

SECURING WEB APPLICATION AGAINST BRUTE FORCE ATTACK USING CONTINUOUS AUTHENTICATION

Project Report in partial fulfillment of the requirement for degree of
Bachelor of Technology

in

Information Technology

By

(Rohan Tyagi (171463))

Under the supervision of

(Mr. Rizwan Ur Rehman)

to



Department of Computer Science & Engineering and Information
Technology

**Jaypee University of Information Technology Waknaghat, Solan-
173234, Himachal Pradesh**

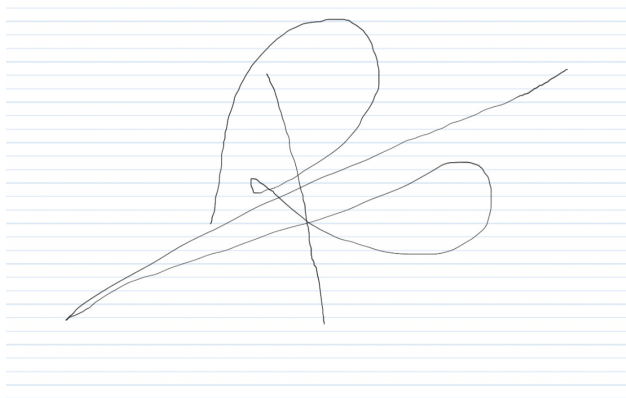
Certificate

Candidate's Declaration

I hereby declare that the work presented in this report entitled “**Securing Web Applications Against Brute Force Attack Using Continuous Authentication**” in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Information Technology** submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Wagnaghat is an authentic record of my own work carried out over a period from August 2020 to December 2020 under the supervision of **Mr. Rizwan Ur Rehman**.

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

This is to certify that the above statement made by the candidate is true to the best of my knowledge.



(Supervisor Signature)



Rohan Tyagi
(Student Signature)

ACKNOWLEDGEMENT

Many efforts have been made to complete this project successfully. But, in order to make this project become a reality, it has taken the kind support of many individuals and resources as well as the help of my colleagues and superiors. It would be an immense pleasure to extend my heartiest and most sincere thanks to all those who were involved in this project directly or through me.

I am highly grateful to Mr. Rizwan Ur Rehman, keeping in mind her keen guidance and constant supervision and support throughout the process of development of this project. I am also thankful to her for providing with all the necessary information and guidance which was required throughout the duration of this project.

I would also like to convey my gratitude towards my peers from Jaypee University of Information Technology, who has been a constant support to me throughout the project. I would also like to appreciate their kind help whenever I required it.

Table Of Content

SR.NO	CONTENT	PAGE NUMBER
1	INTRODUCTION	6-7
2	LITERATURE SURVEY	8-29
3	DATA COLLECTION	30-39
4	API AND DATASET	40-47
5	DATA PROCESSING AND RESULT	48-49
6	CONCLUSION	50-51

ABSTRACT

This project “Securing Web Applications against Brute Force Attack Using Continuous Authentication” is aimed to contribute towards cyber-security. This project describes how fraudsters try to get illegal access to web portals using brute force, different types of brute force attacks and how continuous authentication can prevent illegal access.

Internet users are concerned about the security over internet. Due to which cyber-security has become need of time. Fraudsters can easily trespass the security set-up established by web server. Cyber laws implemented by various federal governments are only effective when helped be spam and theft preventive mechanisms.

Brute force is one of the methods used by fraudsters to carry out malicious activities. It is important to study the types and structure of brute force.

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Brute Force Attack is the concept of trying different usernames and passwords over a platform by hackers in order to get illegal access to data. It can also be specific for a particular targeted account. There are different types of Brute Force Attack. Continuous authentication is a method of verification aimed at providing identity confirmation and cyber-security protection on an ongoing basis.

1.2 OBJECTIVE

The main objective of this project is to develop a model that is to contribute towards cyber-security. Machine Learning is an important segment of data science, this is a reason we have used this technology to predict authenticity of a user using a developed model. Behavioral patterns are also used to predict the category of a user.

1.3 MOTIVATION

We chose this project to contribute a model that could help digital world in preventing frauds and scams. It would help users to maintain a secure and safe connection with their server portal. Our model will also ensure efficiency, thus making connectivity between server and user further stronger. We suggest a method that could be used by cyber-police to prevent crimes using illegal and unauthentic access to end user's private data.

1.4 METHODOLOGY

Research related to brute

Preventive measures to brute

Continuous authentication

Keystroke Dynamics

API

Dataset

1.5 **DELIVERABLES OF THE PROJECT**

We will develop a model that would be helpful in preventing illegal access to web applications by fraudsters.

CHAPTER 2: LITERATURE SURVEY

BRUTE FORCE ATTACK

A web application is software that operates over a web server. It uses a client-server application structure. The main services of web applications are hosted by an external server. Online banking, web-mail, online transaction and online retail applications. Web applications have become necessity for internet users. Fraudsters or hackers can use tricks to get unauthorized access to web server. Therefore it is necessary to use effective measures to secure web application servers.

Brute Force Attack is the concept of trying different usernames and passwords over a platform by hackers in order to get illegal access to data. It can also be specific for a particular targeted account. It works by trying different possible combinations for password. Leak of information or stolen database makes it easier for attackers to get access. Number of failed attempts to get access depends on the password's length and server's security. Most online servers and portals implement countermeasures to prevent such attack. This leaves attackers with very least opportunity to get access. Fraudsters want to get access steal and utilize personal or financial data of an authorized user from their authorized online accounts or from server resources.

Brute force attack can be used for various purposes such as spreading false news, carrying out phishing attacks or selling information to third party sources. Motive such attack depends on the fraudsters who carry out such attempts. It has also become necessary to determine the possible the motive behind attack. Motive behind the attack can help us discover the loop holes in our connection.

Time taken to guess correct pair of password and username determines the success rate of brute force attack. Less time for attack indicates the weakness in the security measures established by the service providers. Encryption level and password-type restrictions should be of a level that leaves less option with fraudsters to guess the correct username and password combination for a particular user.

If the password is strong and has a good amount of random characters in it, then it might take more than million attempts for an attacker to guess the correct pair of ID and password for a particular user. This forces the attackers to try some different and unique techniques to guess particular credentials. They can also use techniques of social engineering to crack the correct password. Due to large amount of time consumption and large number of continuous attempts it attackers have started using different techniques to make the process of guessing easier and quicker. They have started using techniques such as Rainbow, Brutus THC Hydra, etc to make guessing more computerized.

Dark web provides some combination of username and passwords over internet which reduces the effort required by fraudsters. Relation between time of guessing a password and length of the password is considered as exponential by most analysts. Due to complexity and difficulty of this attack, attackers mainly use automaton or AI based technique to make a successful attempt.

Data stored in database should also be safe and secure as it can be common target center for fraudsters to know credential of users. There are different categories of attacks that we need to study and analyze the preventive technique that would be strong enough to prevent an illegal access.

Bots use can ease down the difficulty to get access. Attacker need to focus on the programming structure of the bots which in turn will conduct the actual attacks. Bots can make use of latest emerging technologies such as Artificial Intelligence, Machine Learning and other data science techniques to predict the possible password on the basis of historical or periodic data provided to them. This data collection can also be done using various programming and data science techniques. Supervised learning can efficiently be used to predict the possible character set in login credentials of a particular user. Fraudsters use bots as it is also provides a cover against cyber police. It makes it difficult to trace a cyber-criminal when he makes use of automated bots rather individuals making random guesses for credentials of user. It is suggested to freeze the account of a particular user when request to access is made multiple times and even beyond a specified limit. However, this can lead to unwanted results and inconvenience for users whose account is freeze.

MOTIVE BEHIND BRUTE FORCE ATTACK

Stealing Personal Information

This attack is mostly used in stealing personal confidential and financial information. It can also be used to steal money and carry out financial transactions. End user of a service provider is mostly affected due to this as it directly affects the user. This information can be further used to steal data from user's account and carry out financial transactions. However, monetary loss is also dependent on other credentials such as CVV and OTP. User's awareness is also important to prevent this loss. Brute force attack along with phishing attack is used to conduct illegal financial transactions from user's account to a fraud account. Users need to avoid sharing information related to user's debit/credit card details as this data can be retrieved from database.

Damaging Reputation

Brute force attack can be used to damage reputation of a competitor or popular brand of particular web service provider. A hacker can also cause harm by altering content of a web portal. This approach is not very common, as service providers have a secure defensive mechanism to deal with such attacks. This can also lead to violation of copyright laws which can also harm reputation of the service provider.

Spread Fake Content

Attackers can use brute force to spread fake content. This can be done to various motives. Spreading fake content can lead to disastrous consequences in a civilized society. This type of attack is very common. Fake content uploaded over a trusted site can also harm the reputation of service provider. Cyber crime branch of India has become very active in recent past to deal with spread of fake content. Attackers can also upload spam or malicious content which violates specific policies of portals. New and better cyber-laws are needed by the society to deal with spread of fake content over internet. Fake news spread can also harm communal harmony in the society so it is important to keep a check on content. Service providers also should add multiple layers of encryptions and security to prevent this spread of malicious and fake content over internet.

Profiting From Advertisement

Various agencies hire hackers who can breach through security and promote their products over various platforms. They can either directly post their content or use a redirect link to third party sources. It has become a very profitable industry. Content for advertisement can be spam or misleading which can also prove to be harmful for the reputation of a reputed service provider or brands. Commission agents use bots to breach through the security and post a particular advertisement.

