METHODOLOGIES FOR USER AUTHENTICATION USING KEYSTROKE DYNAMICS

APARNA BHATIA

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2018

METHODOLOGIES FOR USER AUTHENTICATION USING KEYSTROKE DYNAMICS

by

Aparna Bhatia

Department of Electrical Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

November 2018

Dedicated to my daughter

Nitya

CERTIFICATE

This is to certify that the thesis titled "**Methodologies for User Authentication using Keystroke Dynamics**" being submitted by **Aparna Bhatia** to the Department of Electrical Engineering, Indian Institute of Technology, Delhi, for the award of the degree of **Doctor of Philosophy**, is a record of bonafide research work carried out by her under our guidance and supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Date

New Delhi

Dr. M. Hanmandlu

Professor

Department of Electrical Engineering

Indian Institute of Technology Delhi

New Delhi-110016

Dr. B.K. Panigrahi

Professor

Department of Electrical Engineering

Indian Institute of Technology Delhi

New Delhi-110016

ACKNOWLEDGEMENTS

First and foremost, I would pay my sincere gratitude to my supervisors, **Prof. M. Hanmandlu** and **Prof. B. K. Panigrahi** for providing me the constant encouragement and guidance at every stage of my research. Without their constant support and guidance, I could not have achieved my research goals. I could not imagine having better supervisors than them for my Ph.D. program.

I extend my thanks to my SRC members, Prof. K. K. Biswas, Prof. Niladri Chaterjee, Prof. Sumantra Dutta Roy and Prof. Brejesh Lall for their fruitful suggestions and critical comments during my research presentations.

I am equally thankful to my seniors, friends and lab mates Amioy Kumar, Jyotsana Grover, Mamta Bansal, Jeevan and Farah. It was a great pleasure to work along with them gaining their experiences and sharing their learning.

Any work would not be possible without love and support from my family. I do not have words to express my gratitude to my mother for being the strong pillar of support at all needy times. This work won't have reached this stage without her everlasting love, support and sacrifices. I owe my most heartfelt thanks to my husband, Nitin Bhatia, for his constant love, encouragement and patience. I can't imagine completing this stupendous task without his support, sacrifice and patience that sustained my efforts and fueled energy for finishing my research goals.

Fondly I thank all the people who have indirectly helped reach the final destination of completion in my Ph.D. journey.

Finally, my greatest gratitude is for the almighty for his blessings and for imparting me the passion and perseverance needed to achieve my goal.

Aparna Bhatia

ABSTRACT

Considering the importance of securing desktops, laptops and mobile phones, this thesis makes several studies aimed at utilizing the behavioral biometrics like keystroke dynamics for the authentication of users for providing secure access to the devices. As keystroke dynamics being a weak biometric, it is combined with a physiological biometric like face to improve its verification accuracy in case of the need for the highly secured access control.

The uncertainty representation in the measurements of keystroke dynamics is the main consideration for investigating different approaches. In the first approach, we employ Generalized Fuzzy Model (GFM) that combines Mamdani and Takagi-Sugeno fuzzy models for the representation of uncertainty. GFM stems from the well-known Gaussian Mixture Model (GMM) which is derived from the timing measurements of the keystroke dynamics, and it helps in obtaining the parameters of GFM. The benchmark CMU dataset comprising several samples with each sample accounting for a timing sequence (or timing measurements) is used to validate GFM on keystroke dynamics in a real-world situation, and its performance is compared with that of GMM.

In the second approach, the Information set concept originated from the information theoretic entropy function called the Hanman-Anirban entropy function is used to represent possibilistic uncertainty in the timing measurements of keystroke dynamics. The information set consists of a set of information values each of which is a product of information source value (i.e., measurement) and its membership function value. Two types of membership functions (MFs): one based on the timing features of all the samples and another based on the timing features of a single sample are computed. These MFs lead to two types of information components (i.e., spatial and temporal uncertainties) which are concatenated and modified to

produce different feature types. Two Component Information Set (TCIS) is proposed for the development of keystroke dynamics based user authentication system. The keystroke features are converted into TCIS features which are then classified by SVM, Random Forest and proposed Convex Entropy Classifier.

