FILE ACCESS AUTHENTICATION USING INTRUSION DETECTION

Enrolment No:	14103178, 14103185
Name of Student:	Saumitra Vikram Singh, Gaurav Sood
Name of Supervisor:	Ms. Amanpreet Kaur

May – 2018

Submitted in partial fulfillment of the Degree of Bachelor of Technology
In
Computer Science Engineering

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING & INFORMATION TECHNOLOGY JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

(I)

TABLE OF CONTENTS

(II)

DECLARATION

We hereby declare that this submission is our own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: Noida, U.P

Signature:

Date: May 2018

Name: Saumitra Vikram Singh , Gaurav Sood

Enrolment No: 14103178, 14103185

(III)

CERTIFICATE

This is to certify that the work titled "**File access authentication using intrusion detection**" submitted by "**Saumitra Vikram Singh & Gaurav Sood**" of B.Tech of Jaypee Institute of Information Technology University, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

Signature of Supervisor

Name of Supervisor; Ms. Amanpreet Kaur

Designation

Date May 2018

(IV)

ACKNOWLEDGEMENT

We take this opportunity to express our profound sense of gratitude and appreciation to all those who helped us throughout the duration of this project. We would like to express our special thanks of gratitude to our mentor **Ms. Amanpreet Kaur,** who gave us the golden opportunity to be a part of this wonderful project in the domain of Cloud Security and for providing guidance and expert supervision for this project. She has set a spectacular example on us young impressionable minds in the duration of creating this project. We are really thankful to her. We are making this project not only for fulfilling a college requirement but to also increase our knowledge. Thanks again to all who helped us in making of this project.

Signature of the Student:

Name of Student: Saumitra Vikram Singh, Gaurav Sood

Enrolment Number: 14103178, 14103185

Date: May 2018

(V)
SUMMARY

In the present time, where there is such a boon of the usage of information technology across the globe, majority of the companies see cloud as an adequate platform on which the can execute their in-house as well as overall operations. This raises plenty security issues as most of the companies have some data which must be kept confidential and private.

This project focuses on providing a complete security solution for an entity for their cloud platforms. Our solution utilizes Database as a service (DAAS) sourced from godaddy.com, for storing files as well as user databases on the cloud. The software has been designed using C#, and visual studio 2010 as the IDE. The various features provided include user and admin login, Encryption, Compression, Folder locks, file sharing and downloading by users. Even on registration of a new user, the user won't be able to access any file until and unless his id is approved by the admin. File sharing by a user can only be to another user registered in the same system and not other.

Login authorization by OTP ensures security on database and user level as one can't access data without proper credentials but also for some reason if they do the encryption ensures that the person can't make any meaningful information of the data.

Machine learning algorithms (Decision Tree, Naïve Bayes and a hybrid of the two) have been used for designing an intrusion detection system. For training of said system dataset from UNSW has been taken which contain values such as IP Addresses, Source Ports, Protocol used etc. For preventing and detecting intrusions the MAC and IP addresses of every user registered is stored and whenever an unauthorized access is attempted the file is not downloaded from the server and also an alert if sent to the admin account.

(VI)

LIST OF FIGURES

(VII)

LIST OF TABLES

(VIII)

LIST OF SYMBOLS & ACRONYMS

DaaS: Desktop-as-a-service

SaaS: Software as a service

Paas: Platform as a Service

Iaas: Infrastructure as a Service

Sql: Structured Query language

IDS: Intrusion detection system

HIDS: host-based interruption framework

NIDS: network intrusion detection systems

DDoS: Distributed Denial of Service

Dos: Denial of Service

AES: Advanced Encryption Standard

DES: Data Encryption Standard

API: Application programming interface

GUI: Graphical user interface

IDE: Integrated development environment

ML: Machine Learning

Chapter-1 Introduction

1.1 General Introduction

The motivation behind this venture is to beat the security issues display in a system. The venture plans to prevent gatecrasher or unapproved individuals from getting to a private cloud. Look into venture address this part as Intrusion Detection. Mounting world can't envision notwithstanding for a solitary day without PC and PC is premise on web. These days secure data of web is ending up noticeably high need. Present day world accentuations in a route by which it can be shield the information and data from any illegal and unapproved get to.

Interruption Detection Systems (IDS) can be varies in different methods and progress with the target to identify suspicious activity in divergent ways. There are two huge classifications of interruption discovery frameworks. One is called arrange based interruption location framework (NIDS) and the other one is host-based interruption framework (HIDS).

As of now, if Internet framework strike, for example, man in the center assault, foreswearing of administration assaults and worms disease, have turned out to be a standout amongst the most genuine dangers to the system security . It is likely possible to distinguish the assaults and unusual practices if there is adequate and proficient strategy and procedure exists for screen and analyze, and it can not just ensure continue cautioning of potential assaults, additionally assist to perceive the reasons, source and areas of the peculiarities. By along these lines, it might help to control the assaults, sooner than they have enough time to communicate over the system.

This report speaks to the technique, in support of recognising system inconsistencies by examining the sudden change of getting to of information. With the correlation of other oddity recognition strategies. We have point of convergence on the dynamic conduct of the system as opposed to utilising the static models. Our procedure and strategy concerns the Auto-Regressive (AR) procedure to demonstrate the fast and unforeseen change of time arrangement information. Our fundamental object is to piece IP address of a man who is not approved to get to a document is as yet attempting to do it. It is valuable in the event of enormous or little associations, schools, universities, business firms, IT firms.

1.2 Relevant current/open problems

- DDoS attacks in cloud computing: Collateral damage to non-targets:

The effects of distributed denial-of-service attacks on cloud computing are not very similar to those in on-premise infrastructure. In the context of DDoS attacks in multi-tenant clouds, instead of just the victim server, multiple other stakeholders are also involved. Some of these important stakeholders are co-hosted virtual servers, physical servers, network resources, and cloud service providers. Damages/effects to these stakeholders include performance interference, web service performance, resource race, service downtime, and business losses.

- Cloud service abuses:

Cloud services can be commandeered to support nefarious activities, such as using cloud-computing resources to break an encryption key in order to launch an attack. Other examples including launching DDoS attacks, sending spam and phishing emails, and hosting malicious content.

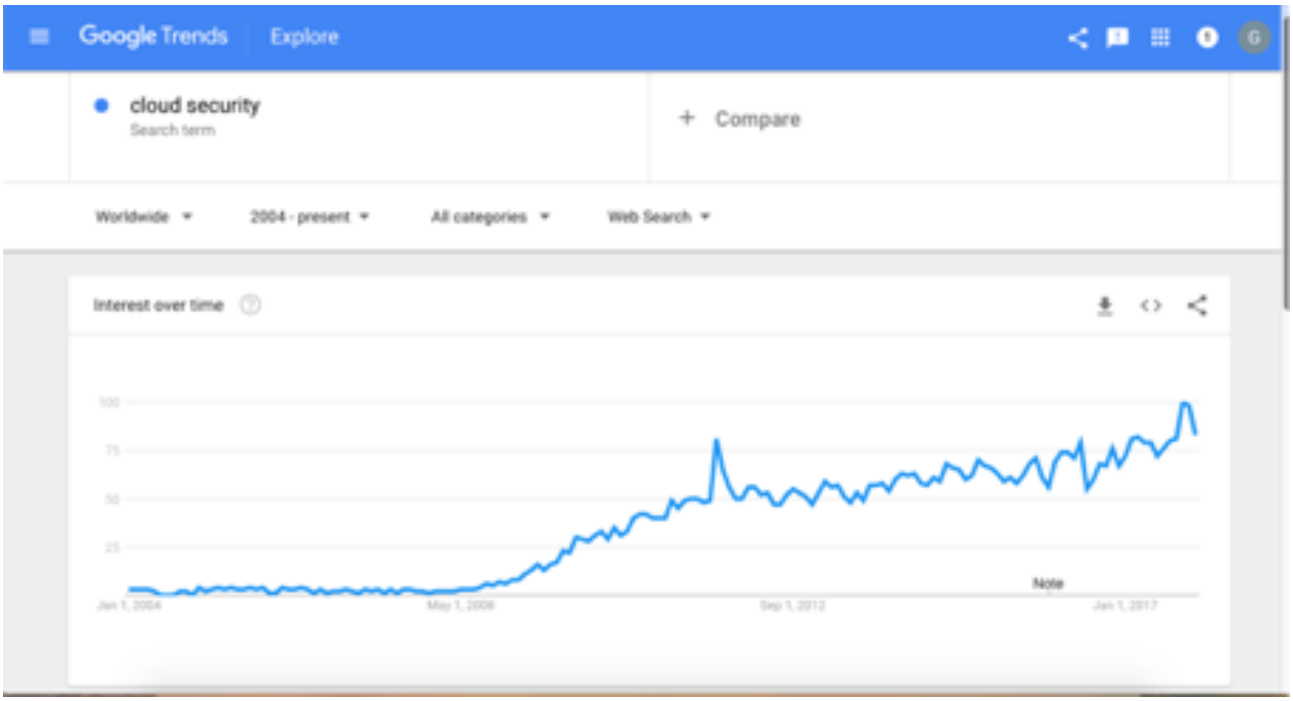

Fig 1.1 Cloud security trend5

1.3 Problem Statement

In ideal IDS should have important characteristics, such as high accuracy detection rate, low false positive and negative rate, and low computational cost. Present IDS should be able to comply with new and upcoming challenges such as.

• Fast expansion in the network systems and in digital devices, like smart phones, tablets, laptops etc.

• Should be able to fight against the increasing number of complex network attacks that are taken place with a lot of planning.

- Occurrence of another interruptions and programming vulnerabilities.
- Low-level identification with free interruption detailing. Regardless of what identification guideline has been connected and what review information has been analyzed in interruption location, the sent interruption recognition instrument or IDS ordinarily distinguishes the interruption altogether from low-level crude data (for instance, crude system movement, have log document record). Despite the fact that low-level data contains the potential ramifications for interruption conceivable outcomes, it is troublesome for the interruption identifier to straightforwardly report the abnormal state caution reflection (that is, the framework has been traded off). An issue in autonomous interruption revealing is when programmer actualizes a similar assault more than once, the IDS will be occupied with producing dreary cautions in a single particular time, in the mean time, the framework manager is overpowered by the surge of repetitive alerts, the programmer can accumulate extra data and make different interruptions.
- Self-existed shortcomings in connected IDS can enable the programmer to debilitate it remotely without warning for the framework head.

Cloud computing has become a thing of upmost importance for business, providing cheap, virtual services that once needed high costing local hardware it cuts out the expensive cost of hardware, but there are also some shortcomings. data in the cloud server needs to be properly encoded so that any unauthorized person cannot get his hands on the private and confidential information of any user . we have proposed a system that has a multi level encryption system, the first level is at the local level where the file is encoded using GZIP stream and AES encryption. then the file is uploaded on the cloud by a authorized user, to become a authorized user, first the user has to fill in his details and a verification link is send to his mail through which he has to confirm his details and once this is done the user will only be authorized after the system administrator verifies him. Each time a user logs in his mac and ip addresses are stored along with the file that he has tried to download or upload. This is part of user level security. the third level of security is when a user tries to access the file he will have to go through 3D security verification, only after which he can proceed.

