Analyzing User Behavior Using Keystroke Dynamics

Naveen Yadav[1], Saurabh Desale[2], Sufiyan Ansari[3], Ijaj Shaikh[4], Prof. Mrs V. C Sanap[5]

[1,2,3,4] Student, Department of Computer Engineering All India Shri Shivaji Memorial Society Polytechnic, Kennedy road, Pune, Maharashtra, India.
[5]Lecturer, Department of Computer Engineering All India Shri Shivaji Memorial Society Polytechnic Kennedy Road, Pune, Maharashtra, India.

---***---

Abstract – *Security is main concern for any organization. Security can be within the organization or outside the organization. To provide security within organization is major issue. So the keystroke dynamics is one of the techniques used in many applications to provide security. Keystroke dynamics is the detailed timing information that describes exactly when each key was pressed and when it was released as a person is typing at a computer keyboard. Work conducted in this field consists of linguistic context(To identify true user based on timing pattern of the user as pattern relate to language phenomena), typing rhythm(judging human behave or through typing speed),touch screen (mobile application) etc .In our system the insider attack is most devastating compared to an outsider and difficult to find as the complete knowledge of the underlying system is well known to the attacker. An Insider is the one who is having the authorized privileges within the organization. A malicious insider can create more damage as well as to the users by stealing the sensitive information. Behavioral bio-metrics is the field of study related to the measure of uniquely identifying and measuring the patterns in human activities. Computer security plays a vital role as most of the sensitive data is stored on computers. Keystrokes Dynamics is a technique based on human behavior for typing the password. Whenever any user logins into the system, username and password combinations are used for authenticating the users. The username is not secret, and the imposter acts as user to guess the password also because of simplicity of password, the system is prone to more attacks. In this case bio-metrics provide secure and convenient authentication. Our system uses a Euclidean Algorithm which is one of the best known classifications and regression algorithm. Researchers have proved that this algorithm will converge to the best possible solution in very less time.*

1. INTRODUCTION

Cloud computing is a general term for anything that involves delivering hosted services over the Internet and managed by the cloud service provider. The increased use of cloud raises the privacy concerns. Security of the data became the major issue as the data resides outside the user premises and managed by a third party. The insider attack is most devastating compared to an outsider and difficult to find as the complete knowledge of the underlying system is well known to the attacker. An Insider is the one who is having the authorized privileges within the organization. A malicious insider can create more damage to the cloud provider as well as to the users by stealing the sensitive information or by a data breach both to the name and fame . For example, a cloud administrator can access all the virtual machines of the users and can steal the sensitive information of the users without their intervention. A variety of risks has been identified is. Identity Management Systems (IDM) and Intrusion Detection System (IDS) are considered as useful tools for taking proactive measures against insider attacks. They can detect abnormal actions like packets containing malicious or unnecessary content and deviations in normal user behavior. IDS are of two types namely Host based IDS and Network based IDS which are used to isolate a host and track them. The tracking will be done either through available signatures or through anomaly using machine learning techniques by tracking the behavior of the attacks. The proposed work follows a host based user profiling technique to trace the user's behavior using a keystroke dynamics. One observation made in cloud is that most of the administration work involves command line interface rather than graphical user interface. Since the command line interface requires lot of key strokes, the proposed approach is well suitable for this environment. If the abnormality in the user behavior is detected, the system is locked so that a malicious masquerader cannot do any modification in the name of others.

Keystroke dynamics is a science of studying about keystrokes that differentiate each user based on their typing speed, latency between keystrokes, and pressure applied on keys etc. Key stroke dynamics fall under non-static bio-metrics which will vary with time. Non-static bio-metrics depends on several environmental, physical and biological factors. In contrast IRIS, finger prints, palm print etc comes under static bio-metrics which stays constant for longer duration but they require extra hardware to achieve it, which is not possible in a cloud based environment. To deal with the non-static biometric nature, different features are to be evaluated to attain proper results. The proposed work uses a Euclidean Algorithm which is one of the best known classifications and regression algorithm to date. Researchers have proved that Euclidean Algorithm will converge to the best possible solution in very less time.

2. PROBLEM STATEMENT :

To design a system to secure bank login for bank employees by differentiating each user based on their typing speed, latency between keystrokes and pressure applied on the keys. It falls under non-static bio-metrics which will vary with time.