In the third approach, the uncertainty in the measurements of keystroke dynamics is dealt with, by formulating the possibilistic Renyi entropy function from the Renyi entropy function using the framework of Hanman-Anirban entropy function. The possibilistic Renyi entropy function is then used to derive the information set features from keystroke dynamics for the authentication of users. A new composite fuzzy classifier is also proposed based on another entropy function called Mamta-Hanman entropy function and applied on the Information set based features. A comparison of the results of the proposed approach with those of Support Vector Machine and Random Forest classifier shows that the new classifier outperforms the other two.

In the fourth approach, the radial basis function network (RBFN) is employed for the representation of uncertainty in the measurements of keystroke dynamics after clustering so as to facilitate the application of RBF (also called Gaussian MF) on the clusters. The coefficients (weights) of RBFN model are computed using the pseudo-inverse method. The model equation is the sum of weighted radial basis function values of the inputs. Next, an Entropy network is developed based on Information set concept. Instead of using RBFN model as the sum of the weighted radial basis function values of the inputs entropy network model takes the product of the coefficients and any kind of information values like sigmoid, energy, composite etc. The performance of Entropy network is shown to be superior to that of RBFN on CMU dataset.

In the above four approaches, the uncertainty representation has been carried out only on the measurements or features of keystroke dynamics. In the fifth approach, the information set based features arising from the uncertainty representation in both keystroke dynamics and face are employed. As the face is affected with variations in illumination and pose, the face images are operated upon by Weber transform and LDP masks to remove the effects of illumination and then information set features are derived and PLPP is applied on these features to remove the effects of pose variations. The features used for the face are almost similar to those of keystroke dynamics but for the membership employed. The fusion of both keystroke dynamics and face is done using different fusion rules, of which composite fusion yields the best results.

Thus, this thesis strives to make keystroke dynamics to the level of a strong biometric by improving its performance using an effective representation.

सार

डेस्कटॉप, लैपटॉप और मोबाइल फोन को सुरक्षित करने के महत्व को ध्यान में रखते हुए, यह थीसिस कई अध्ययन करता है जिसका लक्ष्य है कि डिवाइसों तक सुरक्षित पहुंच प्रदान करने के लिए उपयोगकर्ताओं के प्रमाणीकरण के लिए कीस्ट्रोक गतिशीलता जैसे व्यवहार बॉयोमीट्रिक्स का उपयोग करना। चूंकि कीस्ट्रोक गतिशीलता कमजोर बॉयोमीट्रिक होने के नाते, अत्यधिक शारीरिक पहुंच नियंत्रण की आवश्यकता के मामले में इसकी सत्यापन सटीकता को बेहतर बनाने के लिए इसे शारीरिक बॉयोमीट्रिक के साथ जोड़ा जाता है।

कीस्ट्रोक गतिशीलता के माप में अनिश्चितता प्रतिनिधित्व विभिन्न दृष्टिकोणों की जांच के लिए मुख्य विचार है। पहले दृष्टिकोण में, हम सामान्यीकृत फ़ज़ी मॉडल (जीएफएम) को नियोजित करते हैं जो अनिश्चितता के प्रतिनिधित्व के लिए मामदानी और ताकागी-सुगोनो अस्पष्ट मॉडल को जोड़ती है। जीएफएम प्रसिद्ध गाऊसी मिश्रण मॉडल (जीएमएम) से उत्पन्न होता है जो कि कीस्ट्रोक गतिशीलता के समय माप से प्राप्त होता है, और यह जीएफएम के मानकों को प्राप्त करने में मदद करता है। एक समय अनुक्रम (या समय माप) के लिए प्रत्येक नमूना लेखांकन के साथ कई नमूनों वाले बेंचमार्क सीएमयू डेटासेट का उपयोग वास्तविक दुनिया की स्थिति में कीस्ट्रोक गतिशीलता पर जीएफएम को मान्य करने के लिए किया जाता है, और इसके प्रदर्शन की तुलना जीएमएम की तुलना में की जाती है।