1.4 Overview of proposed solution approach and Novelty/benefits

Cloud computing is turning into one amongst the foremost valuable innovations for business, providing Low-cost, virtual services that once needed high costing, native hardware. Cloud computing cuts out the high price of hardware, however there also are some drawbacks to that, info within the cloud must be properly encrypted in order that malicious users cannot get their hands on confidential information as well as only authorised people are able to access information on the cloud server. we have proposed a system that has a multi level encryption system, the first level is at the local level where the file is encoded using GZIP stream and AES encryption. then the file is uploaded on the cloud by a authorised user, to become a authorised user, first the user has to fill in his details and a verification link is send to his mail through which he has to confirm his details and once this is done the user will only be authorised after the system administrator verifies him. each time a user logs in his mac address, ip address, source port number and destination port number are stored along with the file that he has tried to download or upload. this is part of user level security. the third level of security is when a user tries to access the file he will have to go through 3D security verification, only after which he can proceed.

on any cloud system there are many such frequent attacks, so to further study the types of attacks as well as the files on which such attacks occur frequently we will have used machine learning algorithms including decision tree and naive bayes, we have developed a hybrid algorithm out of the two (Decision tree and naive bayes), by first calculating the probability using naive bayes of an attack on that file and further using decision tree to find the entropy.

One of the most important analysis challenges during this field is that the inconvenience of a comprehensive network based mostly knowledge set which might replicate fashionable network traffic eventualities, immense styles of low footprint intrusions and depth structured info regarding the network traffic.

we have used “UNSW-NB15 dataset ” which is a comprehensive data set for network intrusion detection systems. This data set has a hybrid of the real modern normal and the contemporary synthesized attack activities of the network traffic.

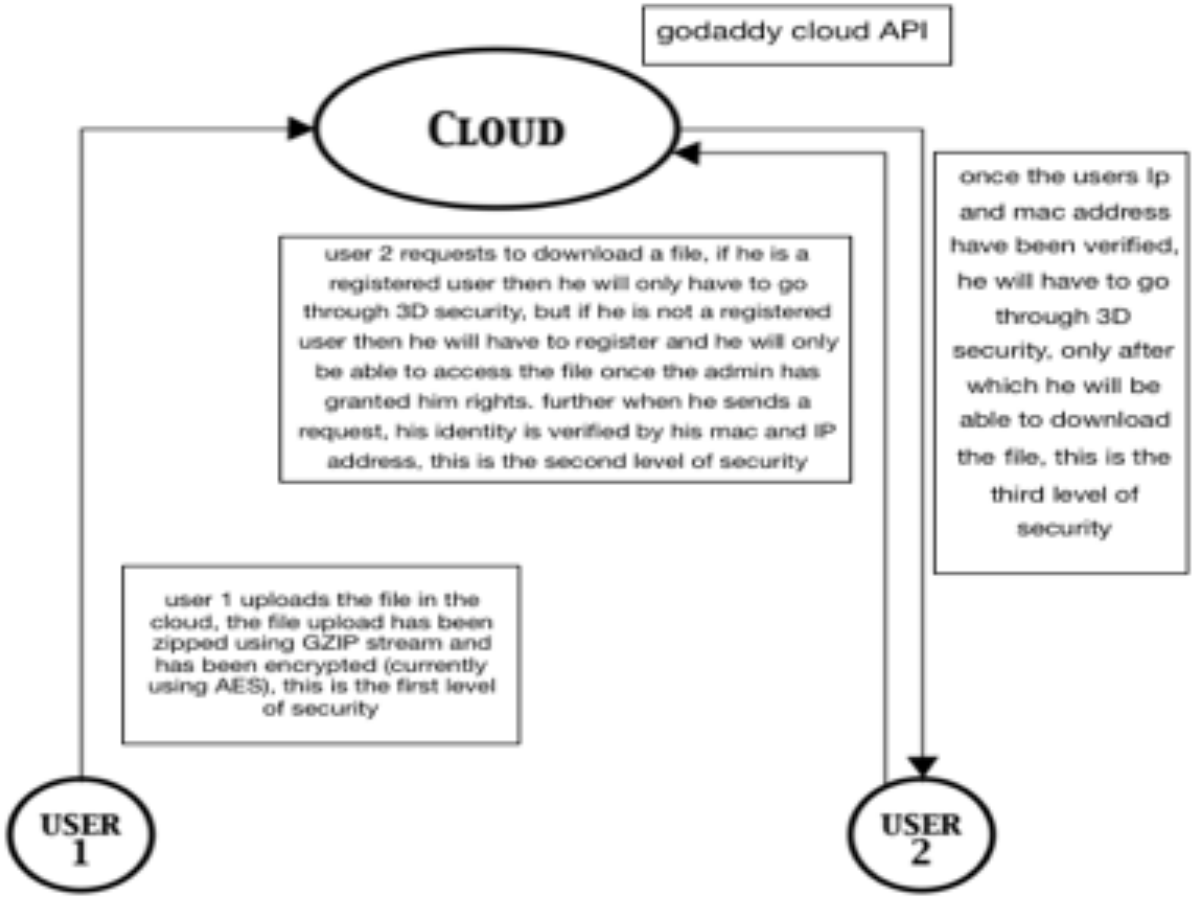

Fig1.2 Design Diagram

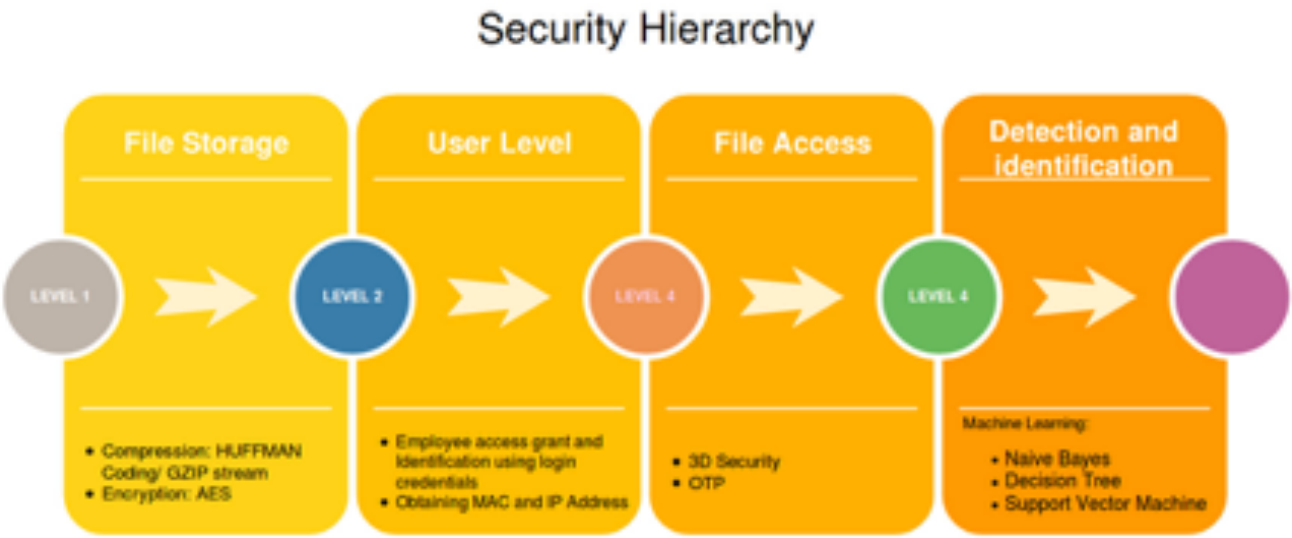

Fig 1.3 Security Hierarchy levels

Chapter-2 Background Study

2.1 Literature Survey

2.1.1 Summary of Papers

Paper 1

TITLE OF PAPER	An Analysis of the Cloud Computing Security Problem
AUTHORS	Mohamed Al Morsy, John Grundy and Ingo Müller
YEAR OF PUBLICATION	2016
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	Proceedings of the APSEC 2016 Cloud Workshop, Sydney, Australia
SUMMARY	While providing a detailed insight in the basic requirements of any cloud Security model and also the different issues faced in each of the delivery model and also elucidating the different the two key characteristics of a good cloud computing model i.e. multi-tenency and elasticity, different probable approaches were given for implementing multi-tenency but a detailed problem statement of elasticity nor a probable approach was highlighted for the characteristic of elasticity. The security vulnerabilities of each model was beautifully presented and elucidated but redundancy and also a very generic undertone of the issues were written. A basic and informative insight in the security implications was shed light on for a clear picture and easy understanding.
Web Link	https://arxiv.org/abs/1609.01107

Table 1: Research Paper 1

Paper 2

TITLE OF PAPER	Data security in decentralized cloud systems – system comparison, requirements analysis and organizational levels
AUTHORS	Mohamed Al Morsy, John Grundy and Ingo Müller
YEAR OF PUBLICATION	2017
PUBLISHING DETAILS	Müller et al. Journal of Cloud Computing: Advances, Systems and

WHERE THIS PAPER WAS PUBLISHED	Applications (2017) 6:15 DOI 10.1186/s13677-017-0082-3
SUMMARY	A very neat comparison of different systems used in different organizations based on very practically relevant metrics which enables to better judge the performance of each system in the user relevant qualities which one may personally require. The system compared are not very common in the Indian market and hence are not familiar so this paper runs the risk os being irrelevant to the Indian consumers and readers. The consideration of decentralization of the organization levels and its impact on the overall trust towards other participants was a new take on a very important factor. Although a concrete winner or best option in neither the systems analysis nor the effect on decentralization was not observed. The cryptic language also posed a problem as it required multiple readings to get an idea of flow of processes. On a whole a very new approach to the Cloud computing problems and an exceedingly interesting article to get research ideologies from
Web Link	https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-017-0082-3

Table 2: Research Paper 2

Paper 3

TITLE OF PAPER	Machine Learning for Anomaly Detection and Categorization in Multi-cloud Environments
AUTHORS	Tara Salman, Deval Bhamare, Aiman Erbad, Raj Jain, Mohammed Samaka
YEAR OF PUBLICATION	2017
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	2017 IEEE 4th International Conference on Cyber Security and Cloud Computing
SUMMARY	The paper was published on a reputed platform i.e. IEEE and garnered a lot of citations and downloads. The accuracy obtained was promising and the categorization accurate. However due to a

	limited dataset 3 of the attacks were not classified. Also a comparison of only two learning models was taken which provide comparison between only them and now their standings on a whole. Due to computational and complexity constraints several of the features in the dataset were discarded and hence not included in the learning models. The models mentioned will require constant updating for new traffic patterns, modifications in the attack types and new evolving attacks. On a whole an excellent paper which provided a keep insight and comparison of different learning models and classification of traffic and the attacks. It also gives an idea on how to implement these models on multi-cloud environments. As students it was a very helpful reading material.
Web Link	http://ieeexplore.ieee.org/document/7987183/?reload=true