Phishing Attack

Brute force attack can be used to carry out phishing attack. Phishing is method which uses various methods of telecommunication to steal personal and financial information of user's of trusted service providers. Using brute force attack fraudsters can promote false and malicious websites which are either directly used by attackers or operated by third party organizations who want this information for personal benefit. Brute force attack can also be used to carry out phishing attack using email. This is also known as email-fraud or CEO fraud. Information that is confidential and can cause monetary loss might get leaked which can incur tremendous loss to an employee or an organization.

Self-Check

Brute force attack can be used to conduct a survey about security and encryption level of a web portal. This tells an organization about the strength of security level set-up of its server. Further results of graph of time taken and strength of password can be used to study the security level of server. This also tells about the behavior of brute force attack. This will help a company to maintain security level for end user thus providing a safe connection.

TYPES OF BRUTE FORCE ATTACK

There are various types of brute force attack that are used by fraudsters. Some of the most common types are:

SIMPLE BRUTE FORCE ATTACK

Attackers make random guesses to get through the log in window to get access. This is generally done without help of any software or programming logic. This attack can succeed if login accounts are protected by weak and simple patterns. This type of attack is very simple and it does not involve any logic. It will be only successful when password is either too weak or there is a leak of password either from user or service provider's end.

For Example:

Username-> "RohanTyagi";

Password -> "Rohan12345"

This can be prevented either by user awareness or by imposing restrictions on passwords by the server administrator. Server side administrator can also suggest some password combinations which a user can use if required.

CREDENTIAL STUFFING

If the username and password combination for a particular user works over a website then attackers try this combination for other websites. There is possibility that user has used same set of login credentials for all websites.

For Example:

AMAZON

Username-> "Rohan1881@amazon"

Password-> "123\$abc"

PAYTM

Username-> "Rohan1881@paytm"

Password->"123\$abc"

GMAIL

Username-> Rohan1881@gmail"

Password->”123\$abc”

This depends on user’s discretion whether he uses same passwords for multiple accounts. Service providers can also instruct user’s to use unique password or a password that has not been used previously.

HYBRID BRUTE FORCE ATTACK

In this type of attacks passwords are mixtures of some common logical words and random characters. In this hit and trial attack, chances of making a hit are high as most users prefer a hybrid mixture of logical words and irregular characters.

For Example-

Username-> “Rohan1881”

Password-> “Rohan1881_&zuq1”

Username->”Rohan4554”

Password->”Rohan4554_kqU%p2”

Hybrid brute force attack can also be done using mixed passwords used by a user over different websites. Hybrid brute force is time consuming but it is one of the most efficient techniques for a brute force attack. Users generally prefer to keep a hybrid password this is also a reason why this technique is more useful

REVERSE BRUTE FORCE ATTACK

It is the reverse of simple brute force attack. Attackers use available leaked passwords. They search for the usernames until they find a match for the available password.

For Example:

Password-> “zqu1321_ui\$@”

Username1-> “Rohan1881”

Username2-> “Rohan1991”

This method is easier to implement than other methods. Guessing a username is much easier than guessing a password for a particular username credential. But this method is least preferred by attackers as password is not easily available on internet and guessing password is main task that has to be performed by attackers.

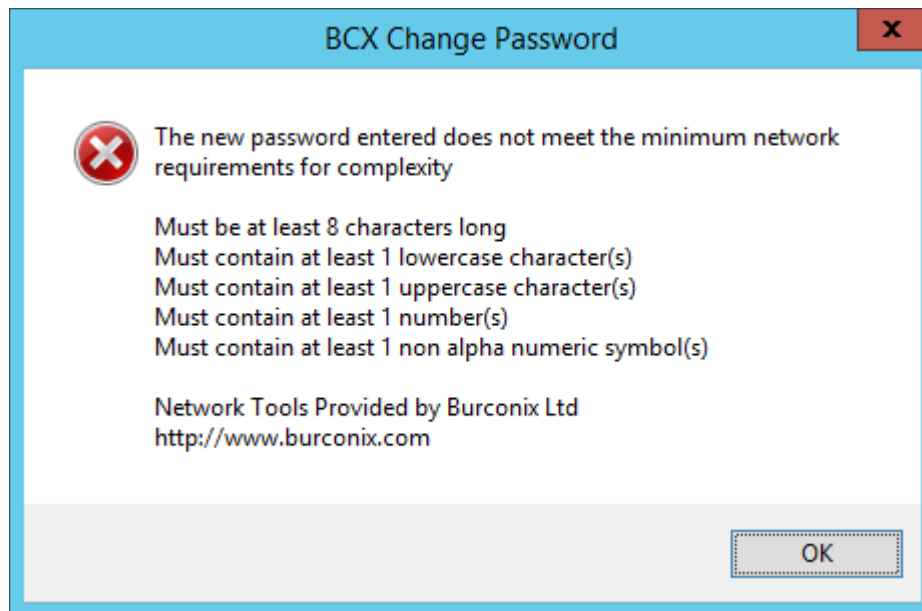
DEFENSE AGAINST BRUTE FORCE ATTACK

Brute force attack is a time consuming concept. There are various methods that can be used to prevent or at least add complexity to the process of attack. Most service providers over internet have made use of these mechanisms compulsory for their users.

Some of the common methods are:

Password Complexity

Complexity of valid passwords allowed can be increased by making it necessary to put each upper case, lower case, numeric and special characters. The end user must be forced to comply with these restrictions in order to create an account. The time to guess correct password increases exponentially with password complexity.




Type of restrictions is to be decided by the web application service provider. Complexity increases the strength of password. Stronger the password is lesser are the chances of hit by bots to get access.

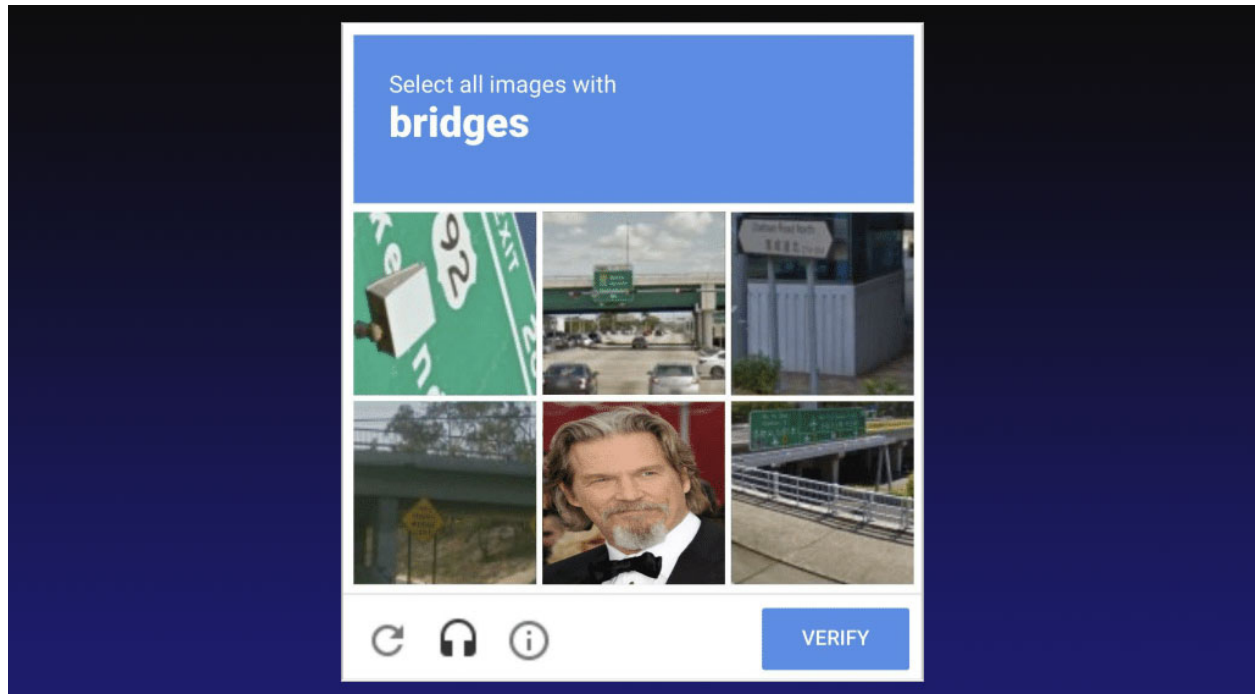
CAPTCHA

CAPTCH stands for Completely Automated Public Turing test makes sure that only humans are attempting to breach through Login window. This defies the chances of using software tools. CAPTCHA is used by most web applications, websites and servers that require login credentials such as username and password. CAPTCHA is very powerful when used with login time.

Please check the box below to proceed.

 I'm not a robot 
reCAPTCHA
Privacy - Terms

CAPTCHAs are also based on images. User has to logically identify the mentioned images as specified. Users are allowed specific number of attempts to correctly identify the correct squares in the given image.



CAPTCHA is a very powerful tool. However, bots can be programmed to identify squares. Some data science techniques and Artificial Intelligence can be used to identify the images.