दूसरे दृष्टिकोण में, जानकारी सेट अवधारणा से उत्पन्न जानकारी सैद्धांतिक एन्ट्रॉपी फ़ंक्शन हनमन-अनिरबान एन्ट्रॉपी फ़ंक्शन नामक कुंजीस्ट्रोक गतिशीलता के समय माप में संभाव्य अनिश्चितता का प्रतिनिधित्व करने के लिए उपयोग की जाती है। सूचना सेट में सूचना मानों का एक सेट होता है जिनमें से प्रत्येक सूचना स्रोत मूल्य (यानी माप, माप) और उसके सदस्यता कार्य मूल्य का उत्पाद होता है। दो

प्रकार के सदस्यता कार्य (एमएफ): सभी नमूने की समय-सारिणी सुविधाओं के आधार पर और एक नमूने की समय-सारणी सुविधाओं के आधार पर गणना की जाती है। ये एमएफ दो प्रकार के सूचना घटकों (यानी, स्थानिक और लौकिक अनिश्चितताओं) का कारण बनते हैं जो अलग-अलग फीचर प्रकारों का उत्पादन करने के लिए संयोजित और संशोधित होते हैं। कीस्ट्रोक गतिशीलता आधारित उपयोगकर्ता प्रमाणीकरण प्रणाली के विकास के लिए दो घटक सूचना सेट (टीसीआईएस) का प्रस्ताव है। कीस्ट्रोक सुविधाओं को टीसीआईएस विशेषताओं में परिवर्तित किया जाता है जिन्हें तब एसवीएम, रैंडम वन और प्रस्तावित उत्तल एन्ट्रॉपी क्लासिफायर द्वारा वर्गीकृत किया जाता है।

तीसरे दृष्टिकोण में, कीस्ट्रोक गतिशीलता के माप में अनिश्चितता का निपटारा हनमन-अनिरबान एंट्रॉपी फ़ंक्शन के ढांचे का उपयोग करके रेनई एंट्रॉपी फ़ंक्शन से संभाव्य रेनई एन्ट्रॉपी फ़ंक्शन को तैयार करके किया जाता है। Possibilistic Renyi एंट्रॉपी फ़ंक्शन का उपयोग उपयोगकर्ताओं के प्रमाणीकरण के लिए कीस्ट्रोक गतिशीलता से जानकारी सेट सुविधाओं को प्राप्त करने के लिए किया जाता है। ममता-हनमन एंट्रॉपी फ़ंक्शन नामक एक और एन्ट्रॉपी फ़ंक्शन के आधार पर एक नया समग्र अस्पष्ट वर्गीकरण भी प्रस्तावित किया गया है और सूचना सेट आधारित सुविधाओं पर लागू किया गया है। समर्थन वेक्टर मशीन और रैंडम वन क्लासिफायर के साथ प्रस्तावित दृष्टिकोण के परिणामों की तुलना से पता चलता है कि नया वर्गीकरण दूसरे दो से बेहतर प्रदर्शन करता है।

चौथे दृष्टिकोण में, क्लस्टर के बाद कीस्ट्रोक गतिशीलता के माप में अनिश्चितता के प्रतिनिधित्व के लिए रेडियल आधार फ़ंक्शन नेटवर्क (आरबीएफएन) नियोजित किया जाता है ताकि क्लस्टर पर आरबीएफ (जिसे गॉसियन एमएफ भी कहा जाता है) के उपयोग को सुविधाजनक बनाया जा सके। आरबीएफएन मॉडल के गुणांक (भार) छद्म-उलटा विधि का उपयोग करके गणना की जाती हैं। मॉडल समीकरण इनपुट के भारित रेडियल आधार फ़ंक्शन मानों का योग है। इसके बाद, सूचना सेट अवधारणा के आधार