Table 3: Research Paper 3

Paper 4

TITLE OF PAPER	Wavelet Transform and Unsupervised Machine Learning to Detect Insider Threat on Cloud File Sharing
AUTHORS	Wangyan Feng, Wenfeng Yan, Shuning Wu, Ningwei Liu
YEAR OF PUBLICATION	2017
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	2017 IEEE International Conference on Intelligence and Security Informatics (ISI)
SUMMARY	The paper while being clear in its language could not elucidate the variables taken for the wavelet transfor. Also no analysis based on any performance or accuracy metric was given for the three unsupervised machine learning algorithms considered.
Web Link	http://ieeexplore.ieee.org/document/8004896/

Table 4: Research Paper 4

Paper 5

TITLE OF PAPER	Improved 3-Dimensional Security in Cloud Computing
AUTHORS	Sagar Tirodkar, Yazad Baldawala, Sagar Ulane, Ashok Jori

YEAR OF PUBLICATION	2014
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	International Journal of Computer Trends and Technology (IJCTT) – volume 9 number 5– Mar 2014
SUMMARY	The paper was clear in its motive and its approach to provide a solution for cloud. The three pronged while very detailed provided no insight on the complexity costs of the solution. Another probable setback was the use of DES instead os AES as AES would have been more optimal and secure as it supports a much larger key lengths and also DES has been proven to be an inadequate encryption technique. The drawback of classification of different data in different rings instead of going in for an onion approach was a archaic decision. The decision of given key management roles to users might be secure but it would be very impractical on Multi- clouds which is generally employed in the real world
Web Link	https://arxiv.org/abs/1404.1836

Table 5: Research Paper 5

Paper 6

TITLE OF PAPER	Intrusion Detection in Computer Networks via Machine Learning Algorithms
AUTHORS	Fatih Ertam, □lhan F□rat K□l□nçer, Orhan Yaman
YEAR OF PUBLICATION	2017
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	2017 International Artificial Intelligence and Data Processing Symposium (IDAP)
SUMMARY	The increase in the number of devices connected to the Internet has allowed the data on the internet to reach very high dimensions. It is especially important for network administrators to determine which data in this high data is normal and which data is an attack. Intrusion detection systems are designed to address these needs of network administrators. In this study, classification studies on NB, bN, RF, SMO and MLP algorithms and KDD Cup

	99 data set which are frequently used in intrusion detection studies were performed in machine learning based approaches. RFR and SMO for probe class, Rf and SMO for U2R class, bN and RF for DDos class and RF and SMO for R2L class were the best values when the FPR values obtained from the classifiers for each class were examined. For MLP, 0 is taken for all classes. When Precision metric values of classifiers are examined, it is seen that for DDoS class, full value for bN, RF, SMO and MLP. T
Web Link	http://ieeexplore.ieee.org/document/8090165/?reload=true

Table 6: Research Paper 6

Paper 7

TITLE OF PAPER	Evaluation of Machine Learning Algorithms for Intrusion Detection System
AUTHORS	Mohammad Almseidin, Maen Alzubi, Szilveszter Kovacs and Mouhammd Alkasassbeh
YEAR OF PUBLICATION	2017
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	Intelligent Systems and Informatics (SISY), 2017 IEEE 15th International Symposium on
SUMMARY	In this paper, several experiments were performed and tested to evaluate the efficiency and the performance of the following machine learning classifiers: J48, Random Forest, Random Tree, Decision Table, MLP, Naive Bayes, and Bayes Network. All the tests were based on the KDD intrusion detection dataset. The rate of the different type of the attacks in the KDD dataset are approximately 79% of DOS attacks, 19% of normal packets and 2% of other types of attacks (R2l, U2R and PROBE). In the experiments 148753 instances of records have been extracted as training data to build the training models for the selected machine learning classifiers. The testing phase is implemented based on 60000 random instances of records. Several performance metrics are computed (accuracy rate, precision, false negative , false positive, true negative and true positive).

Web Link	http://ieeexplore.ieee.org/document/8080566/authors

Table 7: Research Paper 7

Paper 8

TITLE OF PAPER	A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection
AUTHORS	Anna L. Buczak , Erhan Guven
YEAR OF PUBLICATION	2015
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	IEEE Communications Surveys & Tutorials
SUMMARY	The paper describes the literature review of ML and DM methods used for cyber. Special emphasis was placed on finding example papers that describe the use of different ML and DM techniques in the cyber domain, both for misuse and anomaly detection. Unfortunately, the methods that are the most effective for cyber applications have not been established; and given the richness and complexity of the methods, it is impossible to make one recommendation for each method, based on the type of attack the system is supposed to detect. When determining the effectiveness of the methods, there is not one criterion but several criteria that need to be taken into account. They include (as described in Section VI, Subsec-tion C) accuracy, complexity, time for classifying an unknown instance with a trained model, and understandability of the final solution (classification) of each ML or DM method. Depending on the particular IDS, some might be more important than others.
Web Link	http://ieeexplore.ieee.org/document/7307098/

Table 8 :Research Paper 8

Paper 9

TITLE OF PAPER	State-of-the-art Survey on Cloud Computing Security Challenges, Approaches and Solutions
AUTHORS	Shahzad, Farrukh

YEAR OF PUBLICATION	2014
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	Procedia Computer Science 37 (2014): 357-362. The 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014)
SUMMARY	The revolution of cloud computing has provided opportunities for research in all aspects of cloud computing. We presented the five essential characteristics of cloud computing, three cloud service models, and four cloud deployment models. Research in the secure cloud storage is compounded by the fact that users data may be kept at several locations for either redundancy/ fault tolerance or because the service is provided through a chain of service providers. We explored the security measures adopted by the largest cloud service provider (Amazon web services or AWS) including their infrastructure security and security best practices followed by AWS.
Web Link	http://www.sciencedirect.com/science/article/pii/S1877050914010187

Table 9: Research Paper 9

Paper 10

TITLE OF PAPER	An Encryption, Compression and Key(ECK) Management based Data Security Framework for Infrastructure as a Service in Cloud
AUTHORS	Sanket Salvi, Sanjay H.A, Deepika K.M
YEAR OF PUBLICATION	2015
PUBLISHING DETAILS WHERE THIS PAPER WAS PUBLISHED	Advance Computing Conference (IACC), 2015 IEEE International
SUMMARY	The framework provides the subscribed users with the privileges of accessing the virtual machines in more secured way by compressing and encrypting the complete VM. It provides an architecture that ensures the complete VM security including data within it. It provides users with essence of privacy protection even

	when third party owns the data. Also we have found that, LZ4 tool for compression provides most optimized performance for the designed framework. Blowfish and LZ4 combined together had provided with consistently faster performance in our proposed framework. It successfully provides a level of isolation and abstraction for the subscribed VMs to avoid data breaches and privacy violation
Web Link	http://ieeexplore.ieee.org/document/7154830/

Table 10: Research Paper 10

2.1.2 Integrated Summary of Literature Studied

Cloud computing is a information technology (IT) paradigm, a framework for validating universal access to shared pools of configurable assets, (for example, PC servers, servers, storage, applications and services),which can be quickly provisioned with negligible administration exertion, regularly finished the Internet. Cloud computing permits clients and endeavors with different figuring capacities to store and process information either in an exclusive cloud, or on an outsider server situated in a server system - in this way making information getting to numerous mechanisms more productive and dependable. Cloud computing depends on sharing of assets to accomplish intelligibility and economy of scale, like an utility.

Software as a Service (SaaS). The ability gave to the buyer is to utilize the supplier's applications running on a cloud foundation. The applications are open from different customer devices through either a thin customer interface, for example, a web program (e.g., electronic email), or a program interface. The purchaser does not oversee or control the basic cloud foundation including system, servers, working frameworks, stockpiling, or even individual application abilities, with the conceivable exemption of constrained client particular application arrangement settings.

Platform as a Service (PaaS). The ability gave to the client is to send onto the cloud framework purchaser made or gained applications made utilizing programming dialects, libraries, administrations, and instruments and tools bolstered by the supplier. The client does not oversee or control the fundamental cloud foundation including system, servers, working frameworks, or capacity, yet has control over the sent applications and perhaps design settings for the application-facilitating condition.

Infrastructure as a Service (IaaS). The capacity gave to the client is to arrangement handling, storage, systems, and other major registering assets where the client can send and run subjective programming, which can incorporate working frameworks and applications. The purchaser does not oversee or control the hidden cloud foundation but rather has control over working frameworks, stockpiling, and sent applications; and conceivably restricted control of select systems administration parts

With the expanding utilization of cloud advances there is an inborn requirement for security on different levels. Security concerns related with distributed computing fall into two general classes: security issues looked by cloud suppliers and security issues looked by their clients (organizations or associations who have applications or store information on the cloud). The developer must guarantee that their framework is secure and that their clients' information and applications are ensured, while the client must take measures to strengthen their application and utilize solid passwords and verification measures.

Multi-tenure and detachment is a noteworthy measurement in the cloud security issue that requires a vertical arrangement from the SaaS layer down to physical foundation (to create physical alike limits among inhabitants rather than virtual limits at present connected).

Security management is extremely basic to control and deal with this number of necessities and controls. The cloud model ought to have a comprehensive security wrapper with the end goal that any entrance to any question of the cloud stage should go through security parts first.

Recently, advances in machine learning techniques have attracted the attention of the research community to build intrusion detection systems (IDS) that can detect anomalies in the network traffic. Learning-based approaches may prove useful for security applications as such models could be trained to counter a large amount of evolving and complex data using comprehensive datasets. Such learning models can be incorporated with firewalls to improve their efficiency. A well trained model with comprehensive attack types would improve anomaly detection efficiency significantly with a reasonable cost and complexity.

The main Features of any security system is to ensure 4 main features namely maintaining integrity of the data, confidentiality, non- repudiation and availability. These problems can be solved by a variety of measures which includes but are not limited to compression and encryption of data, user authentication while registration and also each login and access attempt.

3D security, a relatively new concept, is a fresh and innovative solution for providing a new

dimension for implementing user authentication. According to this technique, access to the cloud is authenticated using a graphical password. The Graphical password is generated by considering many aspects and confidential inputs of images.

Chapter-3 Analysis, Design and Modeling

3.1 Overview of Project

This project focuses on providing a complete security solution for an entity for their cloud platforms. Our solution utilizes Database as a service (DAAS) sourced from godaddy.com, for storing files as well as user databases on the cloud. The software has been designed using C#, and visual studio 2010 as the IDE. The various features provided include user and admin login, Encryption, Compression, Folder locks, file sharing and downloading by users. Even on registration of a new user, the user won't be able to access any file until and unless his id is approved by the admin. File sharing by a user can only be to another user registered in the same system and not other.

Login authorization by OTP ensures security on database and user level as one can't access data without proper credentials but also for some reason if they do the encryption ensures that the person can't make any meaningful information of the data.

Machine learning algorithms (Decision Tree, Naïve Bayes and a hybrid of the two) have been used for designing an intrusion detection system. For training of said system dataset from UNSW has been taken which contain values such as IP Addresses, Source Ports, Protocol used etc. For preventing and detecting intrusions the MAC and IP addresses of every user registered is stored and whenever an unauthorized access is attempted the file is not downloaded from the server and also an alert if sent to the admin account.

Further, we have implemented real time tracking of the hits from any IP address on any device that is registered on the network and are adding them to the database to expand our database and get a more accurate result. This feature allows our project to be fully operational in any corporate environment now and not relying only on the UNSW dataset.