3. SYSTEM SPECIFICATIONS :

3.1 Hardware Requirements:

Processor: Pentium 4, 2 GHz and above

RAM: 1 GB

Disk: 40 GB

3.2 Software Requirements :

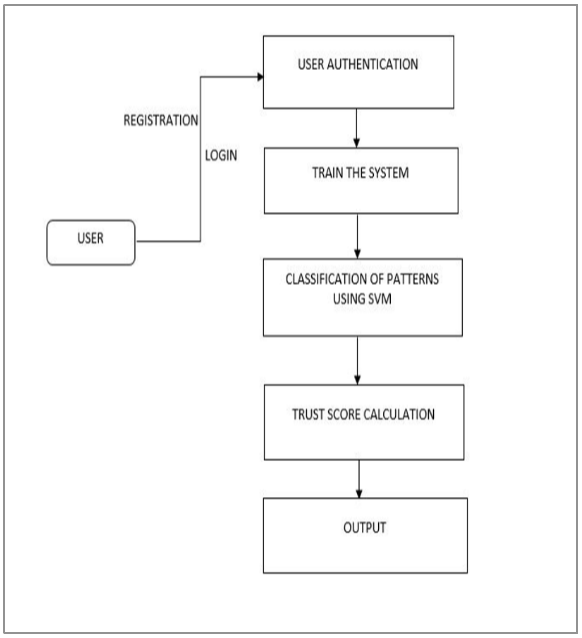
Front end : Java (JDK version 7)

Back end : MySQL IDE : Net Beans

Operating System : Windows XP/Vista/7

Documentation : MS-Office

4. IMPLEMENTATION AND DESIGNING



4.1 User Authentication :

If user is new user then registration for that user is done. Text containing all alphabets from A to Z and numbers 0-9 is used while registration so that pattern for each alphanumerical can be generated. If user is already registered then it's login details are asked to enter. Output of this module is account confirmation for new user and authentication for existing user.

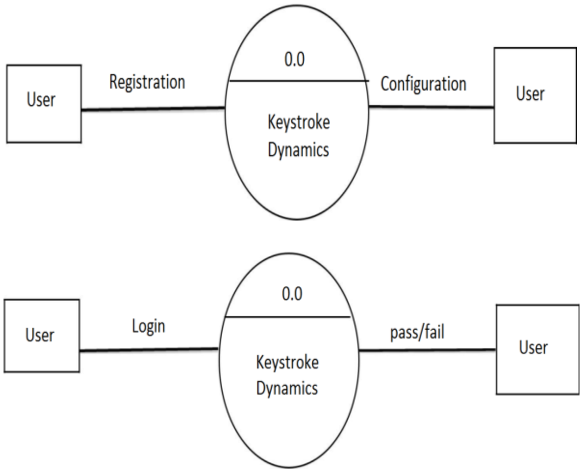


Fig. DFD Level 0 for Overall Process

4.2 Train the System :

Random text is selected from registration or login details. From this text patterns are generated like _x001D_right time, dwell time etc. These patterns are stored in database in specific string format. Every user has its own patterns stored in database.

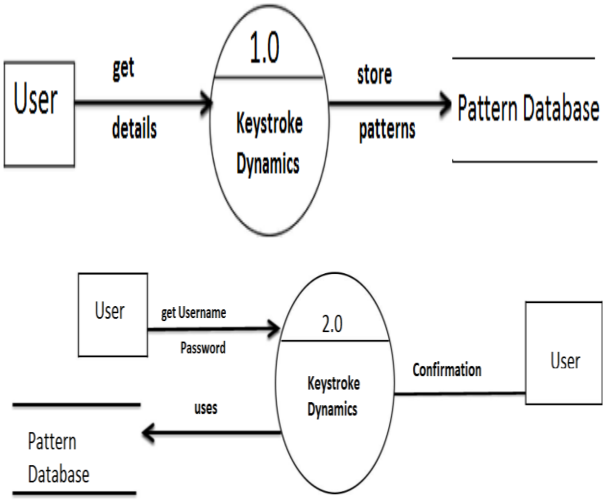


Fig. DFD Level 1 for Overall Process

4.3 Classification of Patterns :

Patterns are generated for the current login details. Patterns are matched with already existing patterns in database for particular database. If patterns are matched then user is valid .

4.4 Trust Score Calculation :

If patterns for particular user are not matched due to some reason then trust score is calculated. Trust score is compared with some threshold value. If value is more than threshold, user is valid.

