CHAPTER 3

KEYSTROKE DYNAMICS – HUMAN AUTHENTICATION SYSTEM

3.1 INTRODUCTION

Typing rhythms form the data, jetting from the interface between users and computers. Upon proper sampling and analysis, they may emerge as functional contrivances in ascertaining personal identity. Not like other access control systems based on biometric features, keystroke analysis has not shown the way to procedures making available an appropriate plane of precision. The cause is in all likelihood the intrinsic variability of typing dynamics versus other very stable, biometric characteristics, such as faces or fingerprint patterns. In this chapter, the investigation is anchored in a promising non-static biometric technique that targets to identify users rooted in analyzing habitual rhythm prototypes in the way they type. Here, some features are extracted that relate to time and depression of keys such as the duration of typed keys, latency between two consecutive keystrokes, Digraph and Tri-graph; thereby make the authentication depend upon the extracted features. Besides time related features, Virtual Key Force (VKF) is taken into account for the authentication. It is also proposed to add uniqueness term in the fitness function of genetic algorithm. The results validate that the use of the Keystroke Dynamics is uncomplicated and professional for personal authentication.
3.2 SCIENCE OF KEYSTROKE DYNAMICS

The biometric features are conveniently divided into two main categories. The physiological features include face, eye, fingerprints, palm topology, hand geometry, wrist veins and thermal images. The behavioural features include voiceprints, handwritten signatures and keystroke dynamics. Unlike other biometric methods, keystroke analysis can be done without the aid of special tools, just the keyboard of the computer where the biometric analysis has to be performed (Bergadano et al 2002). Biometric features are interesting for computer security because, on the one hand, they are sufficiently unique to be used to recognize legal users of a systems and to reject impostors and, on the other hand, they cannot be forgotten, lost, overheard, stolen or extorted (Bergadano et al 2002). Keystroke dynamics is considered as a strong behavioural biometric based authentication system. The behavioural biometric of Keystroke Dynamics uses the manner and rhythm in which an individual types characters on a keyboard or keypad (Gunetti et al 2005). The keystroke rhythms of a user are measured to develop a unique biometric template of the users typing pattern for future authentication. Raw measurements available from almost every keyboard can be recorded to determine Dwell time (the time taken to press a key) and Flight time (the time between “key up” and the next “key down”). The recorded keystroke timing data is then processed through a unique neural algorithm which determines a primary pattern for future comparison. Similarly vibration information may be used to create a pattern for future use in both identification and authentication tasks. Data needed to analyze keystroke dynamics is obtained by keystroke logging. Normally, all that is retained when logging a typing session is the sequence of characters corresponding to the order in which the keys were pressed and timing information is discarded. The latest research centres on the use of this keystroke dynamic function, which is normally discarded, to verify or even
try to determine the identity of the person who is producing those keystrokes. This is often possible because some characteristics of keystroke production are as individual as handwriting or signature. The techniques used to do this vary widely in power and sophistication and range from statistical techniques to neural-nets to artificial intelligence. The time taken to get to and depress a key (seek-time) and the time taken to hold down the key (hold-time) may be very characteristic for a person regardless of the typing speed. Common “errors” may also be quite characteristic of a person and there is an entire taxonomy of errors: “substitutions”, “reversals”, “drop-outs”, “double-strikes”, “adjacent letter hits”, “homonyms” and “hold-length-errors” (shift key held down too long or too short at a time). Thus the patterns of errors might be sufficiently different to distinguish two people. Moreover, unlike other biometric systems, which may be expensive to implement, keystroke dynamics is almost free as the only hardware required is the keyboard. There are two approaches in keystroke authentication.