3.2 Functional and Non-Functional requirements

3.2.1 Functional Requirements

3.2.1.1 Monitoring (resource, service, and dead/alive)

The requirement is to monitor the usage and collect data on the status of all of the network components/resources on the cloud and collect them. and further determine the need for a recovery of disaster option or need to distribute the load. the following requirements are necessary for monitoring the status

- Collect resource information after fixed intervals of time which includes information about each server and each memory/storage unit for each functionality provided by the system on the cloud to check the compliance of each service and decide weather the load needs to be distributed to keep the performance up to mark.
- It must be possible to collect this periodic information and share it with other cloud systems through a common format. in order to keep check in the services spanning the cloud system interwork.
- It must be possible to arrange secure environment for the service and for monitoring attacks on the cloud system from external sources or from internal sources and take actions accordingly

3.2.1.2 Provisioning

To determine the resource essential for the system (Quantity, type of resources etc.) on the basis of quality demands of the service and the resources necessary to keep up the performance when a degradation in performance is detected. The functional requirements for provisioning for providing and maintaining cloud performance are:

- It must be feasible to detect a bottleneck in the system when the load on the cloud varies and to plan accordingly as the operating characteristics varies from app to app.
- It should be attainable to create numerous plans like giving priority to exactness or immediate responses, since the exactness of the resource needed and the time allowed for creating a choice differs between the provisioning for determinative the initial resource configuration of a system on the cloud.

3.2.1.3 Resource management

To manage the resources those are within our own cloud system securely or in other cloud system for providing each service. The requirements are as follows:

- It should be feasible to characterize extra/supporting information in a standard manner, like the resources status and its type. For managing the resources over multiple clouds into an integrated

form.

- It should to be attainable to manage numerous resource configurations for every service, like servers, storage units and networks, in association.
- It should be attainable to upgrade the information regarding the configuration of resources over multiple cloud systems at the same time as the events (securement or release of resources from other cloud systems) between cloud systems.
-It should be possible to store information regarding the change in the cloud system to understand the difference in resource configuration when it has been changed

3.2.1.4 Releasing resources

This function is to evaluate that distribution or recovery isn't any longer required based on watching results and to unharness excess resources, when disaster recovery or load distribution has been adopted. the need for releasing secured resources once they become to be spare are represented below.

-It shall be potential to close up virtual machines or applications that were activated once the secured resources began to be used, to update resource management info, and to fully delete or collect transferred information. It shall be attainable to release networks when servers and storage units are free. It shall even be possible to gather any work remaining in different cloud systems.

3.2.1 Non-Functional Requirements

Availability- The time in terms of percentage is what availability is, the time that the cloud vendor assures of the availability of the cloud services. This includes schedule maintenance down-times. Also 99% of the availability sounds safe which is almost equal to 3.5 days of potential down time annually. 99% could also be taken as 8 hours of downtime. This is best when considered as a part of disaster recovery in the avenue of availability, this is usually when more than one data (physical) center is where it is residing. It is especially affirmative when data residency is a grave concern when legal and regulatory reasons interfere.

Elasticity (Scalability) – How easy is it to bring on line or take down compute resources (CPU, memory, network) as workload increases or decreases.

Interoperability – It talks about utilizing services from numerous cloud providers how pundit it is to shift workloads between cloud providers. It also discusses about if you wish to migrate from one cloud provider to other.

Security – Security levels and standards are being discussed in detail; their level in public/private clouds not in the data center and also taking in the consideration of physical security to the club providers data centers and networks. It is essential to keep in mind about data residency here as well.

Adaptability – The level of ease of extension, addition and growth of services required in a business. For instance, if I wish to switch my business processes or connect to new back vendor external API's what levels of adaptability and ease would I witness.

Performance – From the infrastructure of the cloud to the workload support system, there has to be check on how well suited is the cloud factors. This particularly a crucial factor when there is workload growth or frequent attacks.

Usability – This entirely depends on what kind of client is using the technology. It could be some business users, developers, architects or IT operations. Considering the fusion of the software and its adaptability is required. There are no hidden rooms in IT industry; everything is transparent and clear for the world to witness.

Maintainability – This is seen from an IT operations and developer point of view
How easy is it to manage (and develop) the cloud services.

Integration – Today is the world of hybrid clouds where it is essential to keep the data and workload in the respective data center while some portion of it can be on public as well as private clouds. Understanding how these clouds integrate is a crucial point.

3.3 Overall architecture with component description and dependency details

-DaaS:

Desktop-as-a-service (**DaaS**) is a type of virtual desktop infrastructure (VDI) in which the VDI is outsourced and taken care of by an outsider like a third party. Likewise called facilitated work area administrations, work area as-a-benefit is as often as possible conveyed as a cloud benefit alongside the applications required for use on the virtual work area.

- SaaS:

Software as a service (SaaS) is a software authorizing and conveyance model in which programming is authorized on a membership premise and is halfway facilitated. It is at times alluded to as "on-request software"

- PaaS

Platform as a service (PaaS) or application stage as a service (aPaaS) is a classification of cloud computing services that gives a stage enabling clients to create, run, and oversee applications without the multifaceted nature of building and keeping up the framework ordinarily connected with creating and propelling an application.

- IaaS

Infrastructure as a service (IaaS) is another type of cloud computing that gives virtualized figuring assets over the web. IaaS is one of the three fundamental classifications of cloud computing services, nearby Software as a server (SaaS) and platform as a server (PaaS).

- GoDaddy cloud server API

The GoDaddy Cloud Server API is an API for developers made by developers. It is as of now in Beta status and is to be a straightforward API with quick provisioning and KVM Virtualization. The objective of this new GoDaddy Cloud Server API is to bring better cloud facilitating arrangements, with enhanced speed and versatility for designer activities and applications.

SalesForce REST API

- REST API gives a capable, advantageous, and basic Web services API for collaborating with Force.com. Its points of interest incorporate simplicity of mix and advancement, and it's a magnificent decision of innovation for use with versatile applications and Web 2.0 activities. Be

that as it may, in the event that you have numerous records to process, consider utilizing Bulk API, which depends on REST standards and streamlined for substantial arrangements of information.

- UNSW-NB15 Dataset

One of the real difficulties in this field is the inaccessibility of a complete system based informational collection, which can reflect present day organize activity situations, tremendous assortments of low impression interruptions and profundity organized data about the system movement. Assessing system interruption location frameworks explore endeavors, KDD98, KDDCUP99 and NSLKDD benchmark informational collections were produced 10 years back. In any case, various current examinations demonstrated that for the present system danger condition, these informational indexes don't comprehensively reflect organize movement and current low impression assaults. Countering the inaccessibility of system benchmark informational index challenges, UNSW-NB15 informational collection is utilized. This informational index has a half and half of the genuine current typical and the contemporary combined assault exercises of the system movement. Existing and novel techniques are used to create the highlights of the UNSWNB15 informational index. This informational index is accessible for examine purposes.