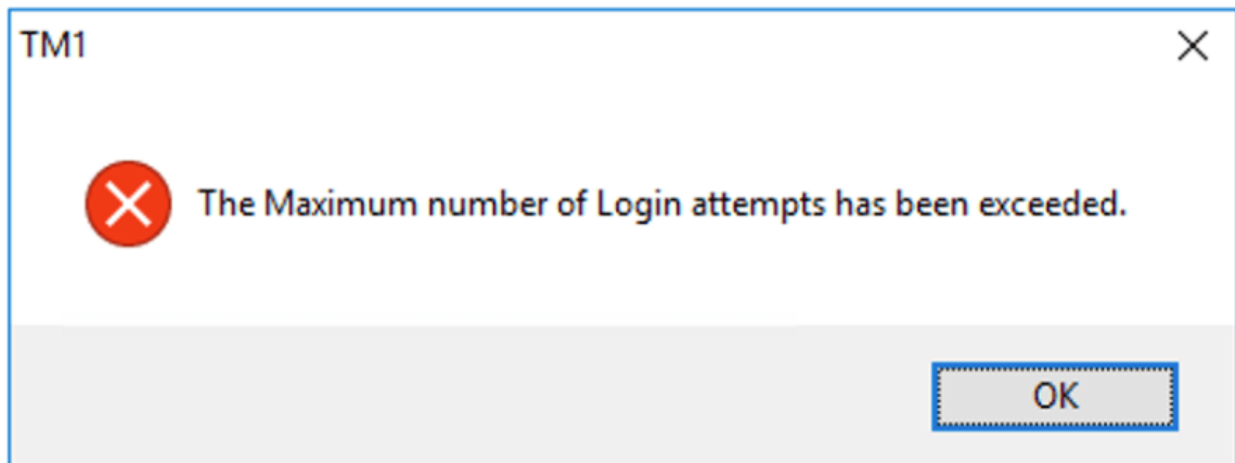
Increasing Password Length

Increasing password length can increase the time required to crack password for a particular user. Service providers can set particular limit below which a user is not allowed to keep password. It is suggested to choose a password that is long enough to prevent any illegal brute force attack attempt.

Length of Password	Combinations	Time to Crack (yrs)	Time to Crack (s)
4	456976	0.0	0.000228488
5	11881376	0.0	0.005940688
6	308915776	0.0	0.154457888
7	8031810176	0.0	4.015905088
8	208827064576	0.0	104.4135323
9	5429503678976	0.0	2714.751839
10	141167095653376	0.0	70583.54783
11	3670344486987780	0.1	1835172.243
12	95428956661682200	1.5	47714478.33
13	2481152873203740000	39.3	1240576437
14	64509974703297200000	1022.8	32254987352
15	1677259342285730000000	26592.8	8.3863E+11
16	43608742899428900000000	691412.1	2.18044E+13
17	1133827315385150000000000	17976714.2	5.66914E+14
18	29479510200013900000000000	467394568.1	1.47398E+16

Login Attempts

Limitation should be applied on Login attempts. After specified number of failed attempts, login access can be temporarily blocked. This can prevent illegal access by limiting the number of trails an attacker can make.



Beyond this limit user should not be allowed to get access to his account and user might have to try the account recovery mechanism to grant access.

CONTINUOUS AUTHENTICATION

Continuous authentication is a method of verification aimed at providing identity confirmation and cyber-security protection on an ongoing basis. Login-based authentication checks a user's identity only once, at the start of a login session, continuous authentication recognizes the correct user for the duration of ongoing work.

Continuous authentication is able to spot the moment at which an unauthorized attacker seizes control of the session, immediately ending the session, logging the account out, and protecting critical systems and data. Continuous authentication is implemented using machine learning (ML) and a variety of factors including behavioral patterns. Continuous authentication functionality constantly collects information about a user's actions and patterns of regular behavior and learns to distinguish between normal and abnormal behavior of a user based on the collected data.

Based on analysis of user behavior, access to a system can be granted or additional user identity verification can be requested. Variances and inconsistencies in behavior and user interaction with a system can be measure or a user's physiological characteristics can be identified continuously during the session. Additionally, if a user behaves badly or is compromised, then access can be revoked and application session ends immediately.

Possible methods of spotting changes include using keystrokes, video, fingerprints, touch (the amount of finger pressure applied) or facial features like eye position, pupil size and how often someone blinks. An application with continuous authentication functionality can continually compute a score to determine how certain it is that the account owner is also the one using the device.

Depending on the score, the user might be prompted to input additional information such as a password, card or fingerprint. Authenticity can be determined either by some continuous or discrete values.

There are various methods used for continuous authentication. Some common authentication techniques are:



Behavioral attributes

Behavioral patterns or gestures such as the typing speed of the user, finger pressure, the way a user moves his mouse, etc. are noted. In this project we have used these features to generate dataset. Then we will apply machine learning algorithms to determine the authenticity of a particular user.

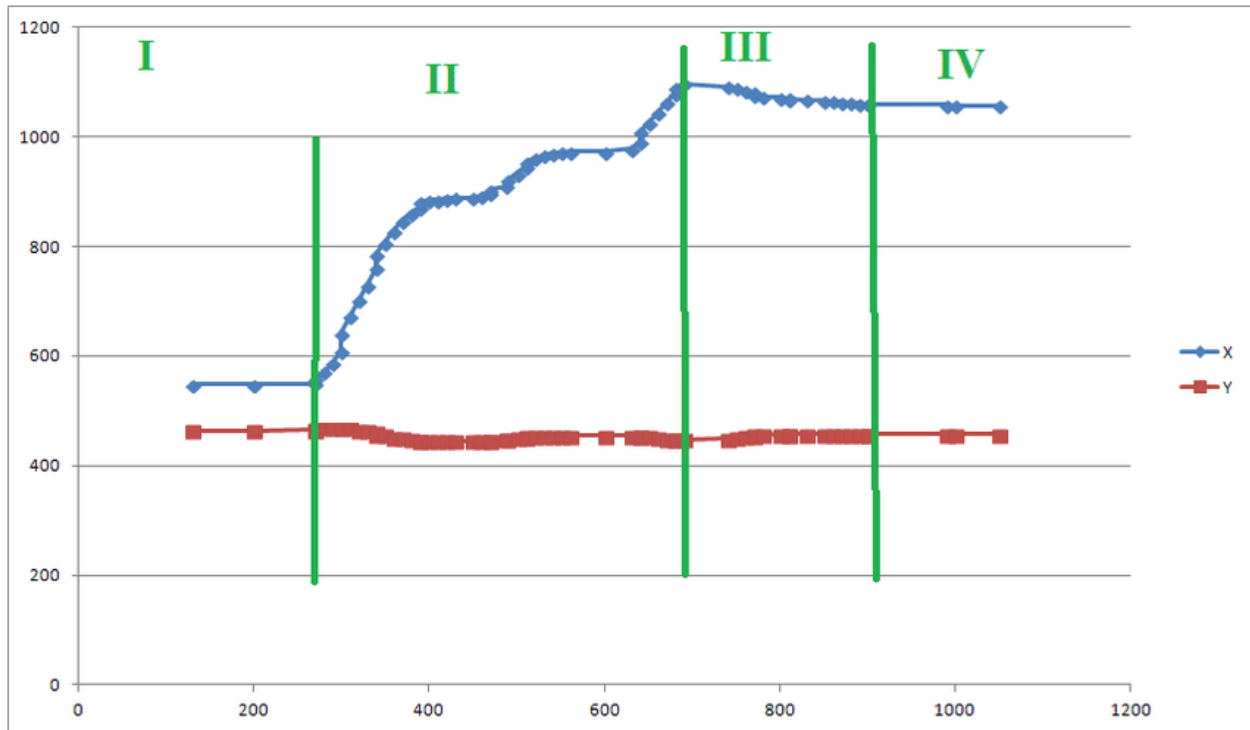
Typed Text	Characters
Average tweet length	60-70
Average sentence	75-100
Phishing email	120
Average Facebook post	155
Maximum tweet length	280
Gettysburg Address	1450
Nigerian prince emails	1500-2500

Any suspicious trends observed in this data can be recorded and observations can be made for behavioral changes over a user's account.

Physical movement

The sensors track the physical movement of the user; the way a user positions his/her device, the speed at which the user moves the mouse pointer, etc. However, these movements and records cannot be fully trusted as physical movements are very random i.e. these movements of user can change with change in time. Physical Movements also fail to determine authenticity when a user himself has allowed an individual to use web service because for any change in physical movement, the administrator will look at it in a suspicious way.

There are many disadvantages of tracking physical movement. However, an attacker cannot replicate physical movement of a user. Due to this reason physical movement is included in the study to determine brute force attempts.



Here we have an image showing trajectory of mouse movement in X-Y axis with some data points showing mouse-clicks.

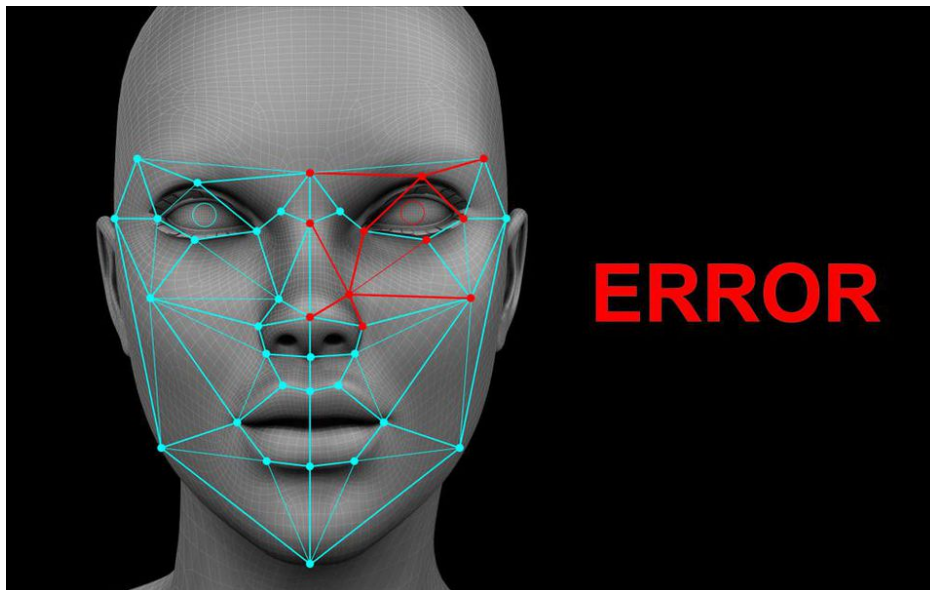
There four phases for change in behavioral pattern of physical movement of mouse in the given diagram.