3.3 IDENTIFICATION AND VERIFICATION

Keystroke dynamic systems runs in two unalike modes – identification mode or verification mode. Identification is the process of trying to identify a user through the scrutiny of a biometric pattern calculated from the person’s biometric features. A sizeable quantity of keystroke data is collected and the user is identified based on previously stored information relating to keystroke dynamic profiles of all users. For every user, a biometric template is formed at the training stage. A pattern that is going to be identified is matched against every acknowledged template, resulting in a score or distance illustrating the similarity between the pattern and the template. The system then assigns the pattern to the user with the closest similarity to the biometric template. In order to prevent the imposter patterns, all the patterns of users not stored in the system, from being fallaciously
identified as genuine, the similarity has to have a high water mark. This is ensured because if the user does not reach this level, the pattern is rejected. Identification in keystroke dynamics implies that the identification of the user solely depends on the measurement of his keystroke dynamics without any additional inputs.

Verification phase is all about checking the identity of the user. The pattern of the user to be verified is compared with the person’s individual template. Keystroke verification techniques are grouped under two domains:

- Static
- Dynamic or Continuous

Static verification approaches analyze keystroke verification characteristics only at specific instances, thus proving as additional security besides traditional username/password as in the case of user login sequence. Static approaches provide more robust user verification in comparison with simple passwords but the drawback lies in the fact that the change of a user could become impossible after the login authentication. Continuous verification, in contrast, monitors the user’s typing behaviour throughout the course of interaction. A real time analysis is made possible in the continuous process since the user is scrutinized ad infinitum. Such a condition ensures that even after a successful login, the access to the system could be thwarted if there is any mismatch in the typing patterns of the user uncovered through persistent observation.

3.4 BENEFITS AND SHORTCOMINGS OF KEYSTROKE DYNAMICS

Keystroke dynamics, which is part of a larger class of biometrics, Behavioural Biometrics, has patterns which are statistical in nature. It is a
commonly held belief that behavioural biometrics are not as reliable as physiological biometrics used for authentication i.e. finger prints or retinal scan or DNA. The reality is those behavioural biometrics use a confidence measurement instead of the traditional pass / fail measurements. As a result, the traditional benchmarks of False Acceptance Rate (FAR) and False Rejection Rate (FRR) have no longer linear relationships. The benefit to keystroke dynamics as well as other behavioural biometrics is that FRR/FAR can be adjusted by changing the acceptance threshold at the individual level. This allows for explicitly defined individual risk mitigation – something which the physiological biometric technologies could never achieve. Following is a summary of the advantages of keystroke dynamics authentication.

- It is not intrusive and computer operators, either professionals or pleasure seekers, in normal course, need to type on a computer keyboard.
- It is low-cost since keyboard attached to a computer is the only hardware needed.
- Keystrokes can be captured continuously – not just at the start-up time alone and therefore there may be adequate opportunities to trigger an alarm in case there is a discrepancy.
- With more businesses shifting to e-commerce, the keystroke biometric in internet applications can make available an effective balance between high security and user-friendliness for customers (Villaniet al 2006).

The keystroke dynamics has disadvantageous aspects also.

- Keystrokes, unlike other biometric, communicate an amorphous and modest quantity of information.
- Keystroke duration and digraph latency are in fact a pretty shallow brand of information.

- Keystroke dynamics, as behavioural biometric, like voiceprints and handwritten signatures, are intrinsically unstable and display a certain degree of inconsistency even without any palpable intention. It is reasonably tricky to have power over the number of milliseconds to hold down a key when typing.

- The unpredictability of typing rhythms may be puffed up by the fact that, during the normal use of a computer, different texts are entered, possibly in different languages (Gunetti et al. 2005).

Keystroke dynamics can be used for various authentication processes like Synthetic Forgeries where keystroke dynamics is robust against synthetic forgery attacks, in which an attacker draws statistical samples from a pool of available keystroke data sets other than the target (Stefan 2010) emotional states where it is used to determine user emotion by analyzing the rhythm of their typing patterns on a standard keyboard, Adaptive Behaviometrics where it is used to verify the user using statistical method, measure of disorder and direction similarity measure that recognized the user based on the adaptive local threshold (Liliana et al. 2011), Analysis of Template Update Strategies where a protocol for evaluating keystroke dynamics template update systems are used. This protocol could also be used for other modalities for which databases captured on several sessions are available (Giot et al. 2011) such as password hardening where the legitimate user’s typing patterns are combined with the user’s password to generate a hardened password that is convincingly more secure than conventional passwords alone (Monrose et al. 2002).