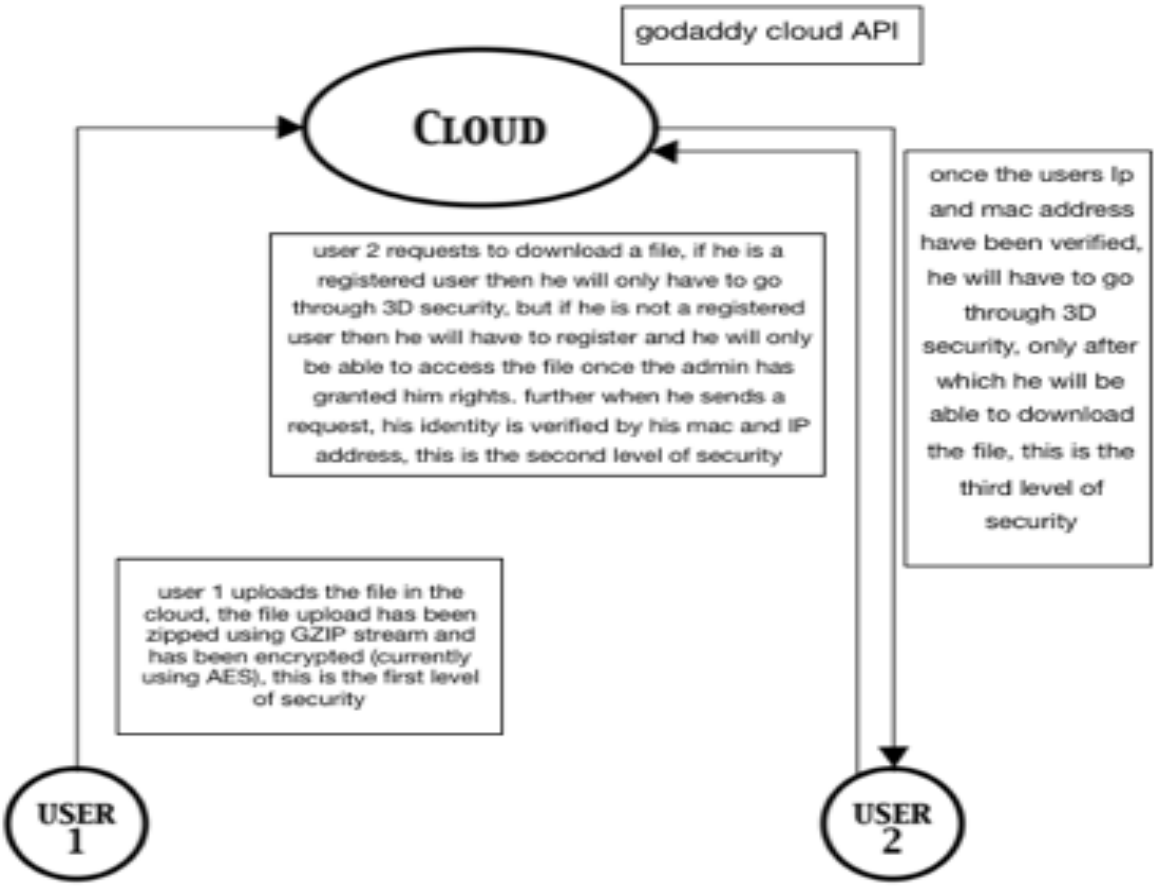

Fig 3.1 Overall Architecture

3.4 Design Documentation

3.4.1 Use Case diagrams

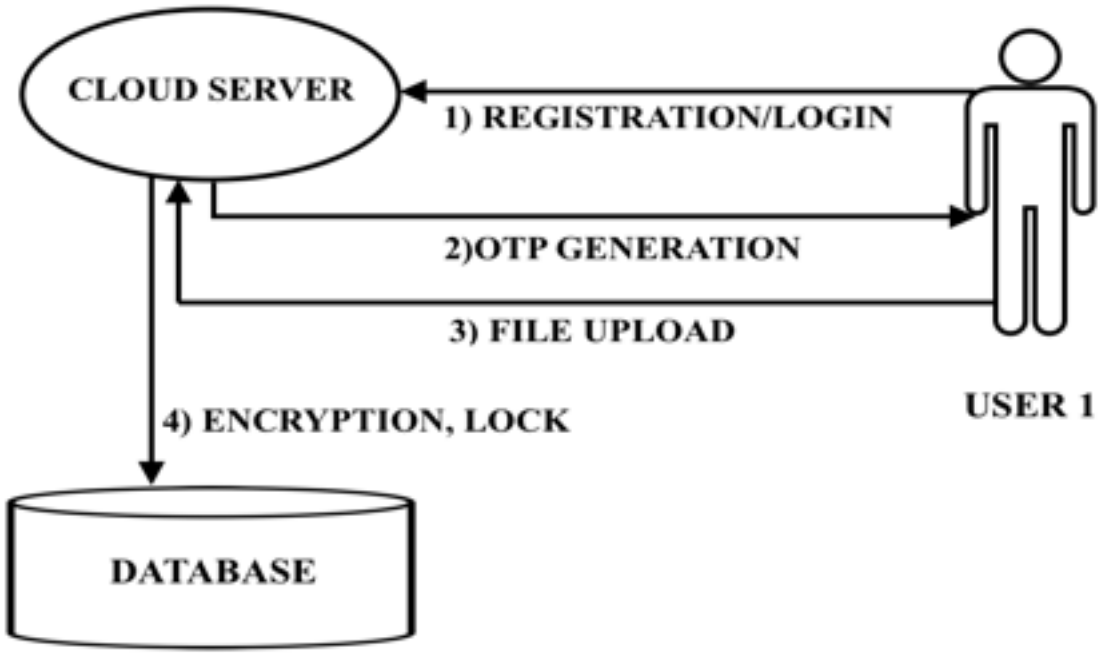

Fig 3.2 File Upload

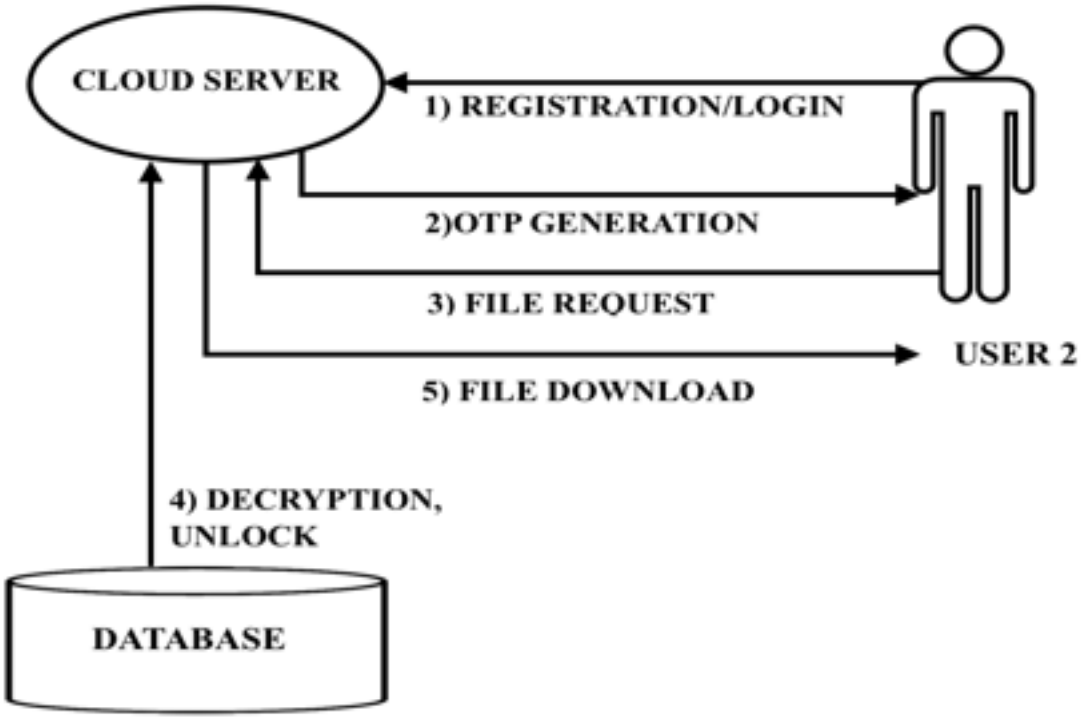

Fig 3.3 File Download

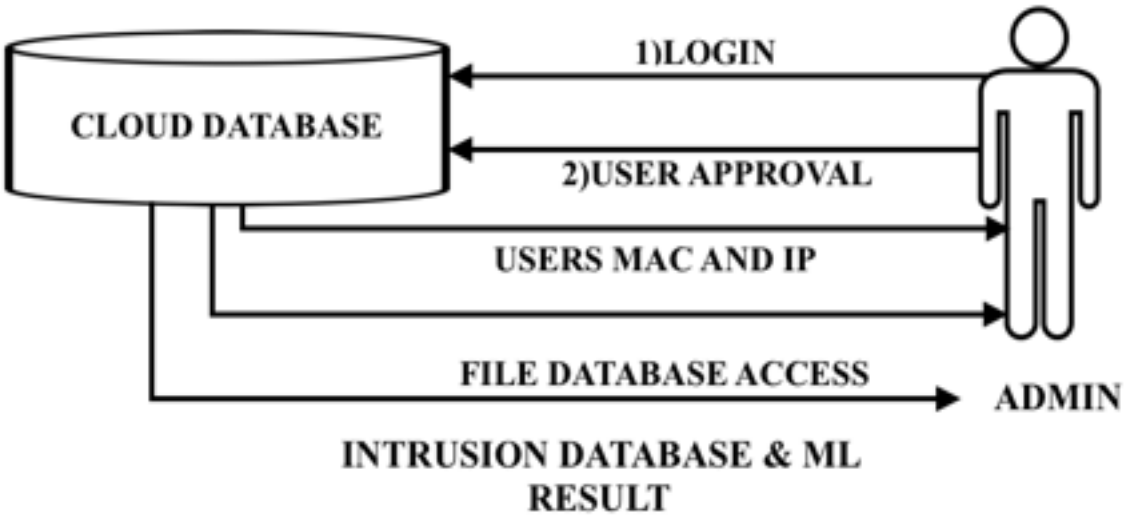

Fig 3.4 Admin Access

3.4.2 Activity Diagrams

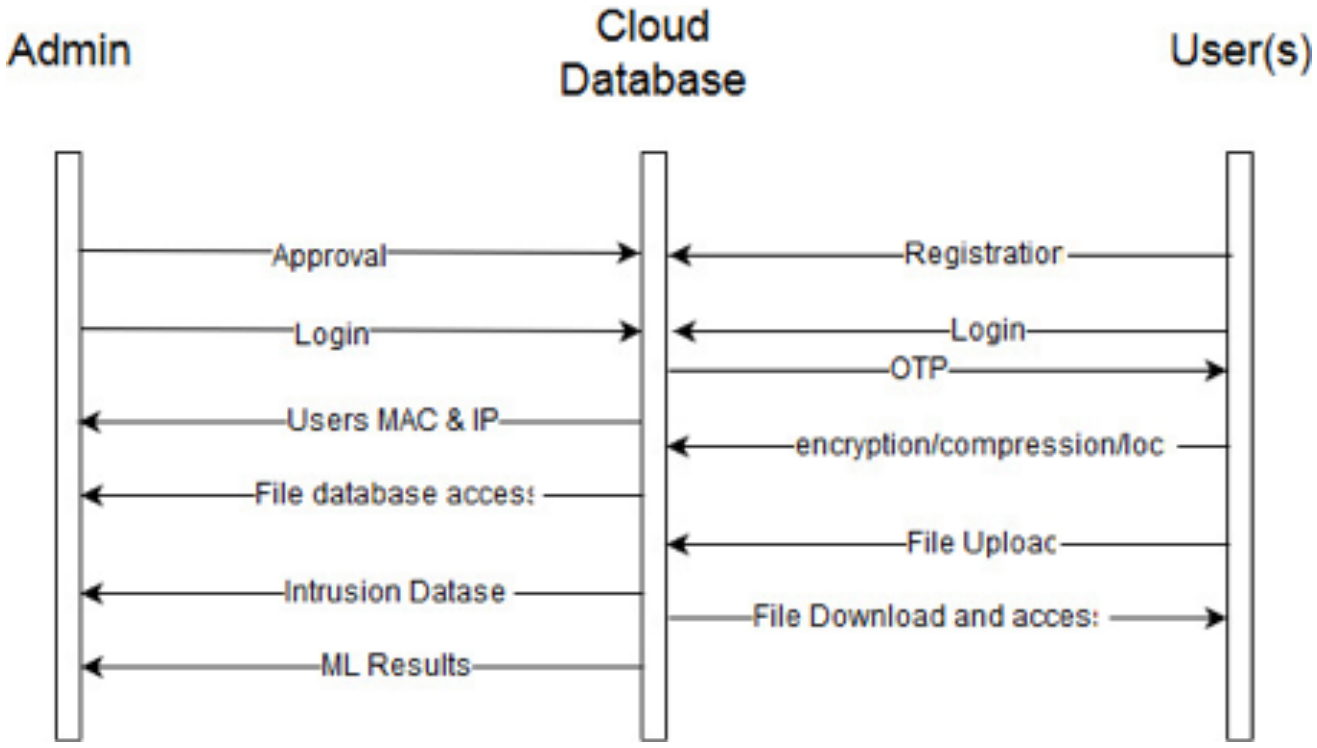

Fig 3.5 Activity Diagram

3.4.3 Sequence Diagrams/Activity Diagrams

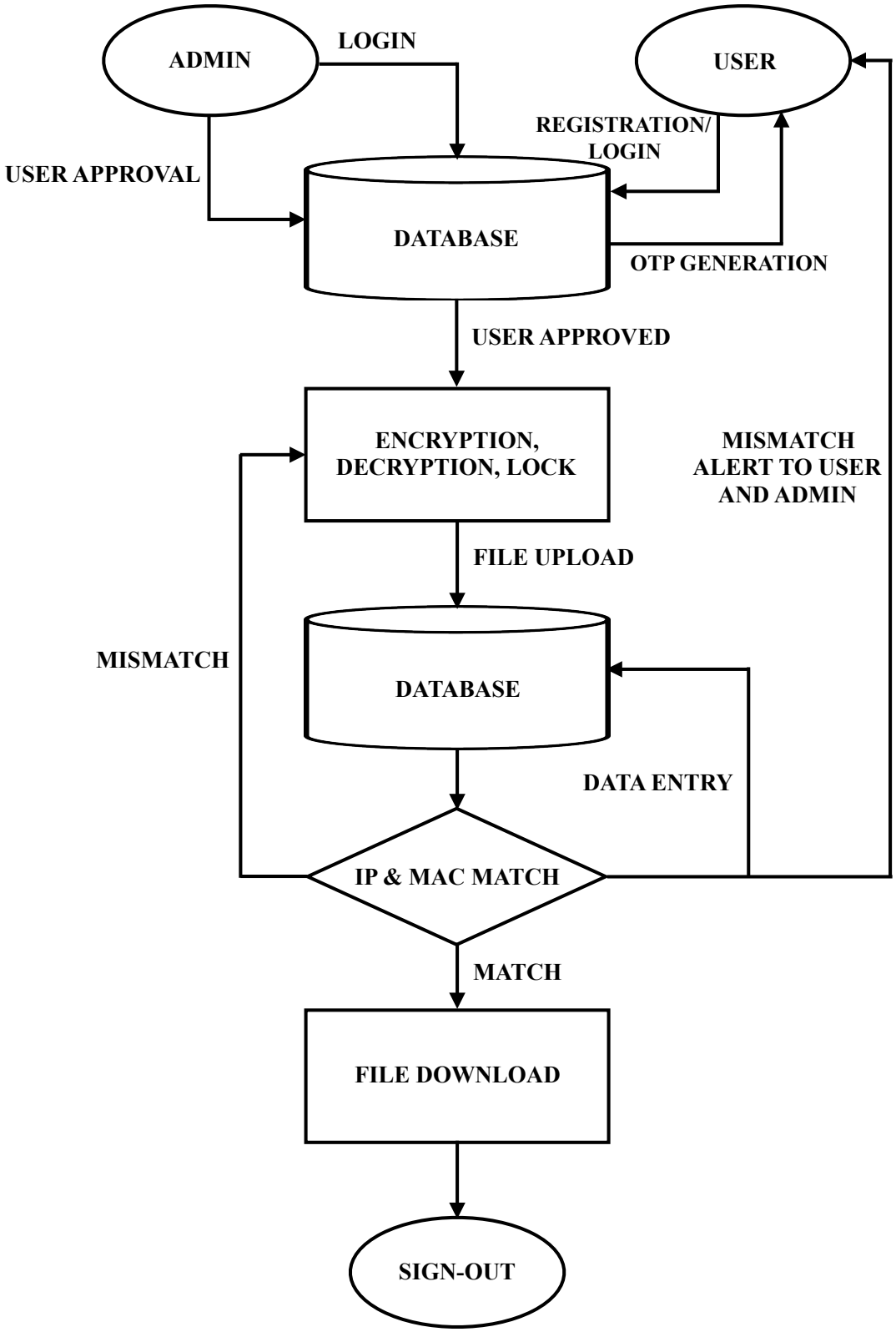

Fig 3.6 Control Flow

3.4.4 Algorithms

-Decision Tree

Decision tree is a structure resembling a flowchart in which each of the parent nodes represents a test on a attribute, and each child nodes represents a outcome of the test. and each of the leaf notes represent a label of the class to which the object belongs. Decision trees are commonly used in operations research and operations management.

Decision tree uses Entropy values. In each step it uses the computed value of the Entropy to decide which child node is better and proceeds with it.

- Naive Bayes

Naive Bayes is a probabilistic classifier used in machine learning, which is based on the principles of Bayes theorem. It is an algorithm, which has been studied in depth in the 1950's. It is still a popular method for baseline text categorization and to label documents weather they belong to one label or another based on word frequency as the feature. Using appropriate pre processing methods, naive Bayes can still be up to the mark with new algorithms such as SVM.

- Hybrid Algorithm

This is a algorithm proposed by us, it is a hybrid between Naive Bayes classifier and decision tree, first the probabilistic model (Naive Bayes) is used to compute the probabilities of an attack based on UNSW database, then those probabilities are further used to classify the attacks using the entropy in decision tree to predict vulnerability, IP addresses that are prone to attack, and IP addresses that are trying to intrude. Currently we have only developed the algorithm, we are yet to test its accuracy and precision under different stress conditions, this will be done in the next phase of the project, in which our algorithm will be stress tested and will be tested using various datasets.

-GZIP

Gzip algorithm is based on the DEFLATE algorithm, which in itself is a combination of two algorithms, that re Huffman coding and LZ77.