पर एक एंट्रॉपी नेटवर्क विकसित किया गया है। भारित रेडियल आधार के योग के रूप में आरबीएफएन मॉडल का उपयोग करने के बजाय इनपुट एंट्रॉपी नेटवर्क मॉडल के फ़ंक्शन वैल्यू गुणांक का उत्पाद लेते हैं और सिग्मोइड, ऊर्जा, समग्र इत्यादि जैसे किसी भी प्रकार के सूचना मान लेते हैं। एंट्रॉपी नेटवर्क का प्रदर्शन दिखाया गया है सीएमयू डेटासेट पर आरबीएफएन से बेहतर।

उपर्युक्त चार दृष्टिकोणों में, अनिश्चितता का प्रतिनिधित्व केवल कीस्ट्रोक गतिशीलता के माप या विशेषताओं पर किया गया है। पांचवें दृष्टिकोण में, कीस्ट्रोक गतिशीलता और चेहरे दोनों में अनिश्चितता प्रतिनिधित्व से उत्पन्न जानकारी सेट आधारित सुविधाओं को नियोजित किया जाता है। जैसे-जैसे रोशनी रोशनी और मुद्रा में भिन्नता से प्रभावित होती है, चेहरे की छवियों को रोशनी के प्रभावों को दूर करने के लिए वेबर ट्रांसफॉर्म और एलडीपी मास्क द्वारा संचालित किया जाता है और फिर सूचना सेट सुविधाओं को व्युत्पन्न किया जाता है और पीओपीपी इन सुविधाओं पर लागू होता है ताकि पॉज़ के प्रभावों को दूर किया जा सके। विविधताओं। चेहरे के लिए उपयोग की जाने वाली विशेषताएं लगभग कीस्ट्रोक गतिशीलता के समान हैं लेकिन सदस्यता के लिए नियोजित हैं। दोनों कीस्ट्रोक गतिशीलता और चेहरे का संलयन विभिन्न संलयन नियमों का उपयोग करके किया जाता है, जिसमें से समग्र संलयन सर्वोत्तम परिणाम उत्पन्न करता है।

इस प्रकार, यह थीसिस प्रभावी प्रभाव का उपयोग करके अपने प्रदर्शन में सुधार करके एक मजबूत बॉयोमीट्रिक के स्तर पर कीस्ट्रोक गतिशीलता बनाने का प्रयास करता है।

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACE	Adaptive Convex Entropy Classifier
ANN	Artificial Neural Network
BP	Backpropagation
CKAA	Clustering based keystroke authentication algorithm
DBN	Deep Belief Networks
DSM	Direction Similarity Measure
EER	Equal Error Rate
FAR	False Acceptance Rate
FIS	Fuzzy Inference System
FRR	False Rejection Rate
GAR	Genuine Acceptance Rate
GFM	Generalized Fuzzy Model
GMM	Gaussian Mixture Model
GMM-UBM	Gaussian Mixture Model with Universal Background Model
GPDF	Gaussian Probability Density Function
K-NN	K-Nearest Neighbor
LBP	Local Binary Pattern

LDP	Local Directional Patterns
LFU	Least Frequently Used
LOOM	Leave One Out Method
MFCC	Mel-frequency Cepstral Coefficients
MISO	Multi-Input Single-Output
MLP	Multi-Layer Perceptron
ML	Mamdani-Larsen
PCA	Principal Component Analysis
PLPP	Pseudo-inverse Locality Preserving Projections
RBF	Radial Basis Function
RBFN	Radial Basis Function Network
RBM	Restricted Boltzmann Machine
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
TCIS	Two Component Information Set
TS	Takagi-Sugeno
UUKL	Up-Up Keystroke Latency
ZMFAR	Zero-Miss False Alarm Rate