User Identification on Smart Phones where keystroke dynamics of a smart phone user can be translated into a viable feature set for accurate user identification (Zahid et al. 2009), Shared Secret where they use GREYC
keystroke database that is composed of a large number of users for validation purposes (Giot et al 2011). User Authentication using Rhythm Click Characteristics, where the rhythms clicked by a mouse as another identifiable factor and mouse clicks can be replaced by a stylus on non-keyboard devices, numeral buttons on mobile phones, or fingers on touch screens to enhance system portability (Chang et al 2011). Anomaly Detection where keystroke dynamics data set is collected, to develop a repeatable evaluation procedure, and to measure the performance of a range of detectors so that the results can be compared soundly (Killourhy 2009).

3.4.1 Historical Perspective

A compendium of information relating to the chronological perspectives of keystroke dynamics authentication approach is presented sequentially to have a better understanding of the present venture.

Karnan et al (2009) have put forward feature subset selection in Keystroke Dynamics for identity verification and it reports the results of experimenting with Ant Colony Optimization technique on keystroke duration, latency and digraph for feature subset selection. The Ant Colony Optimization was used to reduce the redundant feature values and minimize the search space. Optimum feature subsets were obtained using keystroke duration values when compared with the other two feature values. In this regard, mean and standard deviation were used to extract the features from the Keystroke duration, latency and digraph.

Raj &Thomson (2009) advocate a biometric identification problem with focus on extracting the behavioral features of the user and the process of using these features for computer security. Homogeneous mouse dynamics biometrics involves a signature based on distinct mouse movement characteristics under different screen resolution and mouse pointer speed.
settings. Several trials were performed under diverse settings to form the mouse dynamics signature of the user. Earlier methods failed to give better results when performed under different screen resolution and mouse pointer speed. The intended method standardizes the user signature irrespective of the setting making it more expedient for security application. The mouse dynamics could be further regulated based on the mouse pointer speed. The combination of standardized screen resolution and homogenized mouse pointer speed would guarantee still better result.

Jamil & Khan (2011) have the proposition of developing biometric access control measure via keystroke pattern recognition and examine its direct connection in preventing electronic identity thefts. They investigated keystroke dynamics as one of the most cost efficient and easy to implement biometrics for online and project based systems. The prominence was rested on both strategies and mathematical models designed for implementing the keystroke pattern algorithms. They elucidated the keystroke recognition's efficacy in a potential business environment and came up with a schema suggestion for all the phases of a process relying on keystroke dynamics. The inference drawn is that the keystroke dynamics is a user friendly biometric authentication technique already available for online applications, web based emailing and other online services. It curtails the negative bearing on the user's privacy and is very undemanding in incorporation. Hence keystroke pattern recognition technique could be used effectively as a safeguard against unauthorized access to computer resources and sensitive data.

Raghu et al (2011) hint at an association of ideas of neural network guided with the timing vectors of the keystroke dynamics and then used to discriminate between the owner and an imposter. They presented an application of neural nets to user identity authentication on computer access security system. A three-layered back propagation neural network with a
flexible number of input nodes was used to discriminate between bona fide users and impostors derived from each individual's password keystroke pattern. Keystroke latency was measured for each user which forms the patterns of keyboard dynamics. System verification performance was improved by setting convergence criteria RMSE to a smaller threshold value during schooling procedure. An imposter could be detected even if the correct password was typed by him based on the very high accuracy of the method. Moreover, the implementation methodology requires no special hardware and could be very effortlessly integrated with the majority computer systems.