Gzip is a 10-byte header format containing compression id, file flags, compression flags, timestamp and operating system ID

-AES Encryption

The AES algorithm also known as Rijndale is an encryption algorithm that was established in the year 2001. It is a block cipher with blocks of size 218 bits i.e. 4 words. The key size can vary between 128, 192 or 256 bits. the algorithm operates on a 4x4 matrix and makes use of the principle of substitution and permutation.

The algorithm has 10,12 or 14 rounds depending on the key size each round contains 4 basic steps that are 1) Substituting bytes 2) Shifting rows 3) Mixing columns 4) Adding the round key. It is a widely used and highly secure encryption algorithm

- DES

Des is a data encryption algorithm. It is a symmetric key algorithm, which was developed by IBM in the 1970's. It is a fiesta cipher, which has 16 rounds and also an initial and final permutation, which are inverse on one another. Each block size is of 64. Although DES is no longer a highly secure algorithm and there are advance algorithms such as AES. DES is still a very important algorithm.

Chapter-4 Implementation and Testing

4.1 Implementation details and issues

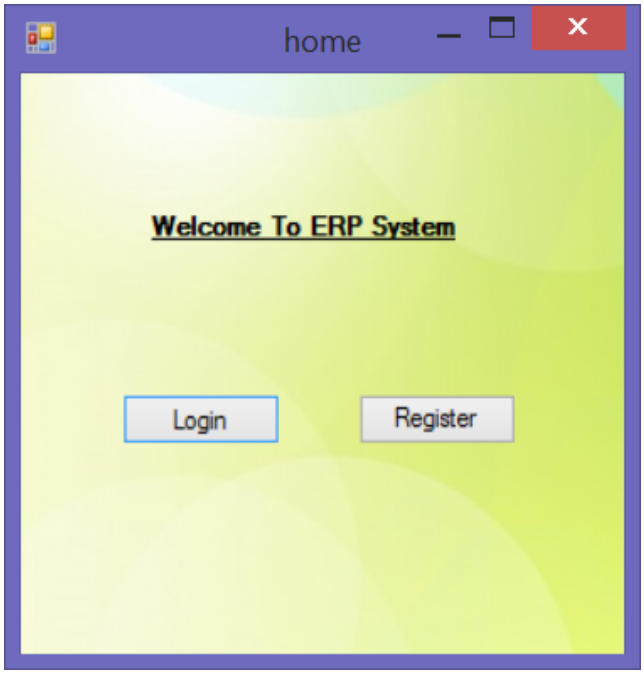

Fig 4.1 Launch Page

Fig 4.2 Login Screen

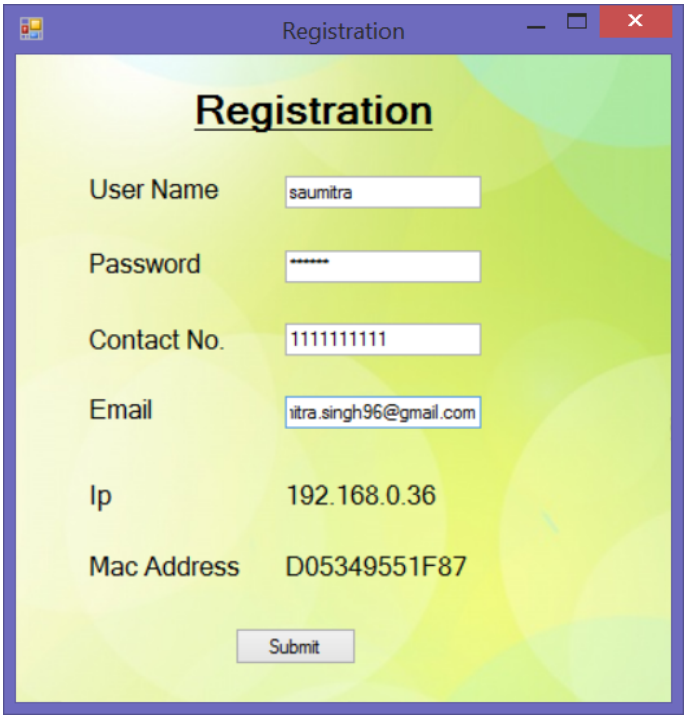

Fig 4.3 New User Registration

After logging in, the administrator has the options of approving users, viewing list of files that have been shared, the attacks that have taken place to access the files by an unauthorised person, and further the administrator can view the classification of these attacks on the basis of 3 algorithms (Naive bayes, decision tree and the hybrid algorithm proposed by us), he can also view the same graphically. and there is also an option to sign out.

Fig 4.4 Admin home screen

the administrator can also view the details of the approved users and the users that are awaiting approval, he can view their email address, contact number, mac address, IP address and their approval status. The admin has the right to change the approval status at any given time.

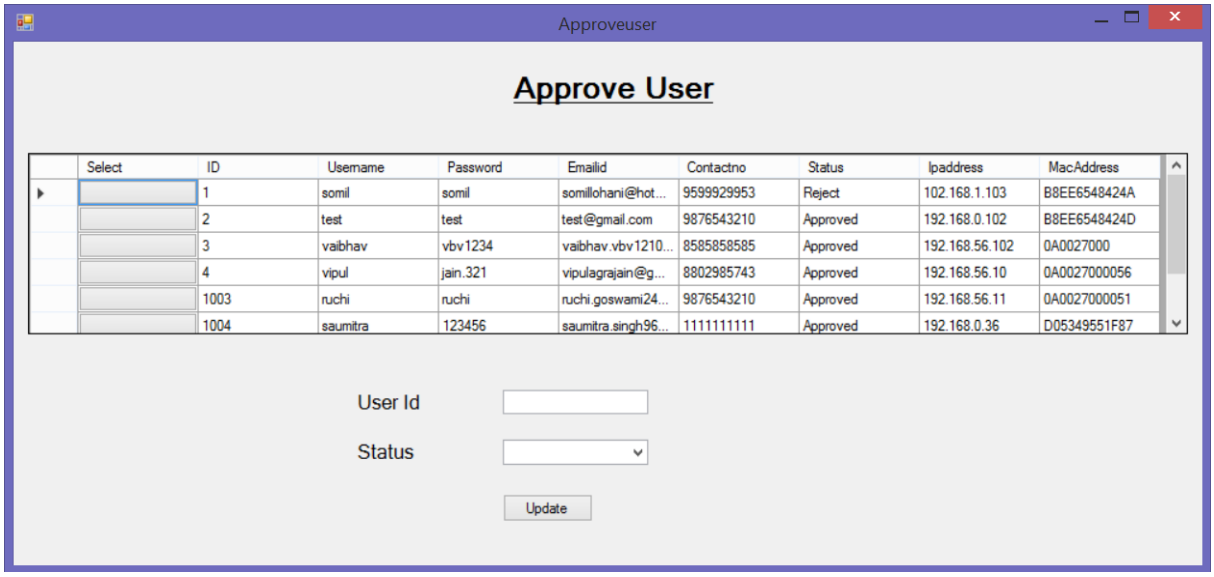

Fig 4.5 Approval Status

The admin may also have a look at the files that are being shared on the network and the details of the sender and the receiver.