Facial recognition

Most devices have this feature for authentication. However, Continuous Authentication follows the way a user glances at his device or the facial expressions he gives while unlocking the device. Artificial Neural Networks, Machine Learning and Artificial Intelligence are used to map the face of a user. Various network points are plotted which are used to recognize the face of an authentic user.



This is more reliable as it is unique for every user. But in some cases it can cause inconsistency such as for twins. This also depends whether a user has allowed access to his front camera or not.



Voice recognition

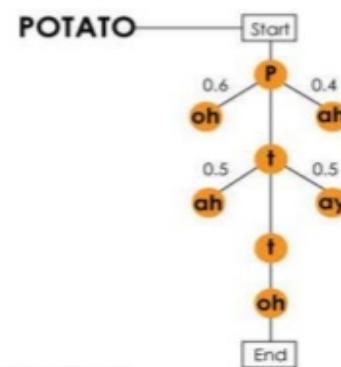
This method of authentication is particularly helpful for banks, call-centers, or services that require the user to interact over a call. Continuous Authentication considers voice pitch, tone, and frequency.

SPEECH RECOGNITION

- Letter Recognition
- Word and Phrase Recognition
- Meaning Interpretation



How Speech Recognition Works



Voice recognition is very common method. This technique is used by most number of online platforms.

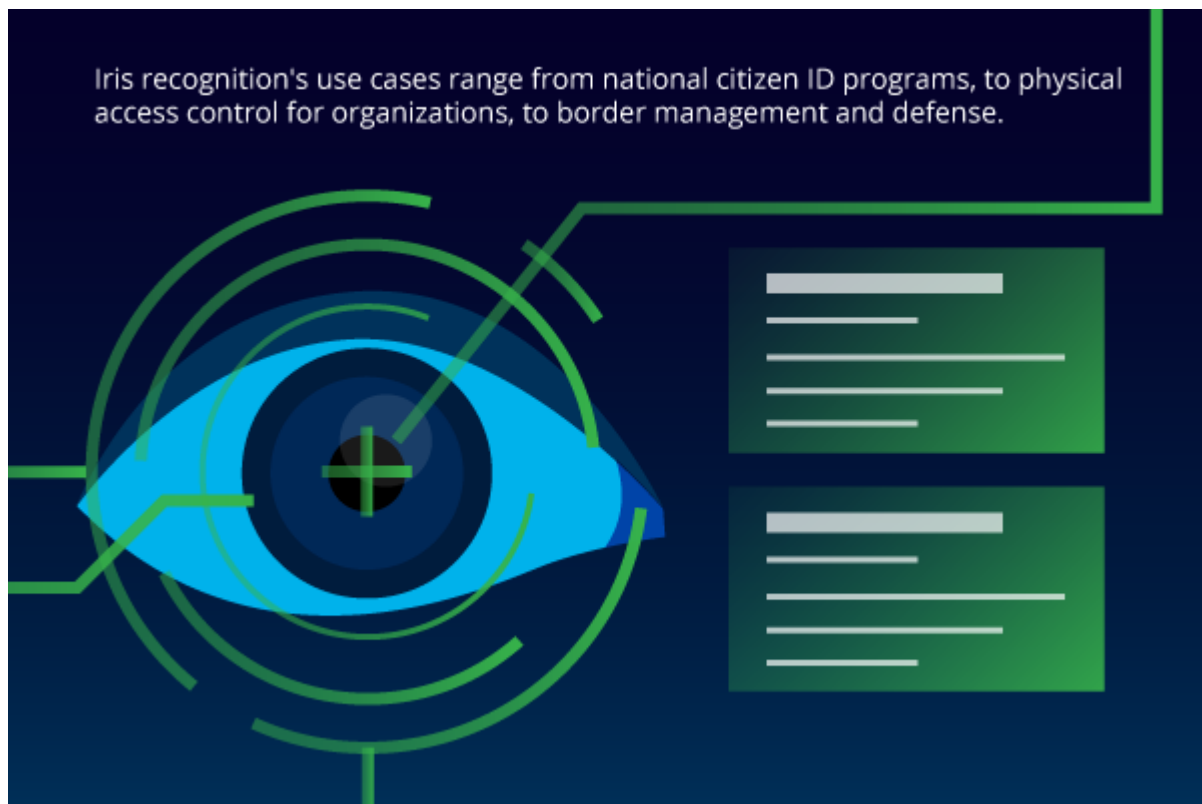
It has some advantages, such as:

- User's Voice is easily accessible.
- Most devices have microphones.
- It is unique for every user except for voice scramblers.
- Pattern matching can be easily done on voice rather than images.
- Changes to user's voice are least expected.
- Most users grant access to web servers to use user's voice.
- It is easier than other methods.
- Any abnormality in voice can alert service provider of illegal access.

Iris Recognition

Iris recognition is a technique which is used to identify a person on the basis of the iris of a person. This is based on a pattern-matching mechanism.

This technique can be used for various purposes such as Social Security Number, to gain access to something, for identification at airports, etc. The hardware and software used for Iris scan is expensive due to which it is not very commonly used.

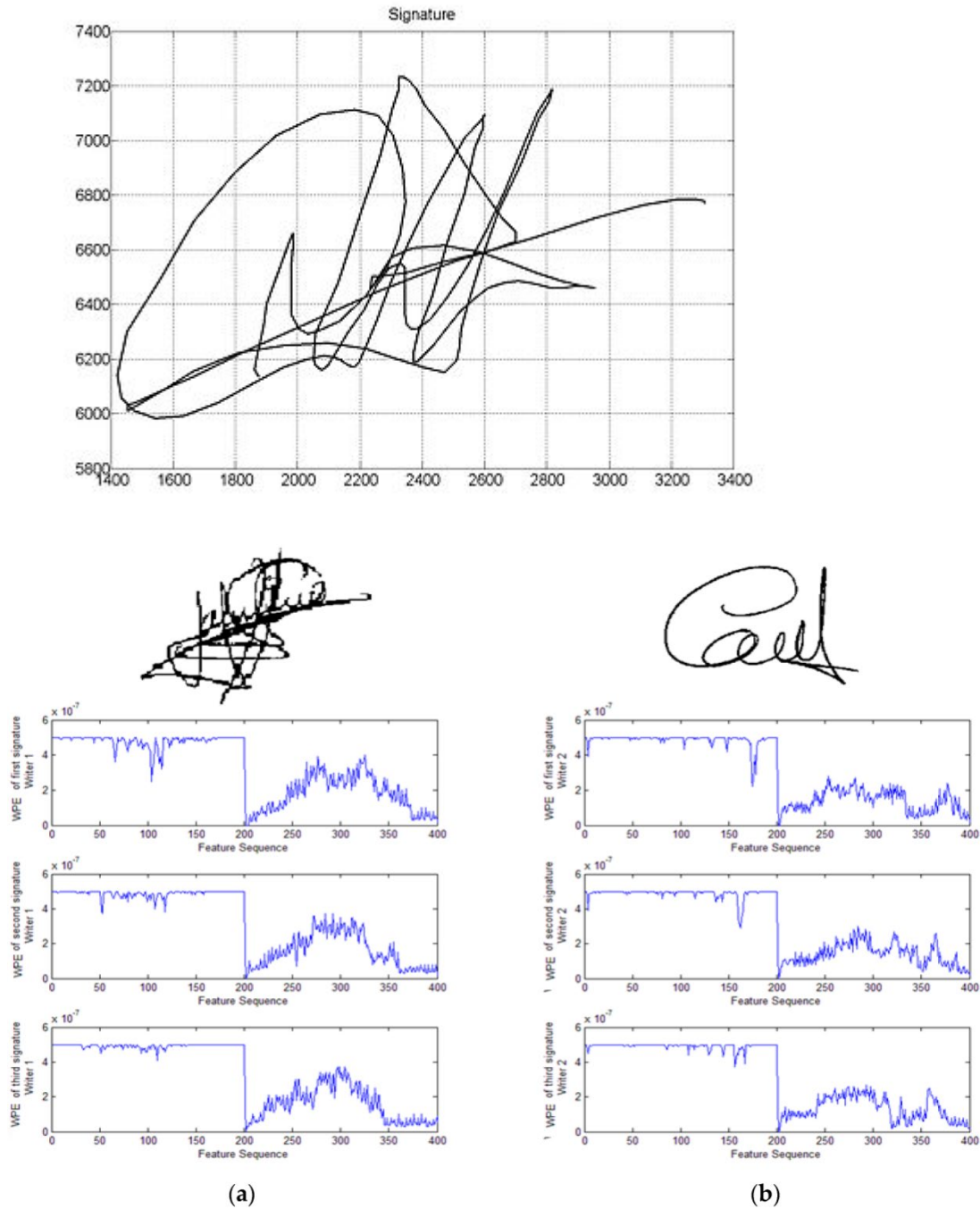


Key features of Iris Scan are:

- It can read through new and clear lenses.
- It needs good lighting in the background of a user for verification.
- Its hardware and software is relatively expensive.
- Image similar to iris of an individual can be used to get illegal access.