Shanmugapriya & Padmavathi (2011) developed a method which has many loopholes i.e. password sharing, shoulder surfing, brute force attack, dictionary attack, guessing and phishing. Keystroke Dynamics is one of the well-known and inexpensive behavioral biometric technologies which identify the authenticity of a user while the user was operating a system using a keyboard. Keystroke features like dwell time, flight time, di-graph, trigraph and virtual key force of every user are exercised in the process. For the purpose of preprocessing, Z-Score method was used. Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) are used in association with Extreme Learning Machine (ELM) for feature subset selection. In order to classify the obtained results, ELM algorithm is used. Comparison of ACO, PSO and GA with ELM respectively is done to find the best method for feature subset selection.

Abualgasim & Osman (2011) have designed a keystroke dynamics biometric as an added security for passwords and Personal Identification Numbers (PINs) used in Points of Sale (POS) and banking Automatic Teller Machines (ATM). They build an algorithm using the keystroke dynamics which attempts to minimize the compromise on security when the imposter
gets hold of both user-ID (user card) and password. Experiments are conducted using passwords of varying length by a group of users. The recommendation is the use of an optimum number of digits in passwords when this biometrics is used. The experimental results are encouraging for PINs of length 4-12 digits where the rate of rejection of legitimate users (FRR) becomes zero while the rate of acceptance of the imposter (FAR) is way below 20%.

Schlenker & Sarek (2011) put forward a scheme to analyze current state in the use of biometrics in computer security. They suggest the most commonly applied anatomical, physiological and behavioral biometric identification methods. The plan is a multifactor system that would verify a number of biometric features simultaneously, thus ensuring greater trustworthiness of identification. The result of the work is a set of methods, which allows unfailing identification of the user in the cosiest way. These methods of data security have been used to enhance the protection of specialized health record. The method contributes to expansion of generally conceived EHR MUDR Concept to other application areas.

3.4.2 Commercial Applications

The keystroke authentication system under study in this thesis has already found its commercial applications in the following domains. TypeWATCH analyzes the free text typing patterns of each user to identify data theft attempts. Intensity Analytics identifies and validates users across applications ranging from compliance and documentation, authentication, forensics to field intelligent applications. AdmitOneSecurity, among other things, detects online frauds. Bio Tracker offers a high level of real time network security. Key Trac offers uninterrupted flow of work with its concealed background keystroke recording. IMagic Software uses Trustable passwords for web authentication and large scale enterprise authentication.
ID Control has a low FRR and FAR for verification and identification. TypeSense of Deepnet Security employs advanced new algorithms to achieve better results. Psylock claims to be the technological leader in keystroke dynamics. Authenware corp., bioChec, Probayes, Delfigo Security and BehavioSec are also in the fray.

3.5 HUMAN AUTHENTICATION SYSTEM BASED KEYSTROKE DYNAMICS APPROACH

Keystroke dynamics is a biometric based on the assumption that diverse people type in matchless characteristic manners. Keystroke dynamics is mostly applied to verification but identification can also be made possible. In the verification phase, it is an established fact that who the user is supposed to be and the biometric system should verify whether the user is who he claims to be. In the identification segment, the biometric system should identify the user, by means of his keystroke dynamics without any supplementary inputs. As a rule, the majority of applications of keystroke dynamics pertain to verification. Keystroke analysis is essentially a form of Pattern Recognition (as is the case with regard to the largest part of biometric techniques). Consequently, it involves the representation of input data measures, extraction of characteristic features and classification or identification of patterns data in an attempt to settle on to which pattern class these data belong.

Figure 3.1 lays bare the basic flow diagram of human authentication system using keystroke dynamics. This authentication system consists of three steps.
Figure 3.1 Basic Flow Diagram

i) Feature Extraction
ii) Pre-processing
iii) Feature Subset Selection (GA)

3.5.1 Feature Extraction

The input data is usually represented by a sequence of typed keys, together with appropriate timing information expressing the exact time at which keys have been depressed and released. From the input data, some of the features will be extracted.