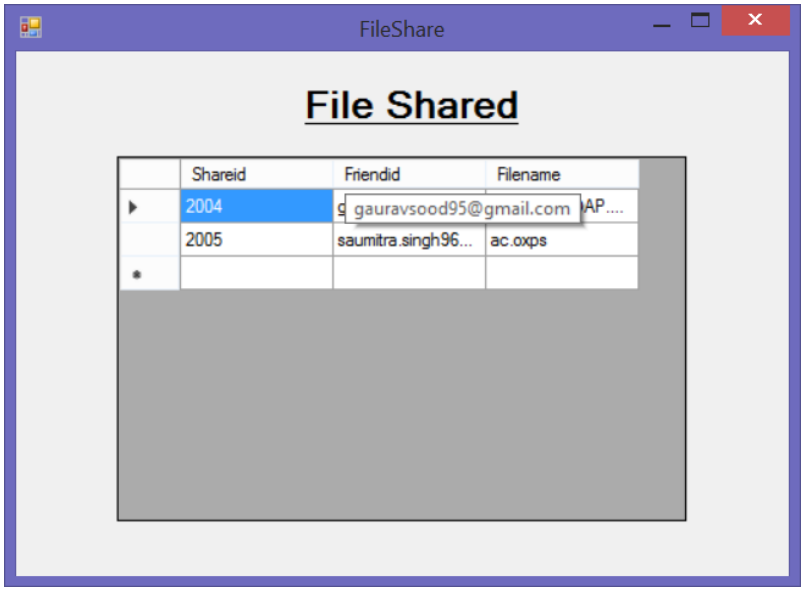

Fig 4.6 Files Shared on Network

A registered user will get an OTP on his email every time he tries to log in. this is an additional security measure, so that the account is not hackable easily.

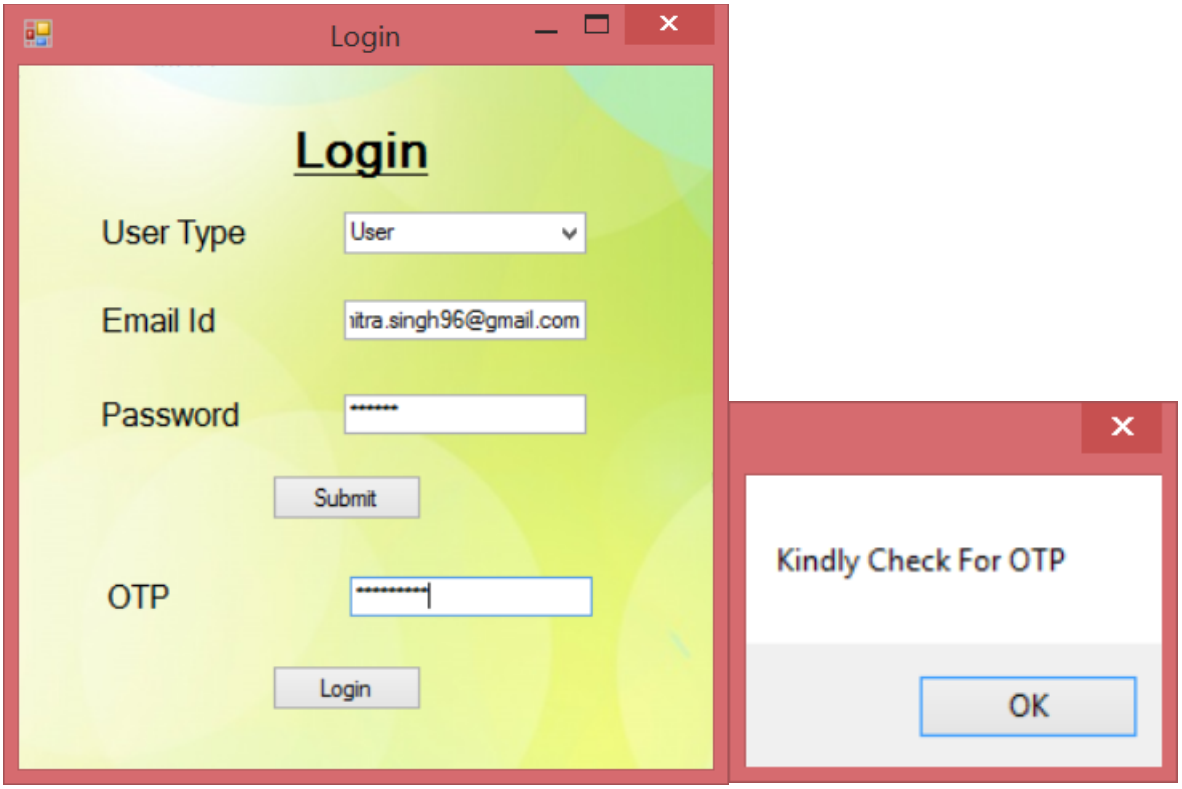

Fig 4.7 User login using OTP

Once logged in, a registered user has the options to lock his file, make a ZIP, share a file or download a file that was shared with him, the user can also encrypt a file that he is sending, he has two options for encryption, he can use either AES or DES encryption techniques

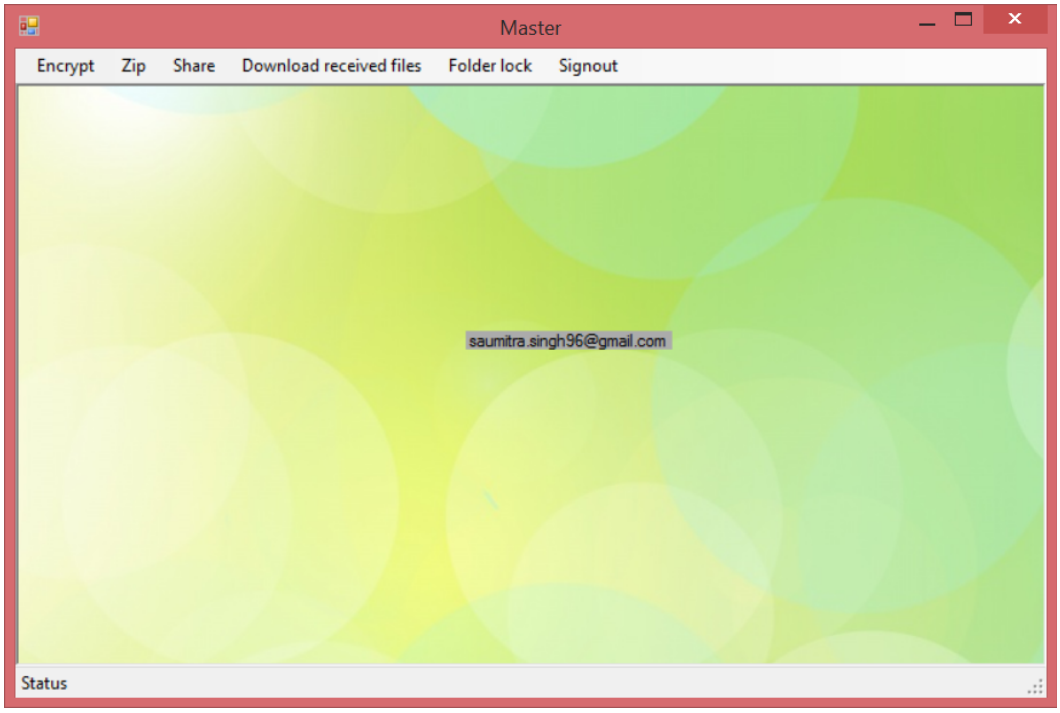

Fig 4.8 User Home screen

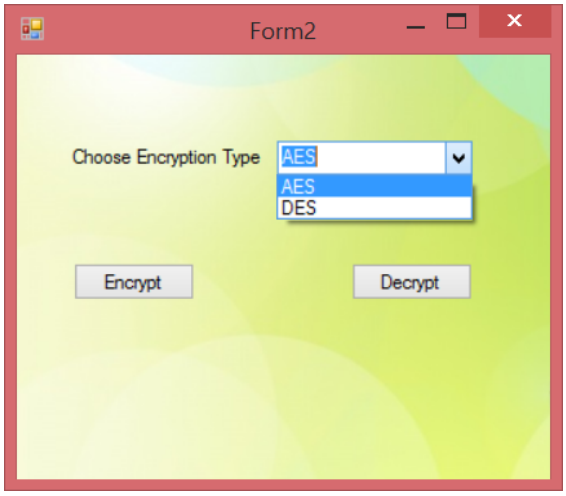

Fig 4.9 File encryption window

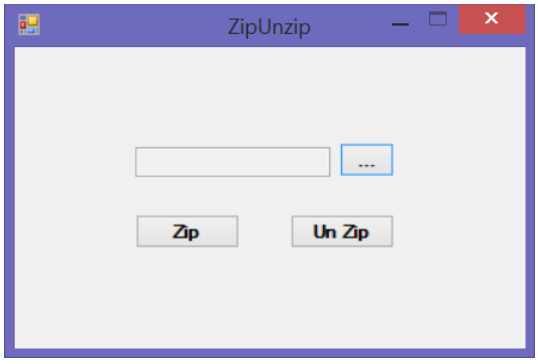

Fig 4.10 File Zip window

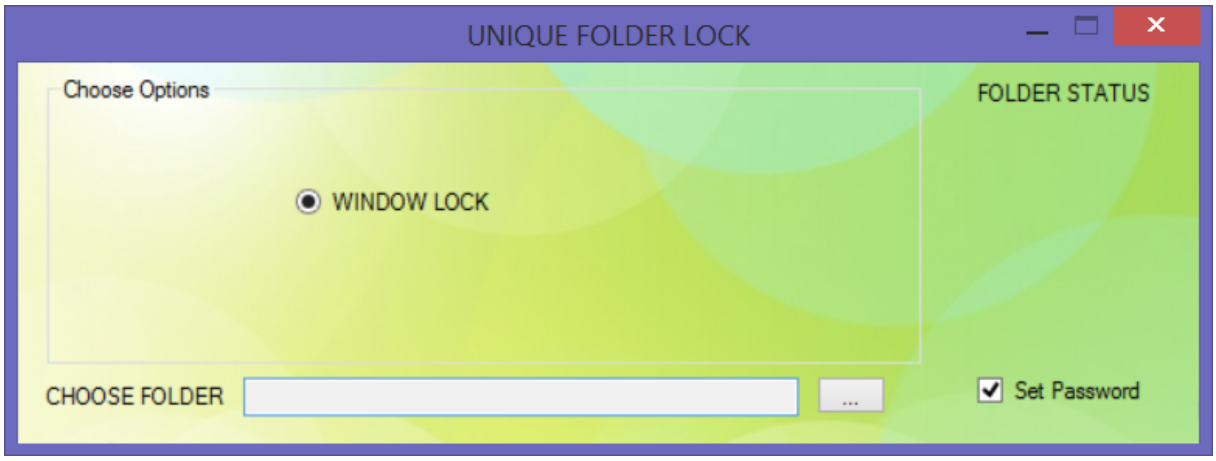

Fig 4.11 File locking window

Fig 4.12 File Sharing Window

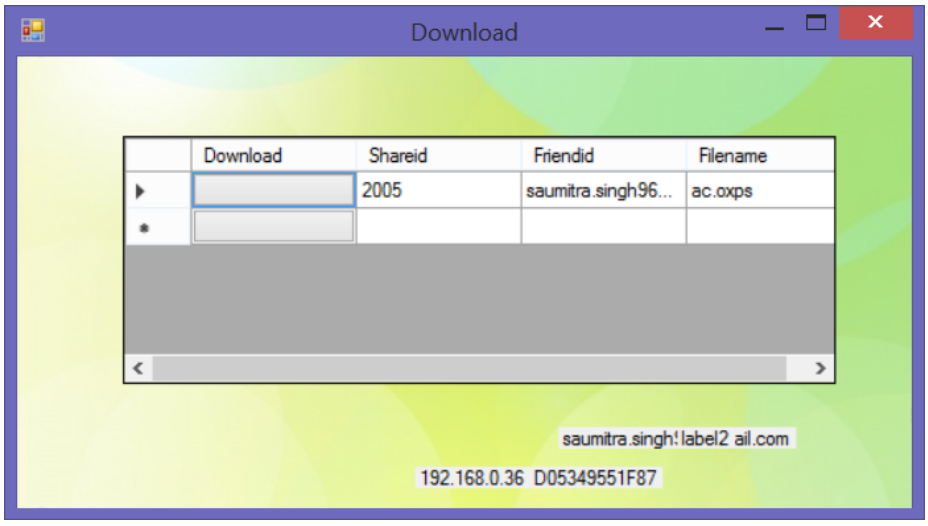

Fig 4.13 Shared file downloading window

The OPT for logging in is sent to the user on his registered email address (as shown in Fig 3.14), if someone other than the authorised person is trying to download the users file then he will receive a notification about the same on his email address (as shown in fig 3.15)