Signature Recognition

Another common technique used for biometric identification is signature recognition. In this method digital signatures of a user are used for authentication.



There are two ways to take digital signatures. It can be taken either through uploaded image of signature or by drawing signatures directly on screen using a special pen called stylus.

Fingerprint Recognition

Fingerprint Recognition is a very common technique used for biometric verifications. It is used everywhere as it is a cheaper and secure way of verifying authenticity of an individual user over a digital platform.

Humans have unique patterns on their fingers which is the reason why finger recognition considered very efficient method. It is available mostly on all the digital devices available in the market.

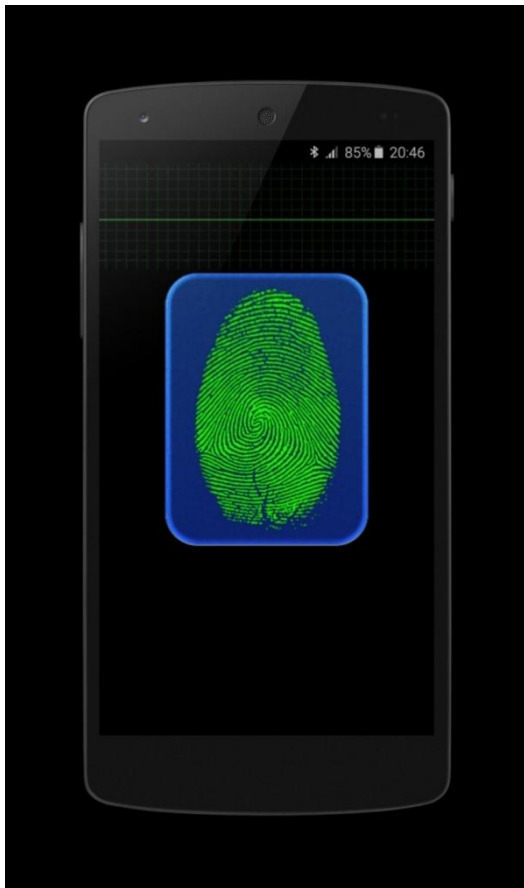


Key features of fingerprint recognition:

- It is most efficient and cheap technique for biometric identification.
- Fingerprints are unique for every individual.
- Digital devices can easily install hardware setup for it.
- It depends on pattern-recognition.
- It is not easy to trick a fingerprint sensor.
- Images cannot be used to get access through this mechanism.

Finger Pressure

Finger pressing can also be used for authenticating a user. It is determined regularly for a particular user. The pressure with which a user presses a sensor is used to determine the validity of an access. It is not very common techniques and it also requires a well designed and programmed model for making predictions which can involve large number of computational and mathematical calculations making it an ambiguous process. This is the reason why it is not very common in use.



Features of Finger Pressure

- It requires large number of computations to make accurate prediction.
- It is uncommon technique.
- It depends mainly on behavior of a user.
- Hardware requires less space to adjust in digital devices.

Geo-Location

Geo-Location is a technique of tracking a user's geographic location. It uses Global Positioning System (GPS) to track Geo-Location of a user. User's location can be used for continuous authentication. Any unexpected changes in location of user can be recorded for keeping a record. Spike in change of location over a small period of time can be considered as suspicious activity and then actions could be taken such as informing user about activity and freezing the account for a certain period.



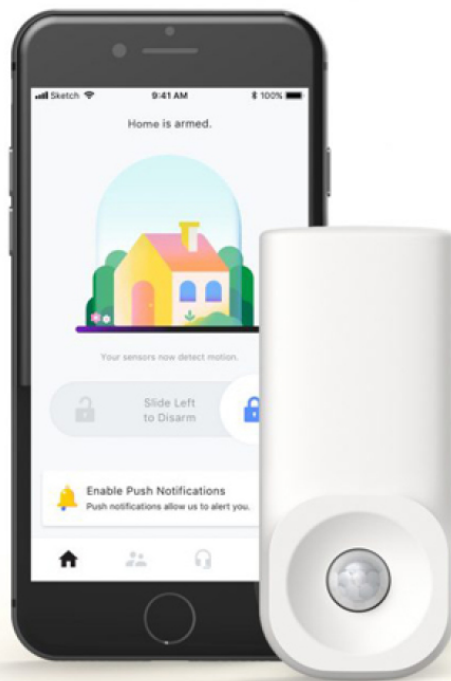
Key features of Geo-Location

- It uses satellite based technology to detect user's location.
- Accuracy of location depends on user's device.
- GPS technology can be easily installed in devices.
- GPS is cheaper than other technologies.
- It is more efficient and common.
- It cannot determine exact identity of a user.

- Geo-Location can only track any suspicious activity.

Motion

Motion can also be used for continuous authentication. But for that the device of a user must have certain sensors that respond actively to any of the device. They not only respond to device but also keep a record of movement of device. This movement can be tracked and recorded. Any suspicious movement when unexpected at a particular place and time can be considered as an attempt to gain illegal access to the user's account. It is the duty of server to provide a particular record of these movements to users.



Key Features of Motion Sensor:

- It depends on the sensitivity of device.
- Atmosphere around the device can also change motion
- It is very unpredictable as it depends on user.
- User can use it for other benefits such as counting steps, walking distance, etc.
- It can be easily installed in a device.
- Special software are needed that can use these sensors.
- Multiple motions sensors are required for more precise and efficient outcomes.

- Each sensor is connected to specific task unit.

CHAPTER 3: DATA COLLECTION

In this chapter we will study certain keystroke dynamic using which we will generate dataset.

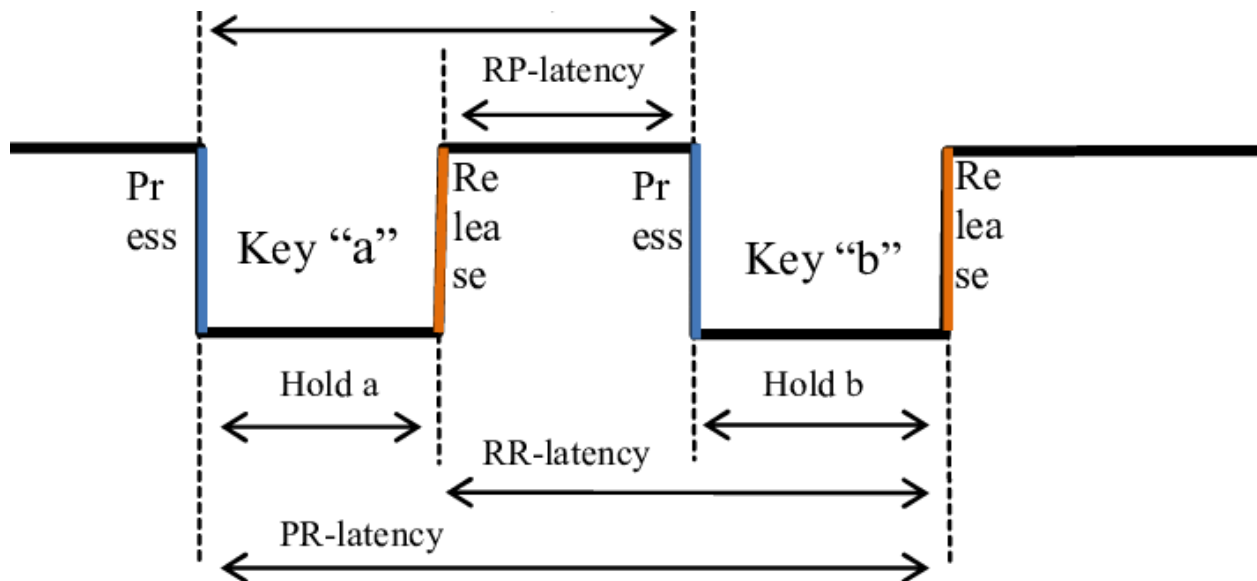
Vibration information may be used to create a pattern for future use in both identification and authentication tasks.

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 keys were pressed and timing information is discarded.

Keystroke dynamics is used for recognition of a person by his or her typing rhythm. Mainly this has been applied to strengthen the username/password log on mechanism. In such a case, not only the value of the password needs to be correct, but also the typing rhythm when entering the password (and maybe also username) needs to be correct.

Key Stroke Dynamics

Key Stroke Dynamics involve the study of certain features that are related to typing of person. These are very efficient and commonly used for continuous authentication. There are various dynamics that we need to study in this section.



Some of them are:

Mean Typing Rate

An individual can type from 35 to 50 words within a minute. However, it is totally dependent on user. Any sudden spike in typing rate for a particular interval of time can be considered as suspicious. Typing rate is also used for other studies and surveys conducted by organizations and companies over end users. It is based on a logical concept that a person with slow typing rate suddenly cannot type fast. In such cases it is assumed that account is accessed by the third party user.