i) Duration of typed keys (how long a key is held down)
ii) Latency between two consecutive keystrokes (the elapsed time between the release of the first key and the depression of the second key)
iii) Digraph
iv) Tri-graph
v) Pressure of keystroke
vi) Force of keystroke.
All the features are not useful and widely used. For measuring Pressure and Force of keystroke special type of pressure or force sensitive keyboard is required. Difficulties of typing text, frequency of word errors, typing rate are useful for long text. Here, the timing features such as Duration or Dwell time, Latency or Flight time, Digraph, Tri-graph are frequently measured from keystroke. Digraphs and Tri-graphs are sequences of two and three characters respectively. Various reasons exist for using digraphs and tri-graphs: keyboards may not have keys to cover the entire character set of the language, input of special characters may be difficult, and text editors may reserve some characters for special use and so on. Tri-graphs might also be used for some EBCDIC code pages that lack characters such as {and}. In addition to the above mentioned timing features, a new feature called Virtual Key force has been introduced.

The virtual key force (VKF) is calculated based on the typing speed and behaviour of the user on the key board. It measures the time taken by the user between releasing one key and pressing another key. It is based on the fact that each user has different typing speed and each user takes his own time to release and press another key. The use of keys and the typing speed and force are unlike for discrete users. Also the time interval taken for the release of one key and pressing of another key is not the same. Let us consider a user typing a word which consists of ten letters. There are nine time intervals between the release of one key and pressing of another key. The average typing speed of the user can be calculated based on these time intervals. Virtual key force can be determined from the key complexity. The key complexity can be calculated thus.

- According to the complexity in the use of keys, key complexity can be determined. It is based on the key position and distance.
• It means that the middle row keys (i.e., the keys from A to L) on the keyboard which are undemanding to cope with by all the users is taken as 0. The key complexity of remaining keys is taken as 1.

Figure 3.2 shows the extracted features: keystroke duration (duration) and keystroke latency (interval) of the word “IVAN.” The complexity label is assigned as CL = (1, 1, 1) i.e. the distance from I and V, V and A, A and N is longer(3.1).

![Figure 3.2 Extracted features of the Keystroke Dynamics](image)

The features extracted for the formation of the pattern profile the Features Vector having keystroke duration and keystroke latency.

Feature vector = \[ I_{tp}, VI_{ts}, V_{tp}, AV_{ts}, A_{tp}, NA_{ts}, N_{tp} \] (3.1)

where, \( I_{tp} \) is the keystroke duration of the key (I) i.e. time that the user takes for pressing and release of the key (I).

\( VI_{ts} \) is the keystroke latency of the keys (V) and (I) i.e. interval of time that the user takes to release the key (V) and press the key (I).
The keystroke duration is just made up of positive whole values. However, keystroke latency could convert positive values into negative. The negative value happens when the user presses the successive key even before the release of the current key. Inexperience or inadequate practice could lead to such pitfalls. Derived from the key complexity and the average time interval taken between releasing a key and pressing another key, an algorithm is formulated to find the Virtual Key Force:

\[
\text{If}(\text{key distance is nearer & time interval is below average})
\]

\[
\text{VKF} = 3
\]

\[
\text{Else if}(\text{key distance is nearer & time interval is above the average})
\]

\[
\text{VKF} = 1
\]

\[
\text{Else if}(\text{keys are longer and the average time interval is below average})
\]

\[
\text{VKF} = 3
\]

\[
\text{Else if}(\text{keys are longer and the average time interval is above average})
\]

\[
\text{VKF} = 2
\]

End

3.5.2 Preprocessing

The extracted feature may provide a great deal of superfluous information. Let us consider the multimodal human verification system that utilizes the combination approach to fusion at the match score level. Min–max and z-score normalization are some of the prevalent techniques used for relevance score normalization in meta search. The most frequently used score normalization technique is the z-score that is calculated using the arithmetic mean and standard deviation of the given data. Given a set of matching scores \( \{z_i\}, i = 1, 2, \ldots, n \), the normalized scores are given by
\[ n_i' = \frac{n_i - \mu}{\sigma} \]

where \( \mu \) is the arithmetic mean and \( \sigma \) is the standard deviation of the given data. Pre-processed results are given to the next step called feature subset selection.

