**M S Ramaiah Institute of Technology**

(An Autonomous Institute, Affiliated to VTU)

MSR Nagar, MSRIT post, Bangalore-54

**Literature Survey**

Title: **Alcatraz, an end-to-end SaaS product for complete computer and network security**

Team:

Akash Gupta 1MS13CS014

Ayush Kumar 1MS13CS034

Bineet Kumar 1MS13CS036

Deepak Jayaprakash 1MS13CS037

**Literature Survey**

Since, each feature of the product is a topic of research and development itself, the literature survey is specified with respect to a feature wise split of the product. We would start with Keystroke Biometrics and list the survey according to importance of the work and time of publication.

Keystroke Dynamics has become a widely researched and active area due to the increasing importance of cyber security and computer or network access control. Most of the existing approaches focus on static verification, where a user types specific pre-enrolled string, e.g., a password during a login process, and then their keystroke features are analyzed for authentication purposes [10]. Only a few research studies address the more challenging problem of keystroke biometrics using “free text”, where the users can type arbitrary text as input. Keystroke dynamics features are usually extracted using the timing information of the key down/hold/up events. The hold time or dwell time of individual keys, and the latency between two keys, i.e., the time interval between the release of a key and the pressing of the next key are typically exploited.

The features extracted from keystroke dynamics pattern in most of researches are timing features. Fig shows the extracted timing features:

1. Key Hold (KD): key delay between pressed and key released.

2. Down-Down Key Latency (DDKL): time in between two consecutive presses.

3. Up-Up Key Latency (UUKL): time between two successive releases.

4. Up-Down Key Latency (UDKL): time in between the current key release and the next key press.

5. Down-Up Key Latency (DUKL): time between the current key press and the next key release.

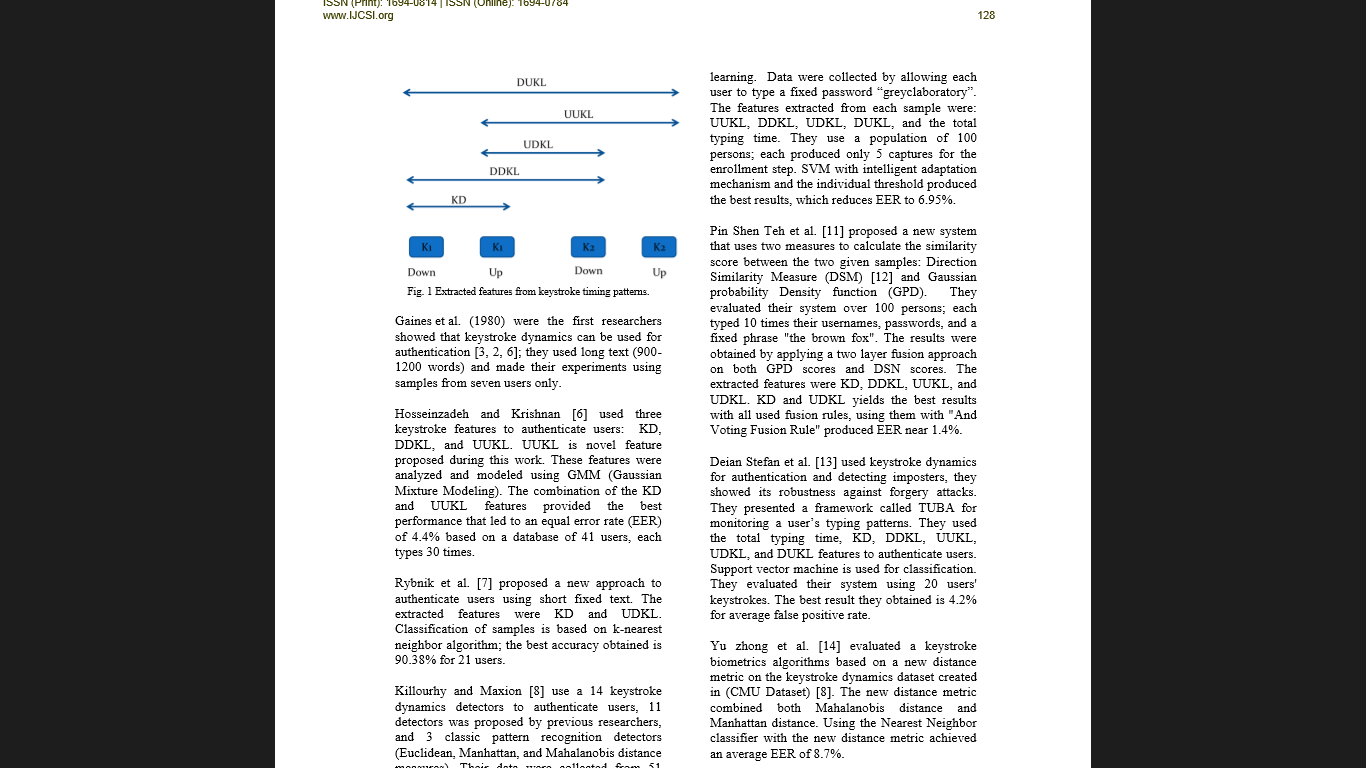


Fig 1: **Keystroke timing features**

Research work on keystroke dynamics all originated from Gaines et al. [16] who did a preliminary study authentication using the T-test on digraph features.

**Monrose and Rubin** few years later extracted keystroke features using the mean and variance of both **digraphs and trigraphs**. Then there were statistical Euclidean distance metrics with Bayesian-like classifiers identified 92% for their small dataset containing 63 users correctly. Over the years, keystroke biometrics research has been implemented in many existing machine learning algorithms and classification techniques.

Researchers have presented their work on choosing different distance metrics, such as the Euclidean distance, the **Mahalanobis** distance and the **Manhattan** distance and have been explored their suitability on the biometrc authentication. For the implementation both classical and advanced classifiers have been used, including K-Nearest Neighbours (**KNN**) classifiers [4], Bayesian classifiers, K means methods [12], **Fuzzy logic, neural networks** , and support vector machines (**SVMs**). A promising research effort in applying keystroke dynamics as a static authentication method originated from the work of Joyce and Gupta [15]. Their approach is relatively simple and yields impressive results.

The NSL-KDD data set is the refined version of the KDD cup99 data set. The original DARPA data set is not used because of a number of prominent reasons. The first important deficiency in the KDD data set is the huge number of redundant records. The existence of these repeated records in the test set, on the other hand, will cause the evaluation results to be biased. Another problem is that the models built over fit the data due to which it cannot identify any new attacks that occur.

Prominent data mining algorithms like Naïve Bayes, Decision Tree, K-nearest neighbor, K-Means and Fuzzy C-Mean clustering algorithms and machine learning strategies like Support Vector Machine (SVM), neural nets have been applied to study the classification.

The entire research can be classified into:

* **Machine Learning Based IDS:** Naïve Bayes, KNN, SVM and the proposed method multi Objective Genetic Fuzzy Intrusion Detection System (MOGFIDS), etc. Systems have been proposed that makes use of the analytical strengths of neural networks to detect stepping-stone intrusion with two schemes. Among the two schemes, one uses eight packet variables and the other clusters a sequence of consecutive packet in round-trip times.
* **The Multiple Classifiers (MCS)** approach was suggested based on pattern recognition distinct feature representation and tested with different fusion rules (Giorgio Giacinto et al 2003). The reported results proved that the MCS approach provides a better false alarm generation than that provided by an individual classifier trained on the overall feature set. Among the fusion rules, the dynamic classifier selection technique provided the best results.
* **Boosting based IDS:** Boosting is a method to develop the accuracy of several given learning algorithms by combining weak classifiers to get one strong classifier. Weiming Hu et al (2008) suggested an Adaboost based intrusion detection algorithm, in which the decision stumps are used as weak classifiers and the decision rules are provided for both categorical and continuous features. Jun Gao et al (2009) presented a distributed IDS framework based on the Model based online Adaboost algorithm and Particle Swarm Optimization-SVM algorithm.
* **Data Mining and Rule based IDS:** In data mining approaches, IDS comprises association rules and recurrent episodes, which are based on constructing classifiers by discovering appropriate patterns of program and user performance. The association rules and frequent episodes are used to study the patterns that illustrate user behavior.