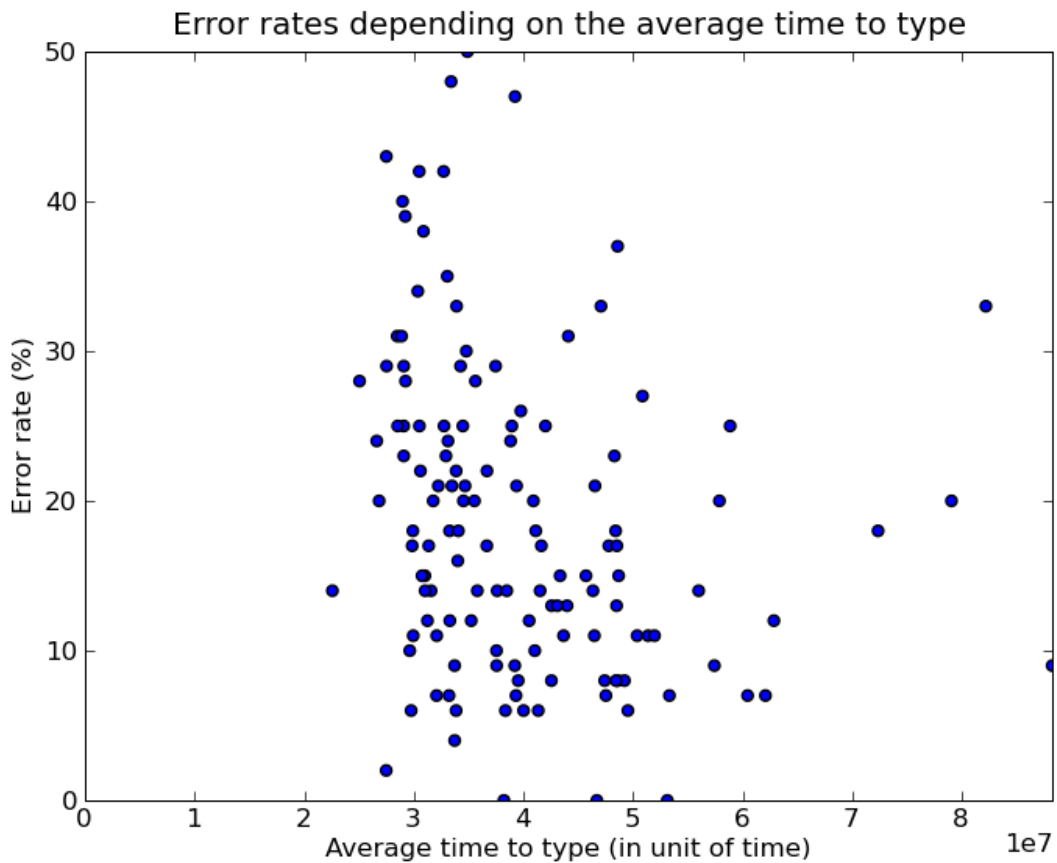


Applications of typing rate:

- Data can be easily collected and stored for analysis purpose.
- Variation in typing rate for a particular user is less.
- Suspicious activity can be traced in accordance with time period.
- Less efficient than some other methods

Mean Error Rate

Mean Error Rate tells about the average number of mistakes in a particular typing session carried out by the user.



This graph show average number of errors that occurred during typing along with interval of time. Error Rate is higher when average typing speed is average

$$\begin{aligned} \text{Net WPM} &= \text{Gross WPM} - \left(\frac{\text{Uncorrected Errors}}{\text{Time (min)}} \right) \\ &= \frac{\left[\left(\frac{\text{All Typed Entries}}{5} \right) - \text{Uncorrected Errors} \right]}{\text{Time (min)}} \end{aligned}$$

Errors are also important as they are used to calculate Net Words per Minute for a particular duration of time. Error rate initially increases when typing rate increases but later decreases with increase in typing rate. It is expected to lie in between a range of 0 to 20 %. Higher error rate is not good for continuous authentication.

Applications of Error Rate

- It is an efficient feature for continuous authentication.
- A user with less error rate is not expected to make typing mistakes.
- A user with high error rate is expected to make typing mistakes.
- Errors are also used for calculating Net WPM.
- It can also be used for various surveys and studies.

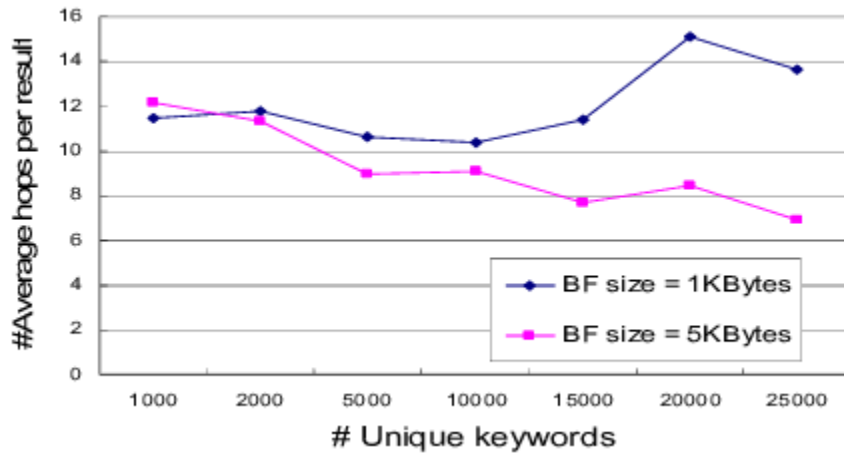
Latency

Latency refers to delay or gap in time between two or more successive events. There are different types of latencies:

- **Digraph Latency**
The digraph latency is latency between two successive key pressing events during a typing session of an end user.
- **Tri-graph Latency**
The tri-graph latency is latency between three successive key pressing events during a typing session of an end user.

- Keyword Latency

Keyword Latency is the latency for an entire word entered during a typing session of an end time user. CPU (Central Processing Unit) determines the length of word by a space between consecutive characters.



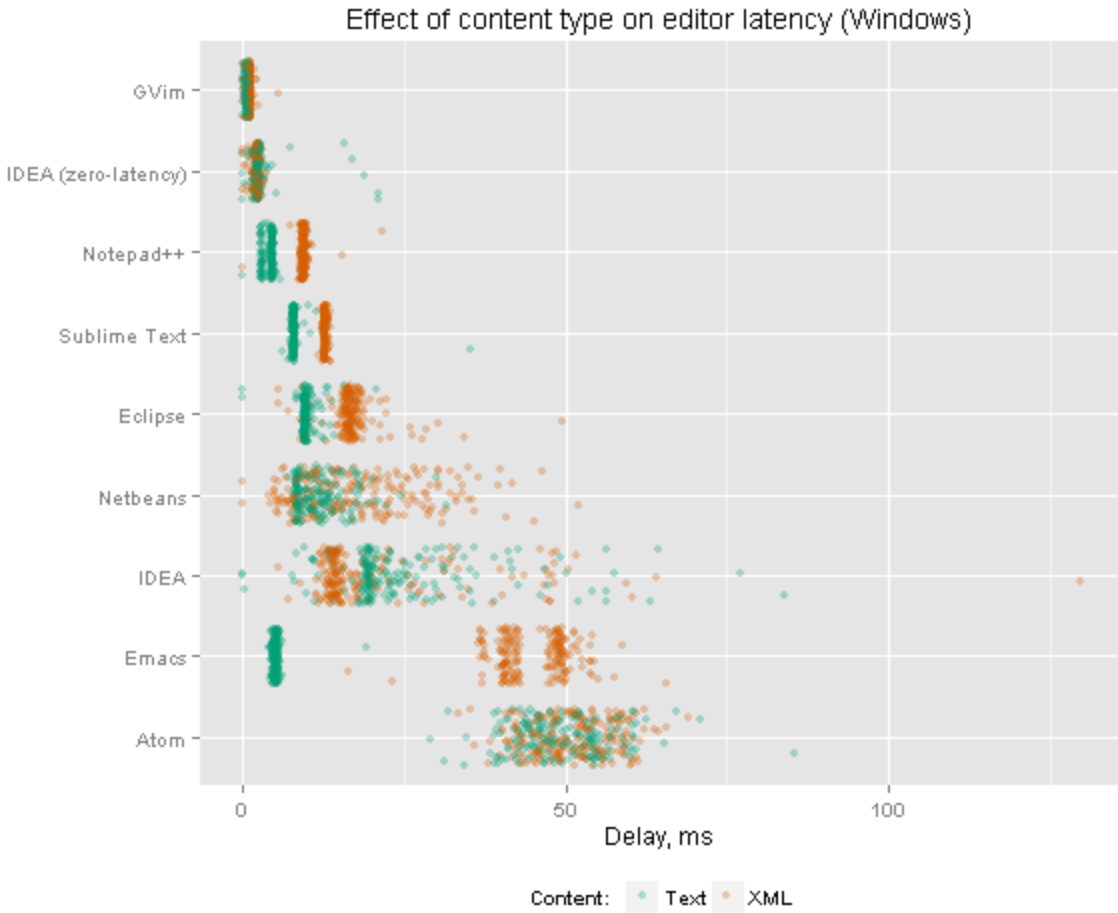
The above graph shows average latency for a given user.