### 3.5.3 Feature Subset Selection

From the reduced set of genes obtained in the previous pre-processing stage, this second stage uses a wrapper approach that combines a GA and an SVM to accomplish the feature subset selection. Feature subset selection is applied to high dimensional data before proceeding to the classification step. The basic idea here depends on using a GA to discover “good” subsets of genes, the reliability of a subset being evaluated by a SVM classifier on a set of training data. During this stage, high quality gene subsets are recorded in an archive for further analysis. At the end of the GA, the investigation of the archived gene subsets is performed: gene subsets are evaluated by putting them side by side and the most recurrently appearing genes are recognized.

Feature subset selection is fundamentally an optimization stumbling block. The selection is concerned with searching the locale of possible features to recognize one that is optimum or near-optimal in accordance with some performance measures, since the objective is to acquire any subset that reduces or improves a specific measure. GA wrapper and ensemble generates a population of accurate classifiers. Genetic Ensemble Feature Selection (GEFS) adds a diversity term in the fitness function of GA. Figure 3.3 shown below is the basic flow diagram of GA using SVM.

Ensemble is a set of classifiers trained differently: by separate data sets or features or models. After individual classifiers are trained, they are
combined either by majority voting or averaging to harvest a single value. The performance of an ensemble classifier has been found to be quite high in practice in a variety of applications. Individual classifiers participating in an ensemble have to be accurate as well as diverse in order to cause an accurate ensemble. Combination is in basic terms natural.

Figure 3.3 Flow Diagram of GA wrapper based feature subset selection

Contrary to the ordinary GA, GA wrapper has to find not only good but also diverse strings. In order to enforce diversity, the fitness function needs a diversity term as in GEFS. In the current project, one step approach is embraced resembling that of GEFS; yet with a more direct diversity term in the fitness function and SVM as base classifier. The "uniqueness" term is implemented here which measures each chromosome in terms of difference from other chromosomes. As more chromosomes with exclusive features are preferred, distinctiveness is simply added to accuracy just as diversity term in GEFS.
3.5.4 Crossover and Mutation

Crossover is a genetic operator used to alter the programming of a chromosome or many chromosomes from one generation to the next. It is analogous to reproduction and biological crossover upon which genetic algorithms are based. Crossover is a process of taking more than one parent solution and producing a child solution from them.

Mutation is a genetic operator used to maintain genetic diversity from one generation of a population of algorithm chromosomes to the next. It is analogous to biological mutation. Mutation alters one or more gene values in a chromosome from its initial state. In mutation, the solution may change entirely from the previous solution. Hence GA can come to better solution by using mutation and mutation occurs during evolution according to a user-definable mutation probability. This probability should be set low. If it is set too high, the search will turn into a primitive random search.

3.5.5 Evaluation by SVM

Support Vector Machines are basically binary classification algorithms. SVM computes the hyperplane that maximizes the margin between the training examples and the class boundary when the data is linearly separable. When the data is not linearly divisible, the examples are mapped to a high dimensional space where such an unscrambling hyperplane can be found. In the wrapper GA/SVM algorithm, an SVM classifier is applied to gauge the quality of a gene subset. The SVM-based classifier is used to ensure the fitness evaluation of each candidate gene subset. One important feature of the GA developed in this composition is the use of an archive to record quality gene subsets discovered during the gene subset selection stage. This archive is then analyzed to identify a small number of highly frequently appearing genes that are used in the final classification stage.
Before defining uniqueness, S-distance between the two chromosomes can be defined. The S-distance between two chromosomes \( i \) and \( j \), \( S(d_{ij}) \) is defined as

\[
S(d_{ij}) = \begin{cases} 
\left( \frac{d_{ij}}{C} \right)^2 & \text{if } d_{ij} < C \\
1 & \text{otherwise}
\end{cases}
\]

where \( d_{ij} \) denotes the Hamming distance between two chromosomes and \( C \) a constant. Figure 3.4 shows the graphical form of equation (3.3).