**References:**

[1] KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[2] Xu, X.: Adaptive Intrusion Detection Based on Machine Learning: Feature Extraction, Classifier Construction and Sequential Pattern Prediction. International Journal of Web Services Practices 2(1-2), 49–58 (2006)

[3] Vipin Kumar, Himadri Chauhan, Dheeraj Panwar, “K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue4, September 2013

[4] Santosh Kumar Sahu Sauravranjan Sarangi Sanjaya Kumar Jena, “ A Detail Analysis on Intrusion Detection Datasets”, 2014 IEEE International Advance Computing Conference (IACC)

[5] Abdoul Karim Ganame, Julien Bourgeois, Renaud Bidou and Francois Spies, “A global security architecture for intrusion detection on computer networks”, Elsevier, Journal of Computers & Security Vol.27, No.1-2, pp. 30-47, 2008.

[6] Azzedine Boukerche, Kathia Regina Lemos Juca, Joao Bosco Sobral, Mirela Sechi Moretti and Annoni Notare, “An artificial immune based intrusion detection model for computer and telecommunication systems”, Elsevier, Journal of Parallel Computing, Vol.30, No.5-6, pp. 629–646, 2004.

[7] Abdun Naser Mahmood, Christopher Leckie, and Parampalli Udaya, “An Efficient Clustering Scheme to Exploit Hierarchical Data in Network Traffic Analysis”, IEEE Trans. on Knowledge and Data Engineering, Vol. 20, No. 6, pp. 752767, 2008.

[8] Ajith Abraham, Ravi Jain, Johnson Thomas and Sang Yong Han, “D-SCIDS: Distributed soft computing intrusion detection system”, Elsevier, Journal of Network and Computer Applications, Vol.30, No.1, pp. 81–98, 2007..

[9]L. C. F. Ara´ujo, L. H. R. Sucupira, M. G. Liz´arraga, L. L. Ling, and J. B. T. Yabu-uti. . “User authentication through typing biometrics features”, In Proc. 1st Int’l Conf. on Biometric Authentication (ICBA), volume 3071 of Lecture Notes in Computer Science, pp. 694–700, 2004.

[10] F. Bergadano, D. Gunetti, and C. Picardi, “User Authentication through Keystroke Dynamics”, ACM Trans. Information and System Security, 5(4), pp. 367–397, 2002.

[11] S. Cho, C. Han, D. H. Han, and H. Kim. “Web-based keystroke dynamics identity verification using neural network”, Journal of Organizational Computing and Electronic Commerce, 10(4):295–307, 2000.

[12] Yu Zong and Yunbin Deng, Anil K Jain, “Keystroke Dynamics for User Authentication”, Int’l Joint Conf. on Biometrics (IJCB), pp. 1-5, 2011.

[13] John Leggett and Glen Williams. Verifying identity via keystroke characteristics. International Journal of ManMachine Studies. Journal of Organizational Computing and Electronic Commerce, 10(6):, 2000.

[14]Kevin S. Killourhy and Roy A Maxion, “Comparing AnomalyDetection Algorithms for Keystroke Dynamics Proc. of the 3rd Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, pp. 61-64, 2007.

[15] R. Joyce and G. Gupta, “Identity authentication based on keystroke latencies”, Communications of the ACM, 33(2):168–176, 1990.

[16] K. S. Killourhy and R. A. Maxion, “Comparing Anomaly Detectors for Keystroke Dynamics”, in Proc. 39th Annual Int’l Conf. on Dependable Systems and Networks (DSN2009), pp. 125-134, 2009.