Sizes of BF are:

1K bytes

5Kbytes

Respectively;

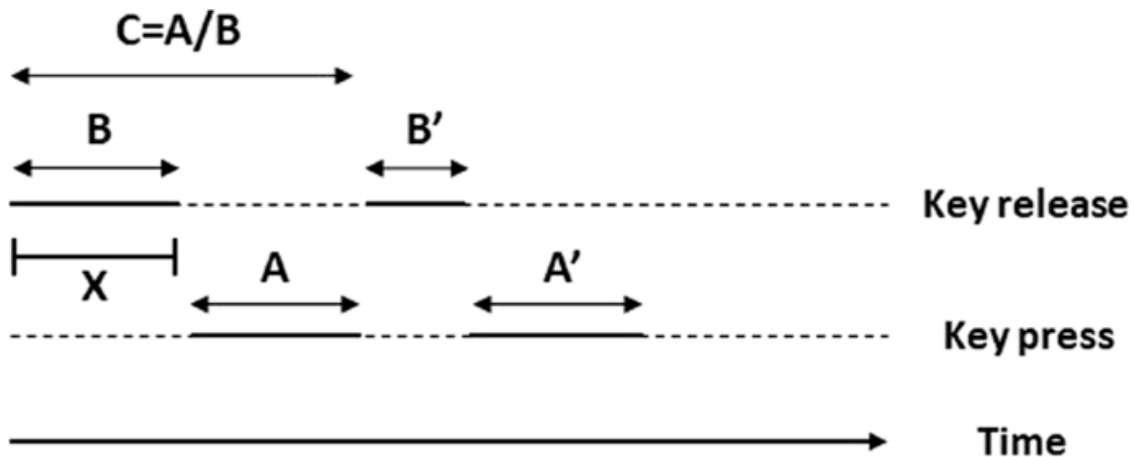
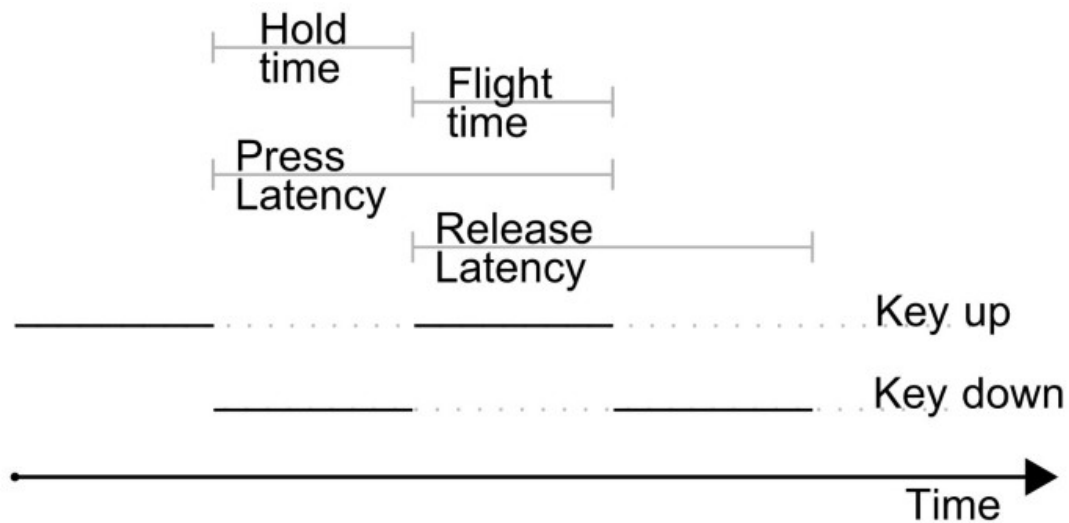


Applications of Latency

- It can be calculated for any number of events.
- Choice of latency depends on the data type.
- Latency is considered as the best feature for data set.
- It can also be used for other survey related purposes.

Flight Time

Flight time is time between two successive keys and is calculated by CPU. It is one of the raw measurements used for key stroke dynamics. It is a very useful keystroke metric that is used for analysis purpose. Flight time is also dependent on typing rate. However, we don't draw any relationship between flight time and typing rate. Some random behavior in typing on user's end can lead to undesired results.

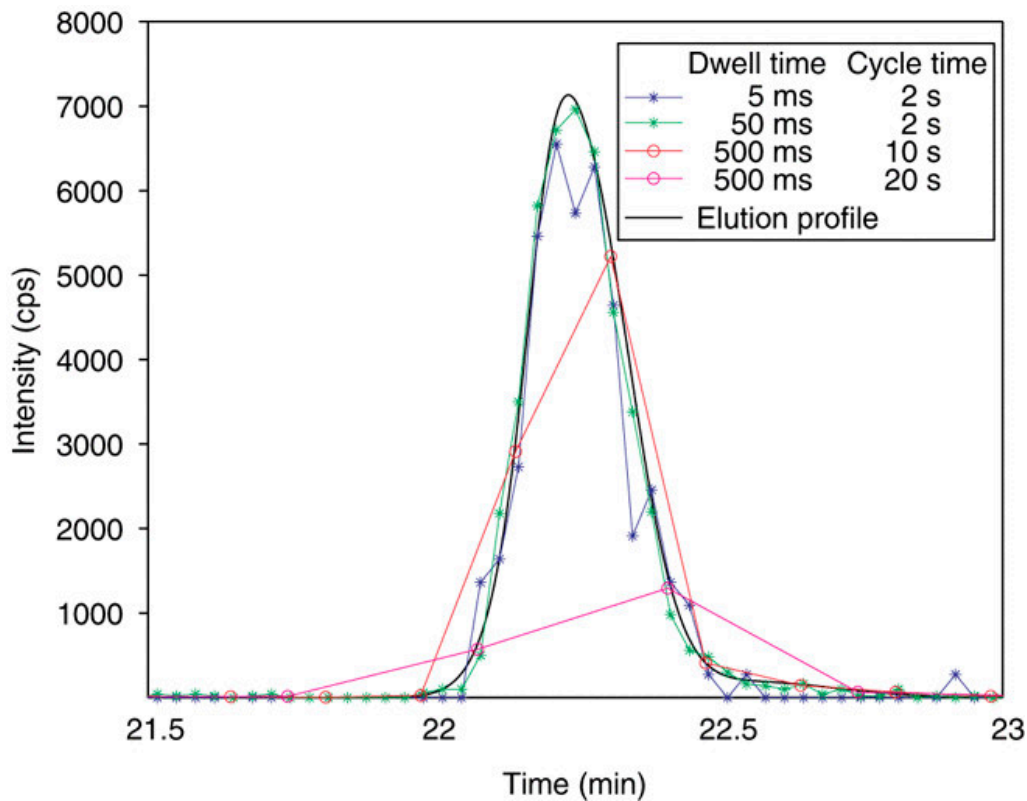


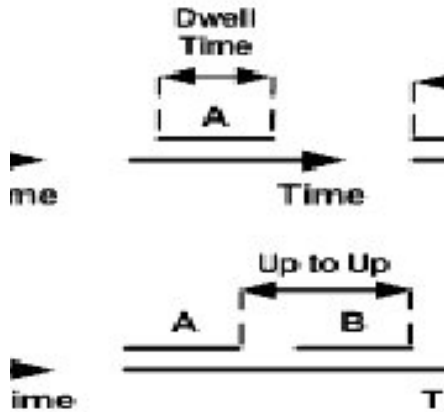
Application of flight time:

- It is the most important feature of keyword metrics.
- It has a deep relation with typing rate.
- There is no existing relationship between error rate and flight time
- Sudden increase in flight time can indicate illegal activity.

Dwell Time

Dwell time is another important feature in keyword dynamics used for the process of continuous authentication. Dwell time is the time for which a particular key is pressed either on phone, keyboard or any digital device. Keys are situated in a specific layout. Due to which dwell time can be different for particular keys. It depends on the reach and effort required for particular press and release of the specified key. Dwell time is the second most important feature for in keyword dynamics.





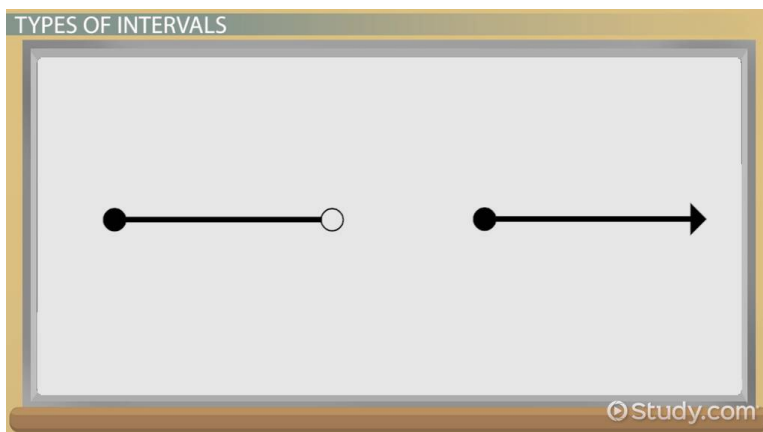
Latency = Dwell time + Flight time;

Applications of Dwell time:

- It is one of the most important features of dynamics.
- Dwell time is highly dependent on typing rate of a person.
- It follows and followed by flight time.
- It is also used by keyboard brands to analyze the quality of their key buttons.

Up to Up/ Interval

This Interval is the difference between flight time of previous key and dwell time of the upcoming key. It is calculated by processors. It requires both dwell and flight time of the keys. The microprocessors installed in the keyboard should be able to respond quickly to the keys pressed.



Application of Up to Up Intervals

- It tells user about the difference in consecutive keystroke action events.
- Calculation of Up to Up interval is little complex.
- It has direct relation to flight time and dwell time.
- Up to Up Interval is not an important feature for continuous authentication.

CHAPTER 4: API AND DATASET

The main motive of our project is to develop a model that would be helpful in preventing brute force attack. There are series of steps which we followed in model development. We will here consider some of the steps used for model development.

STEPS OF MODEL DEVELOPMENT

1. Knowledge related to brute force attack and continuous authentication
2. Keyword Metrics
3. API [Application Programming Interface]
4. Implementation of metrics
5. Data Set Generation
6. Implementation of Supervised Machine Learning
7. Predictions using developed model

In this section of project we have covered segment 1 to 5. Remaining Machine Learning application segment will be covered in further half of our project.