![Figure 3.4 S(d_{ij}) against d_{ij}](image)

Now the uniqueness of \( x^{th} \) chromosome is defined as an arithmetic average of S distances to all other chromosomes.

\[
U(x) = \frac{\sum_{x \neq j} S(d_{xj})}{n - 1}
\]

The fitness of a chromosome, i.e. a subset of genes, is assessed by the classification rate on the initial datasets. In other words, a subset of genes leading to a high classification rate is considered to be better than a subset
leading to a low classification rate. The fitness of chromosome \( x \) is defined as a simple sum of accuracy and uniqueness, which can be mentioned as,

\[
Fitness(x) = A(x) + U(x)
\]

The accuracy \( A(x) \) can be represented as

\[
A(x) = 1 - FRR
\]

where, \( FRR \) is the false rejection rate as only the user’s patterns are available in the training.

3.5.6 Outcome Estimation

The human authentication making use of keystroke dynamics is put into operation using MATLAB and realized that the projected practice is much better than other human authentication using keystroke dynamics. A data set is given as input for recognition (Killourhy & Maxion 2009). It is a benchmark data set for keystroke dynamics. The data is made up of keystroke-timing information from 50 subjects (typists). All subjects typed the same password and each subject typed the password 400 times over 8 sessions (50 repetitions per session). The break between the sessions is at least one day to facilitate capturing a few day-to-day deviations of each subject's typing. The password (.tie5Roanl) is chosen in such a way to be representative of a strong 10-character password. The raw records of all the subjects' keystrokes and time stamps were analyzed to create a password-timing table. The password-timing table encodes the timing features for each of the 400 passwords that each subject typed.

In this scheme, an efficient Genetic Algorithm is used for feature reduction. 5×10 chromosomes are engaged in initial population and again
5×10 chromosomes are taken up for second iteration for each person. In this analysis, features of 30 persons are referenced for recognition. As a result, 10×10 chromosomes total are taken into account for each person. The total number of best chromosomes selected is 1500. The total unique index of best chromosomes is 485. By means of these data, the performance of the proposed system is explored.

3.6 PERFORMANCE ANALYSIS

In the study of the performance of the intended process of human authentication system, two terms are taken for reflection

- Feature reduction rate
- Error authentication rate

3.6.1 Feature Reduction Rate

Feature selection is also acknowledged as variable selection or feature reduction or attributes selection or variable subset selection. This is the technique of opting for a subset of relevant features for building robust learning models. Subset selection searches the set of possible features for the optimal subset. The feature reduction rate can be calculated for the rate of reduction of the total features. This rate of reduction can be calculated with the help of the total number of features and the number of features selected for authentication. The feature reduction rate can be calculated using the following formula.

\[
Feature\ reduction = \frac{Total\ no.\ of\ features - no.\ of\ features\ selected}{Total\ no.\ of\ features}
\]
A comparison of the proposed approach is set with three other practices to analyze the performance. They are:

- Genetic Algorithm (GA)
- Particle Swarm Optimization (PSO)
- Ant Colony Algorithm (ACO) with Extreme Machine Learning (ELM)

Extreme Machine Learning (ELM) technique is used as an objective function in GA, PSO and ACO for feature selection. Particle swarm optimization (PSO) is a population dependent stochastic optimization approach. The system is initialized with a population of arbitrary solutions and optima are looked for through updation of generations. The possible solutions in PSO are called particles. The entire quantity of particles follows their coordinates in the problem situation which are related with the best possible solutions (fitness). An additional “best” value that is tracked by the particle swarm optimizer as the best value, is achieved at any point by any particle in the neighbourhood of the particle and this location is called Pbest. If a particle considers all the population as its topological neighbours, then the best value is a global best and is called Gbest.

Genetic Algorithm is employed as a computer model in which a population of conceptual representations (called chromosomes) of candidate solutions (called individuals, creatures or phenotypes) to an optimization complexity progresses near better solutions. Fitness function of the GA is determined by the ELM.