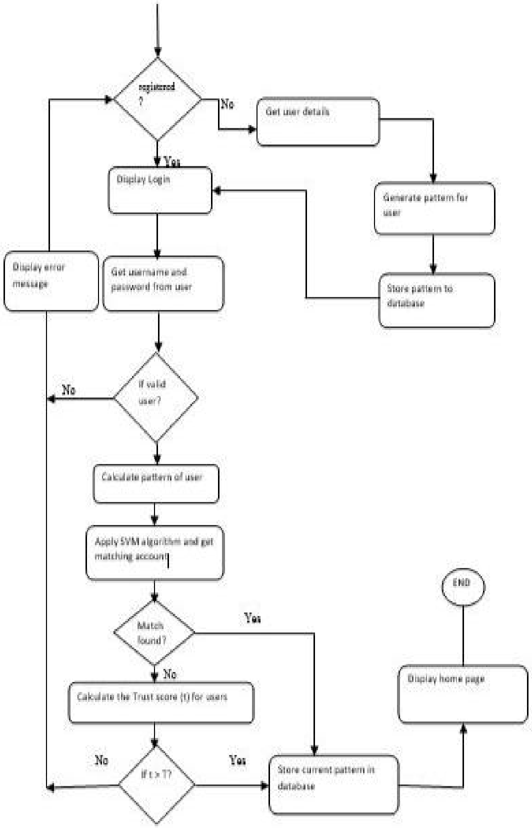

registered ?
No
Yes
Get user details
Generate pattern for user
Store pattern to database
Display Login
Display error message
Get username and password from user
If valid user?
No
Calculate pattern of user
Apply SVM algorithm and get matching account
Match found?
Yes
No
Calculate the Trust score (t) for users
No
If t > T?
Yes
Store current pattern in database
Display home page
END

Fig. Activity Diagram

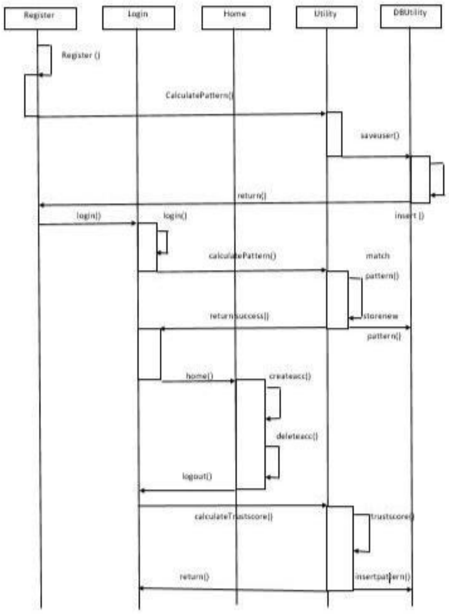


Fig. Sequence Diagram

4.5 Use Case Diagram:

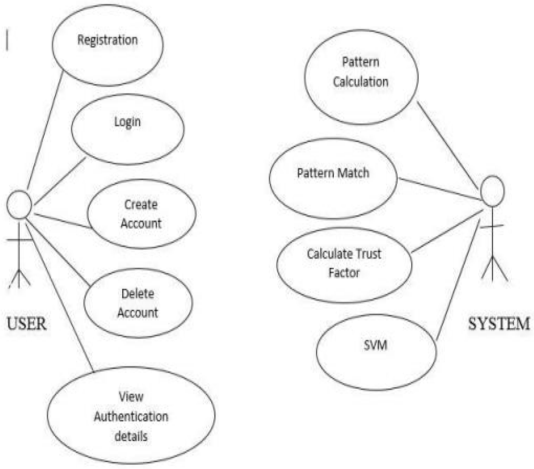


Fig. Use Case Diagram

5. CONCLUSION

So we have created a a system to secure login for hospital's receptionists by differentiating each user based on their typing speed, latency between keystrokes and pressure applied on the keys. It falls under non-static bio-metrics which will vary with time.

6. REFERENCES

[1] Mell, Peter, and Tim Grance. "Effectively and securely using the cloud computing paradigm." NIST, Information Technology Lab (2009).

[2] Rocha, Francisco, and Miguel Correia. "Lucy in the sky without diamonds: Stealing confidential data in the cloud." Dependable Systems and Networks Workshops (DSN-W), 2011 IEEE/IFIP 41st International Conference on. IEEE, 2011.

[3] Teh, Pin Shen, Andrew Beng Jin Teoh, and Shigang Yue. "A survey of keystroke dynamics bio-metrics." The Scientific World Journal 2013(2013).

[4] Ngugi, Benjamin, Beverly K. Kahn, and Marilyn Tremaine. "Typingbio-metrics: impact of human learning on performance quality." Journal of Data and Information Quality (JDIQ) 2.2 (2011): 11.

[5] Stolfo, Salvatore J., Malek Ben Salem, and Angelos D. Keromytis. "Fog computing: Mitigating insider data theft attacks in the cloud." Security and Privacy Workshops (SPW), 2012 IEEE Symposium on. IEEE, 2012.

[6] Rocha, Francisco, and Miguel Correia. "Lucy in the sky without diamonds: Stealing confidential data in the cloud." Dependable Systems and Networks Workshops (DSN-W), 2011 IEEE/IFIP 41st International Conference on. IEEE, 2011.