Wrapper Model	297
6.1	Contingency Table	319
6.2	ROC Space	324
6.3	Binary Classifiers	325
6.4	ROC Curve	327
6.5	Comparison of AUC for Two Classifiers	330
6.6	ROC Curve for Participant 1	337
6.7	Example Demonstrating Best Classification	347
6.8	Example Demonstrating Average Classification	347
6.9	Example Demonstrating Worst Classification	348
6.10	Keystroke Dynamics	350
A.1	Reported Security Breaches (1988-2003)	456
A.2	Reported Vulnerabilities (1995-2008)	457
A.3	Number of Internet Users (December 1995-June 2002)	459
D.1	Best Classification Performance – Participant 52	484
D.2	Good Classification Performance – Participant 18	484
D.3	Good Classification Performance – Participant 27	485
D.4	Good Classification Performance – Participant 60	485
D.5	Average Classification Performance – Participant 38	486
D.6	Average Classification Performance – Participant 49	486

D.7 Average Classification Performance – Participant 61 487

D.8 Worst Classification Performance – Participant 3 487

D.9 Worst Classification Performance – Participant 12 488

D.10 Worst Classification Performance – Participant 74 488

Acknowledgments

I would like to express my sincere gratitude to my primary supervisor, Dr Andrew Turk, whose analytical skills proved most valuable and insightful during the course of this project. Dr Turk has been inspirational in encouraging me to strive for the highest standards, whilst undertaking rigorous and honest research. He has also been very helpful in regard to maintaining self-discipline (in relation to one's research work ethic), and a congenial and cooperative attitude with one's colleagues.

This research involved many technical aspects, and thanks are extended to my secondary supervisor Mr Shri M. Rai for his assistance with these matters. Shri Rai supervised my honours research and was therefore familiar with previous work in the area of keystroke dynamics. However, his proficiency in mathematics and science became truly valuable when working with fingerprint data and data fusion issues.

My thanks to friend and colleague, Dr Christian Payne. Christian first kindled my interest in computer security, and this area of research remains as interesting and challenging today as ever.

Thanks to Dr Lance Fung for suggesting the inclusion of fingerprint recognition for this project. His suggestion was inspirational and opened up new areas of investigation. Dr Fung also arranged for the purchase of the fingerprint scanner and the software development kit that were used for the experiment.

To my dear friend Dr R. (Chandra) Chandrashekhara of SwanLotus, thank you for your assistance with interpreting the point pattern matching algorithm used in the experiment to align fingerprint features. Your contribution was invaluable, and your friendship is priceless.

To my family and friends, thank you for the support that only loved ones can give.

To my dearest friend Paramahansa Yogananda, thank you so much for showing me a purpose to life and guiding me through its many, and varied, joys and tribulations.