The ant colony optimization (ACO) technique has been inspired by the investigation of real ant colony’s foraging activities and on the principle that ants can frequently identify the shortest path between the food source and
the nest. An ant identifying an already laid trail can identify the thickness of pheromone trail. It chooses with a very high probability to follow the shortest path and strengthen that trail with its own pheromone. The huge quantity of pheromone is on a certain path and a greater prospect is that an ant chooses that path and the path’s pheromone trail will become stronger. Finally, the ant colony together plots the shortest path that has the highest pheromone density.

Table 3.1 Comparison of Reduction Rate

<table>
<thead>
<tr>
<th>Method</th>
<th>Total no. of Features</th>
<th>No. of Features Selected</th>
<th>Feature Reduction Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA with ELM</td>
<td>43</td>
<td>28</td>
<td>34.88</td>
</tr>
<tr>
<td>PSO with ELM</td>
<td>43</td>
<td>30</td>
<td>30.23</td>
</tr>
<tr>
<td>ACO with ELM</td>
<td>43</td>
<td>23</td>
<td>46.51</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>2550</td>
<td>112</td>
<td>95.6</td>
</tr>
</tbody>
</table>

Table 3.1 shows the reduction rates of the previous methods and the proposed method. The study of the table clearly implies that the proposed method has the highest reduction rate as compared to all other authentication methods. Figure 3.5 shows the comparison graph of the proposed method with the existing methods.

In the work entitled “An efficient feature selection technique for user Authentication using keystroke dynamics” by Shanmugapriya & Padmavathi (2011), it was noted that ACO with ELM produced Feature Reduction rate as 46.51% as its best. My present work has resulted 95.6% Feature Reduction rate with SVM with GA method. This research work was performed with increase in the total number of features and optimized the selected number of features and hence has resulted in the very good results of 95.6% feature reduction rate.
3.6.2 Rejection and Acceptance Error Rate

Rejection of Valid Users or False Non-Match Rate is the probability that a valid user's keystroke will incorrectly be considered a non-match for his reference template. Acceptance of invalid users or false match rate is the probability that an individual's template will incorrectly be considered a match for a different individual's keystroke sample. The error rate can be found from the following formula.

\[
Error \ rate(\%) = \frac{Total \ no. \ of \ errors}{Total \ no. \ of \ samples} \times 100
\]

Table 3.2 shows the values of the acceptance and rejection rates for both the proposed model of authentication system and the existing system. The existing system considered for comparison is Keystroke Authentication using Bayesian classifier.
### Table 3.2 Comparison of Error Rate

<table>
<thead>
<tr>
<th>Type of Error</th>
<th>Existing Method</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Sample</td>
<td>Errors</td>
</tr>
<tr>
<td>Rejection of valid users</td>
<td>539</td>
<td>44</td>
</tr>
<tr>
<td>Acceptance of invalid users</td>
<td>768</td>
<td>22</td>
</tr>
</tbody>
</table>

**Figure 3.6 Comparison of Error rate for rejection of valid users**
From Table 3.2 and Figures 3.6 and 3.7, it could be said that the proposed method is more efficient as the percentage of error rate is very less.

3.7 CONCLUSION

Security and authentication are the most significant challenges in computer systems or networks. Many techniques exist based on different biometric features like fingerprint and iris. But most of these can be easily cracked and the operational costs are very high. To overcome these drawbacks, keystroke pattern is identified. Keystroke features like dwell time, flight time, di-graph, tri-graph and virtual key force of every user are experimented in this paper. A GA based wrapper approach is applied to keystroke dynamics based authentication. As compared to the previous works, an additional input is introduced in the form of diversity of the population with the addition of a term in fitness function that measures the uniqueness of a chromosome. In this context, one class SVM is utilized as base classifier and forced diversity is obtained through the uniqueness of each chromosome. This renders a rather problematical post processing unwarranted. When the
aimed proposition and the other existing procedures are compared, the feature reduction rate is very high while the error rate is very small in the intended model. Accordingly, the planned scheme of human authentication system is efficiently helpful and successfully operational.