API

We have developed an Application Programming Interface in python programming language.

Reason for choice of Python-

- Syntax of Python is easier than other languages.
- It allows multi-programming paradigm.
- It also facilitates connectivity to database.
- It allows us to develop both API and GUI.
- Has large number of modules.
- Python allows better support for Machine Learning.
- Python is dynamically typed language.
- It is platform independent.

An Application Programming Interface is a link between two applications that supports easier and faster communications between the two connected applications. There are some salient features of every API. The choice of an API depends on the type of task assigned to a particular programmer.

Features of our API:

- It is an interface that links two applications.
- There is form at the front end.
- There is a programming logic based on data retrieval.
- We are supposed to try different number of login credentials over a database.
- The data entered will be retrieved at the back end of our interface.
- Data collected is then collected in to single dataset.
- API also displays the value of features for a particular runtime.

Libraries Used

numpy

Is a library in python that allows us to perform mathematical computation on matrices and arrays.

tkinter

Is a python library that allows us to build a GUI [Graphical User Interface] and link some events and tasks to the GUI.

Important Functions Used:

numpy:

- `reshape()`
- `transpose()`
- `broadcast()`
- `concatenate()`
- `split()`
- `resize()`
- `insert()`
- `delete()`
- `update()`
- `append()`
- `hsplit()`
- `unique()`
- `stack()`
- `vstack()`

tkinter

- pack()
- grid()
- place()
- title()
- geometry()

Miscellaneous functions

- eval()
- int()
- chr()
- input()
- list()
- locals()
- map()
- object()

Tkinter

It is a module that allows programmers to build a Graphical User Interface using python programming language.

Steps to build a GUI:

1. from tkinter import *
2. Create Main window
3. Choose and drop widgets in the window
4. Add Action listener classes.

Important Widgets

- Button
- ListBox
- Label
- Menu
- CheckBox
- RadioButton
- Scale
- Text

Dataset

The dataset is generated (“Dataset (1) .csv”) which contains over 5,000 data entries of login credentials.

This data has values entered by multiple attempts over API [Application Programming Interface]. There are various features that could be used for our project but we have used three most important features.

Dataset has 6 features.

[Each of three type 2 times]

The features are:

- DWELL_TIME (PW1)
- DWELL_TIME (PW2)
- FLIGHT_TIME (PW1)
- FLIGHT_TIME (PW2)
- AVG_KEYSTROKE_TIME (F)
- AVG_KEYSTROKE_TIME (D)

DWELLTIME(PW1)	DWELLTIME(PW2)	FLIGHTTIME(PW1)	FLIGHTTIME(PW2)	AVG.KEYSTROKE_TIME(D)	AVGKEYSTROKE_TIME(F)
75	66	1206	4137	70	2671
90	74	823	2736	82	1779
53	84	1166	3534	95	2933
99	73	802	2642	86	1722
67	61	898	3036	64	1967
64	59	1266	2852	61	2059
86	73	1206	1405	79	1305
71	79	1181	3954	75	2567
59	81	961	2869	70	1915
59	99	1205	1597	79	1401
51	70	861	1431	60	1146

Hence the generated dataset will be used for further analysis.

CHAPTER 5 : DATA PREPROCESSING AND RESULT

1. Data Preprocessing

- 1.1 Handling Missing Data
- 1.2 Feature Scaling
- 1.3 Encoding Categorical Data
- 1.4 Visualize the Correlation Matrix

2. Applying Machine Learning Algorithm

- 2.1 Split the data into training and testing data set
- 2.2 Import Sklearn and Machine Learning Inbuilt Algorithm
- 2.3 Apply KNN Algorithm :
 - 2.3.1 We Apply Knn Algorithm on our dataset . Knn Algorithm uses Equilidean Distance formula to provide a particular class to a test label.
 - 2.3.2 We assume and take a value as k .
 - 2.3.3 Whenever a test set arrives we calculate their k nearest neighbours values and if maximum nearest neighbours are from class a then we provide test set to class a otherwise we provide test set to class b.
- 2.4 Apply Naive Bayes Algorithm :
 - 2.4.1 We Apply Naive Bayes Algorithm on our dataset . Naive Bayes uses bayes theorem probability formula to provide a particular class to a test label.
 - 2.4.2 First we calculate probability of test label that it belongs to class a then we calculate probability of test label that it belongs to class b.
 - 2.4.3 If probability of class a is higher then test set belongs to class a otherwise test set belongs to class b.

3. Result

- 3.1 We evaluate confusion matrix by both algorithms.
- 3.2 We evaluate accuracy, precision, recall and f1 score by both algorithms.
- 3.3 We compare both algorithms based on these features
- 3.4 We visualize the result

4. Outcome :

4.1 After Getting Results of Both Algorithms We Get The Outcome That Which Particular WebPage Is Facing Brute Force Attack And Which WebPage Are Secured Based On Their Labels.

4.2 By Getting Their Performance Matrix We Analyze Which Algorithms Gives Us A Better Outcome About The Brute Force Attack.

4.3 Using Knn Algorithm We Get :

Accuracy : 0.88

Precision : 0.91

Recall : 0.88

F1 Score : 0.89

4.4 Using Naive Bayes Algorithm We Get :

Accuracy : 0.97

Precision : 0.82

Recall : 0.85

F1 Score : 0.83

CHAPTER 6: CONCLUSION

CONCLUSION AND SUMMARY

- Through this paper we have studied various research related to cyber-security.
- We studied the nature of brute force attack.
- Then we discussed possible goals of a brute force attack.
- Then we further discussed the use of continuous authentication in preventing brute force attack.
- It was followed by some preventive measures to prevent this attack.
- We also considered keystroke dynamics.
- Then we build an API.
- We entered multiple login credentials over various attempts.
- At last we generated dataset.
- we will use this dataset to predict a brute force attack

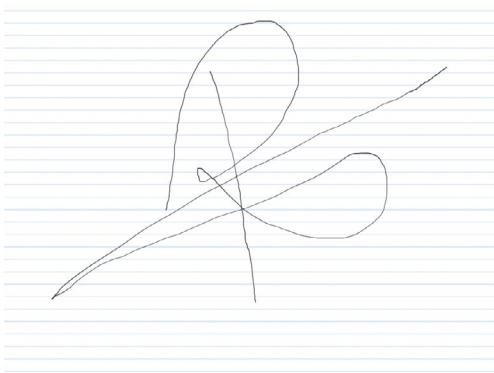
Application

- Awareness

- Cyber-security

REFERNCES

1. <https://www.kaspersky.com/resource-center/definitions/brute-force-attack>
2. <https://www.varonis.com/blog/brute-force-attack/>
3. <https://www.forcepoint.com/cyber-edu/brute-force-attack>
4. <https://www.imperva.com/learn/application-security/brute-force-attack/>
5. <https://www.cloudways.com/blog/what-is-brute-force-attack/>

A handwritten signature in black ink on a background of horizontal blue lines. The signature is stylized and appears to be a cursive or semi-cursive script.

(Supervisor Signature)

Rohan Tyagi

(Student Signature)

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: ...18-06-2021.....

✓

Type of Document (Tick): PhD Thesis M.Tech Dissertation/ Report B.Tech Project Report Paper

Name: Rohan Tyagi Department: IT Enrolment No 171463

Contact No. 7807051838 E-mail. 171463@juitsolan.in

Name of the Supervisor: Mr. Rizwan Ur Rehman Title of the

Thesis/Dissertation/Project Report/Paper (In Capital letters): Securing Web Application Against

Brute Force Attack Using Continuos Application

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/ revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages = 46
- Total No. of Preliminary pages = 44
- Total No. of pages accommodate bibliography/references = 2



Rohan Tyagi
(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at 11(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor)

Signature of HOD

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Generated Plagiarism Report Details (Title, Abstract & Chapters)	
Report Generated on	<ul style="list-style-type: none"> • All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String 		Word Counts	
			Character Counts	
		Submission ID	Total Pages Scanned	
			File Size	

Checked by
Name & Signature

Librarian

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File)

through the supervisor at plagcheck.juit@gmail.com

ORIGINALITY REPORT

11 %
SIMILARITY INDEX

10 %
INTERNET SOURCES

1 %
PUBLICATIONS

5 %
STUDENT PAPERS

PRIMARY SOURCES

1 searchsecurity.techtarget.com 3 %
Internet Source

2 Submitted to Jaypee University of Information Technology 2 %
Student Paper

3 pdfs.semanticscholar.org 2 %
Internet Source

4 Submitted to Liberty University 1 %
Student Paper

5 www.forcepoint.com <1 %
Internet Source

6 whatis.techtarget.com <1 %
Internet Source

7 research.ijcaonline.org <1 %
Internet Source

8 brage.bibsys.no <1 %
Internet Source

9 repository.kisti.re.kr <1 %
Internet Source

10 Umid Kumar Dey, Alaa Noor. "Comparative Exploration Of Prediction Algorithms For Sentiment Analysis Using NLP", 2019 International Conference on Computer Communication and Informatics (ICCCI), 2019 Publication <1 %

11 www.coursehero.com Internet Source <1 %

12 phoenixnap.com Internet Source <1 %

13 www.filemakr.com Internet Source <1 %

14 www.researchsquare.com Internet Source <1 %

15 epdf.tips Internet Source <1 %

16 www.ncfs.ucf.edu Internet Source <1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On