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

**Dr. Annapurna P Patil J Geetha Reddy**

Professor, Assistant Professor,

Department of Computer Science and Engineering, Department of Computer Science and Engineering,

Ramaiah Institute of Technology Ramaiah Institute of Technology

Bangalore - 560054 Bangalore - 560054

**Ayush Kumar Bineet Kumar**

Student, Department of CSE, RIT Student, Department of CSE, RIT

**Aman Mehta Deepak Jayaprakash**

Student, Department of CSE, RIT Student, Department of CSE, RIT

***Abstract--***The demand for modern tools and techniques to restrict access to applications and services which contain delicate data is increasing exponentially each year. People can no longer rely on older methods of computer and network security as they fail in dealing with the challenges posed.

The product is aimed at providing an architecture and open source code to developers so that they can embed these features into their applications to enhance the security. The services provided are top notch and cover the broad spectrum of computer and network security. All the features of the product involve the application of **Data Mining** and **Machine Learning** techniques onto the domain of **Computer Security**.

The most prominent features of the product are listed below:

- **Keystroke Biometric Authentication**: Ensembles result of 3 statistical models and classifies user based on Dynamic Thresholding

- **Network Intrusion Detection:** Data mining on NSL-KDD data set, build 3 models, accuracy comparison and data visualization.

- **Anti-key logging:** Uses authentication based on GET request exchange and adds a double layer of AES + Hashing to guarantee both Data Integrity and Security.

- **O-Auth with Hashing:** Traditional O-Auth modified with additional layer of Hash to impart Digital Signature.

The complete product with the features is based on dynamic-dispatch, a kind of “on-demand software” inspired ideology, hence the term SaaS. The entire product is released as open source with detailed documentation on GitHub. It is released as a stable live version with a guide for developers on how to contribute.

The product when compared to existing software applications like Snort, SpyShelter and other biometrics have advantages like

- No external hardware

- One complete encapsulated bundle

- Open source code

- Cloud Integration

- Detailed documentation and support

*Keywords: Keystroke dynamics, zero variance removal, O-Auth, Network Intrusion Detection, Anti-key logging, Statistical modeling*

1. INTRODUCTION

The biggest necessity of a product of this nature arises from the fact that developers prioritize attributes such as functionality, UI, memory and efficiency while designing and creating applications of both small and large scale. But they comparatively turn a blind eye or assign a low level priority to security. This is understandable because making an application/ service secure requires a lot of programming effort and expertise. The objective of the product is hence to provide an overall architecture, design and code to application developers so that they can embed these onto their modules.

The features are also on a demand and get basis so only those services which are required can be taken. Each feature is an application of statistical modeling and ML algorithms on layers of computer security. The services use highly efficient mechanisms ranging from Dynamic Thresholding, Zero-variance removal, AES, Hashing, Data Mining to Ensemble learning, asynchronous communication and accuracy analysis.

The entire product is versioned and released on GitHub. Each feature is a novel implementation because all of them are not a single model-classifier service. They are an ensemble learnt multi classifier accumulation, wherein each classifier/ statistical model is designed to outperform one another so that the drawbacks of one are overcome in another thereby leading to a more accurate result. Enough mathematical support is given to all the algorithms used. Some of the concepts used are linear programming, statistical mean, median, variance, k-cross validation, iterative averaging, etc.

1. LITERATURE SURVEY

The features offered in the product belong to well researched and hot topics lying in the domains of Biometrics, Data Mining, Machine Learning, Statistical modelling and Computer/ Network Security. Prominent work has been done in the field of keystroke biometrics. This is because of the rising significance in the area of digital signatures, data integrity and security. Many existing approaches implement the static text method of keystroke dynamics [5]. The existing live implementations for “free text” are very hard to find, which can be attributed to the low accuracy rate and other problems like bias, over-fitting, false recognition rates and non-significance

First true work on keystroke dynamics originated from Monrose and Rubin [23]. Their learning models were implemented by using statistics like digraph/ trigraph mean and variance

Researchers have presented their work on choosing different distance metrics ranging from Manhattan to Euclidian and combined them with mean, median and tested their suitability on the biometric authentication. For the implementation both traditional and modern classifiers have been used, including knn[5], Bayes Point Machines, K-mediods/means methods [11], genetic algorithms, neural networks , and SVMs if the attributes are non linearly dependant.

Network Intrusion Detection happens to be one of hot topics in the domain of Network (both wired and wireless) security. 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 existing work can be broadly classified into:

- Machine Learning Based IDS

- The Multiple Classifiers (MCS)

- Boosting based IDS

- Data Mining and Rule based IDS

There has been quite few implementations of Anti-Key logging systems and have been offered as service or an application supporting some Operating Systems. The research done on this include but is not limited to Behavior based detection technique using KLIMAX: Kernel- Level Infrastructure for Memory and execution profiling, Anomaly Based Detectors, Signature based detectors, Subnet Mask Scanners and Host Analyzers.