**M S Ramaiah Institute of Technology**

(An Autonomous Institute, Affiliated to VTU)

MSR Nagar, MSRIT post, Bangalore-54

**Synopsis for final year project**

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

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

1. **Problem Statement**

The product is developed as a part of the final year project. It 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 (as of now) include **Keystroke Biometric Authentication, Network Intrusion Detection, Privacy respecting Cloud, Secure File/Mail transfer**, etc.

Each feature is a problem statement and a topic of research by itself. But to simplify things, we build the applications using web technologies to demonstrate the overall architecture.

**Keywords:** Data mining, ML, computer security, biometrics, classification, adaptive dynamics, Leave One Out Method,(LOOM), signature-based, IDS, etc.

1. **Abstract**

The demand for modern tools and techniques to restrict access to applications and services which contain delicate data is increasing exponentially each year. Traditional methods such as PINs, tokens, or passwords fail to keep up with the challenges presented because they can be lost or stolen, which compromises the system security. But even the most powerful cryptographic systems fail to prevent unauthorized access.

Developers need to address the issue of security at each module while developing applications involving customer info. The aim of the project is to provide the basic features needed to make a system secure. Hence, we have services ranging from Biometric Authentication to Network Intrusion Detection Systems.

Here is a fairly detailed description of the services offered as a part of the product.

* **Keystroke Biometric Authentication:** Biometrics, defined as the physical traits and behavioral characteristics that make each of us unique, are a natural choice for identity verification. Biometric attributes become the most optimal and ideal candidates for authentication since they cannot be stolen, lost or impersonated. They are based on “who” is the person or “how” the person behaves, present a significant security advancement to meet these new challenges.

The most promising approach has been Keystroke biometrics which refers to the habitual patterns or rhythms an individual exhibits while typing on a keyboard input device. Compared to other biometric schemas, keystroke has the primary advantages that:

1. No external hardware like scanner or detector is needed. All that is wanted is a keyboard.

2. The “rhythm” or the pattern of the users is a very reliable statistic.

3. It can easily be deployed in conjunction with existing authentication systems.

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 threshold calculation is what makes the **model adaptive** and different than other existing models and algorithms. The window for error is the space in which he is permitted to cause any errors. This is decided by a method called **Leave One-Out-Method (LOOM).**

* **Network Intrusion Detection**: In the era of information society, computer networks and their related applications are becoming more and more popular, so does the potential threat to the global information infrastructure to increase. Therefore, the role of Intrusion Detection Systems (IDSs), as special-purpose devices to detect anomalies and attacks in the network, is becoming more important. IDSs are impressive since it can detect, prevent and possibly react to attacks in an efficient manner.

In this feature we shall be providing an architecture for the developers to build potential algorithmic models like Naïve Bayes, Decision Trees and Random Forests along with detailed code for each of the steps involved like

1. Data set Description
2. Classification of attributes
3. Explanation of attacks
4. Pre-processing and dimensionality reduction
5. Building classifiers and testing results
6. Data visualization

* **Privacy respecting Cloud:** Cloud storage is easy to come by. Dozens of services shovel tons of free space just for signing up. But which of those services are looking at the files that are uploaded, and most importantly, which services encrypt your personal data so no onecan look at it. Privacy is even more important when it comes to cloud storage. Trusting the service you sign up for to keep your files safe and secure and away from prying eyes is what is needed. Whether you use your cloud storage for music, tax returns, or backups, it's still important to know that your provider isn't rifling through your files to make sure the music isn't pirated.

Examples for such infrastructure are SpiderOak, Wuala, MEGA, etc.

But the primary problem with these clouds is that they are all proprietary and their architectures are not disclosed. Hence, we provide a general and simpler one to demo how the internal schema works in this part of the project.

These are the 3 primary features of the product. Other features like Secure Mail/File transfer, etc will be added on to the existing scope based on the progress.

1. **Technologies used:**

The base of the project lies on applying the algorithms of **Data mining, Machine Learning and Statistical Modeling** onto the domain of **Computer Security**. A detailed description of tech used is given below:

Tech used in

* **Keystroke Biometrics:** Statistical models like Euclidian, Manhattan and **LOOM** for threshold calculation.

Tech stack used:

* Front end tech: HTML5, CSS3, Bootstrap, jQuery
* JavaScript and Ajax: to gather the key up time from the user and parse it.
* PHP and R: To perform file operations and build models
* Mysql: Database operations
* **Network Intrusion Detection:** Signature based anomaly methods, elimination of redundancy, pre processing and attribute selection, implementation of classification algorithms( Naïve bayes, Decision Trees and random Forests) and data visualization.

Tech stack used:

* Front end tech: HTML5, CSS3, Bootstrap, jQuery
* Back end: PHP+ Mysql
* R libraries: library(caret), library(e1071), library(randomForest), library(rpart , rpart.plot), etc

The tech stack used for other features would be similar to what is listed above.