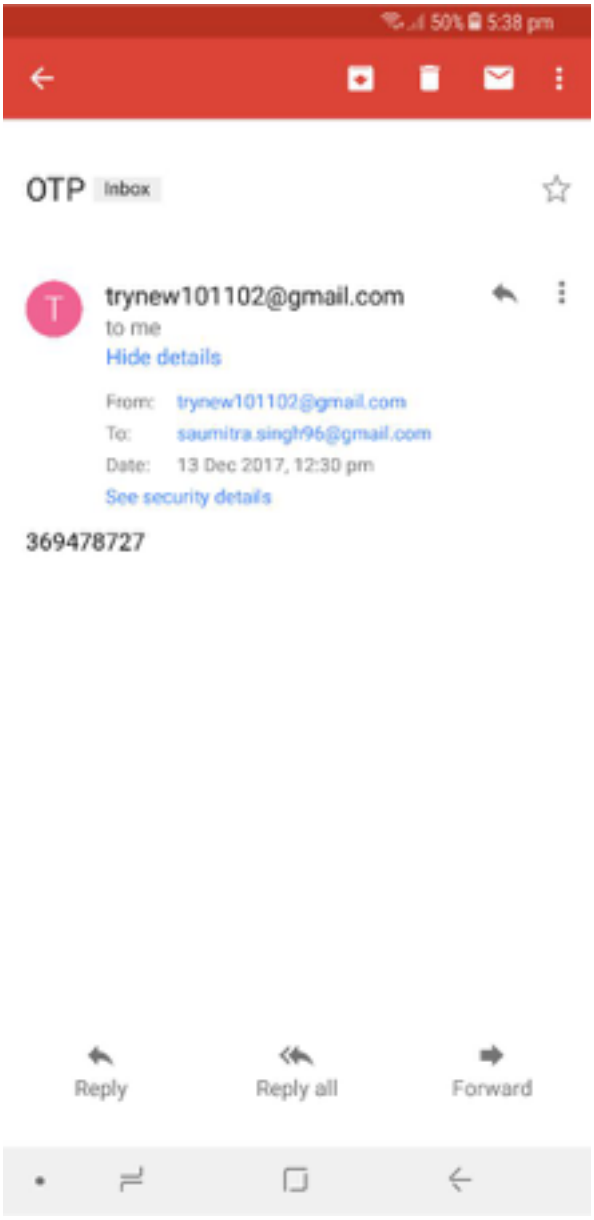

Fig 4.14 OTP for user log in

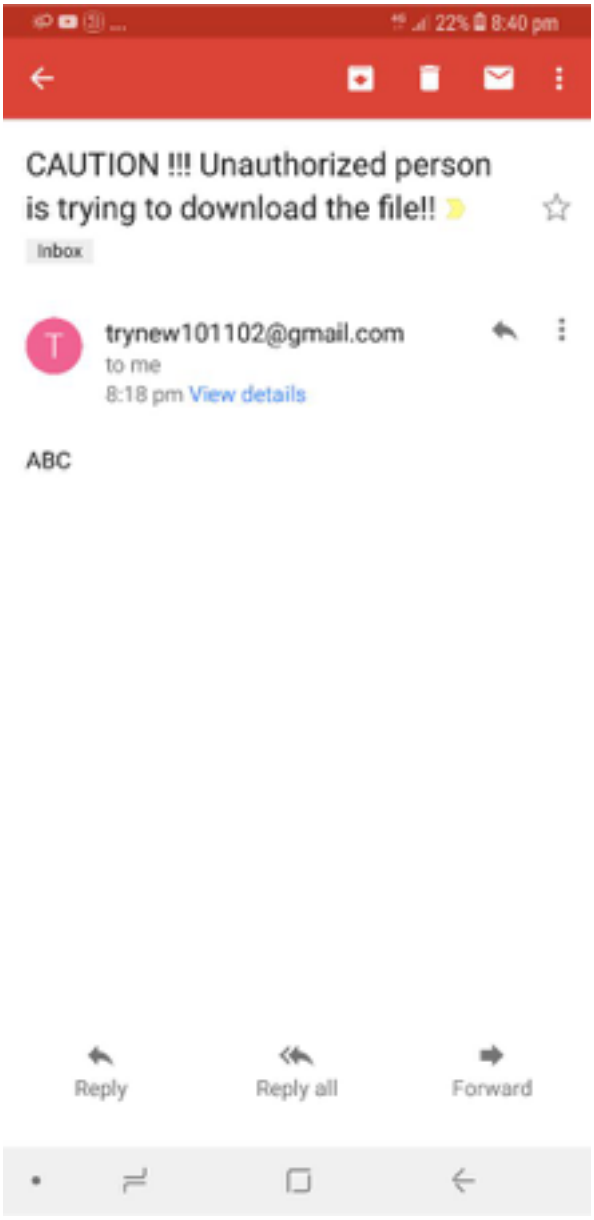

Fig 4.15 Unauthorized access notification

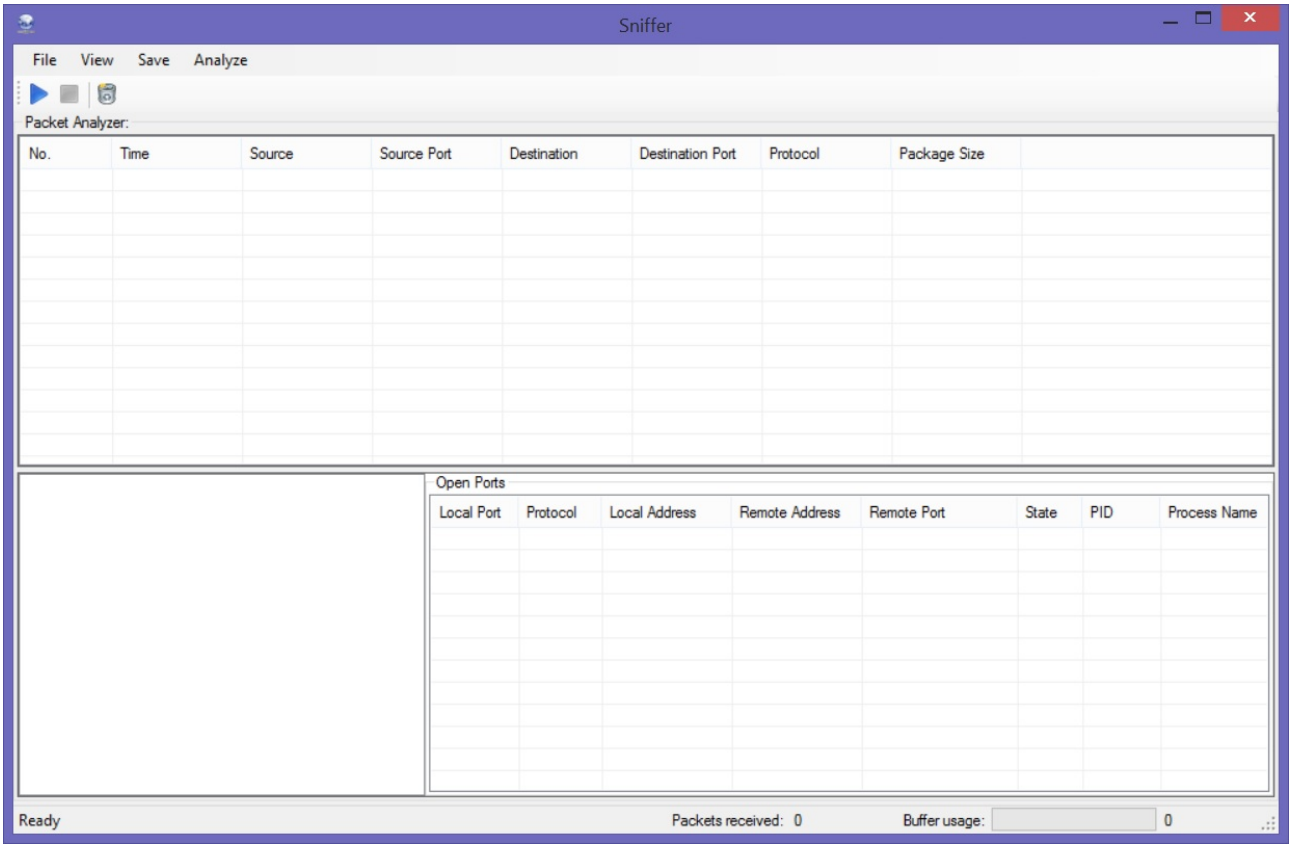

Fig 4.16 Real Time data packet tracker 1

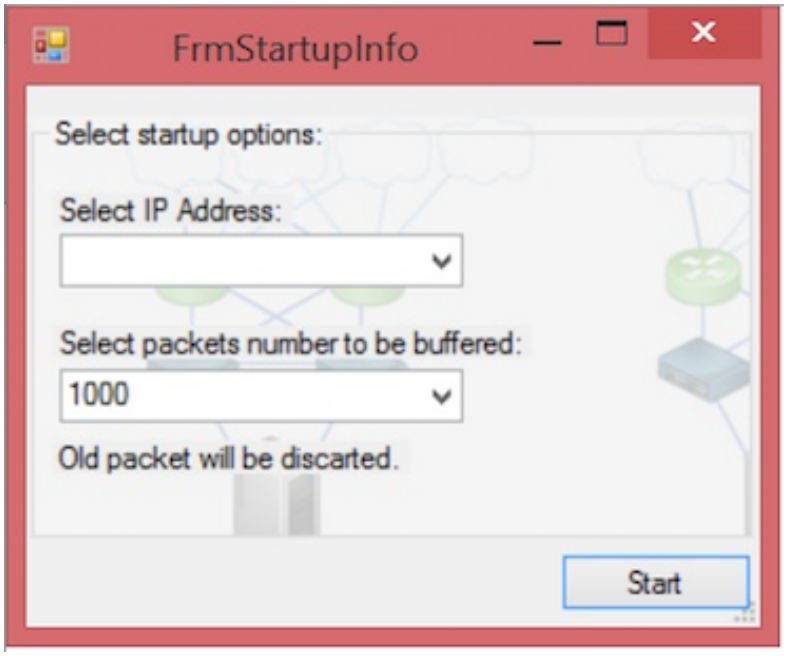

Fig 4.17 Real Time data packet tracker 2

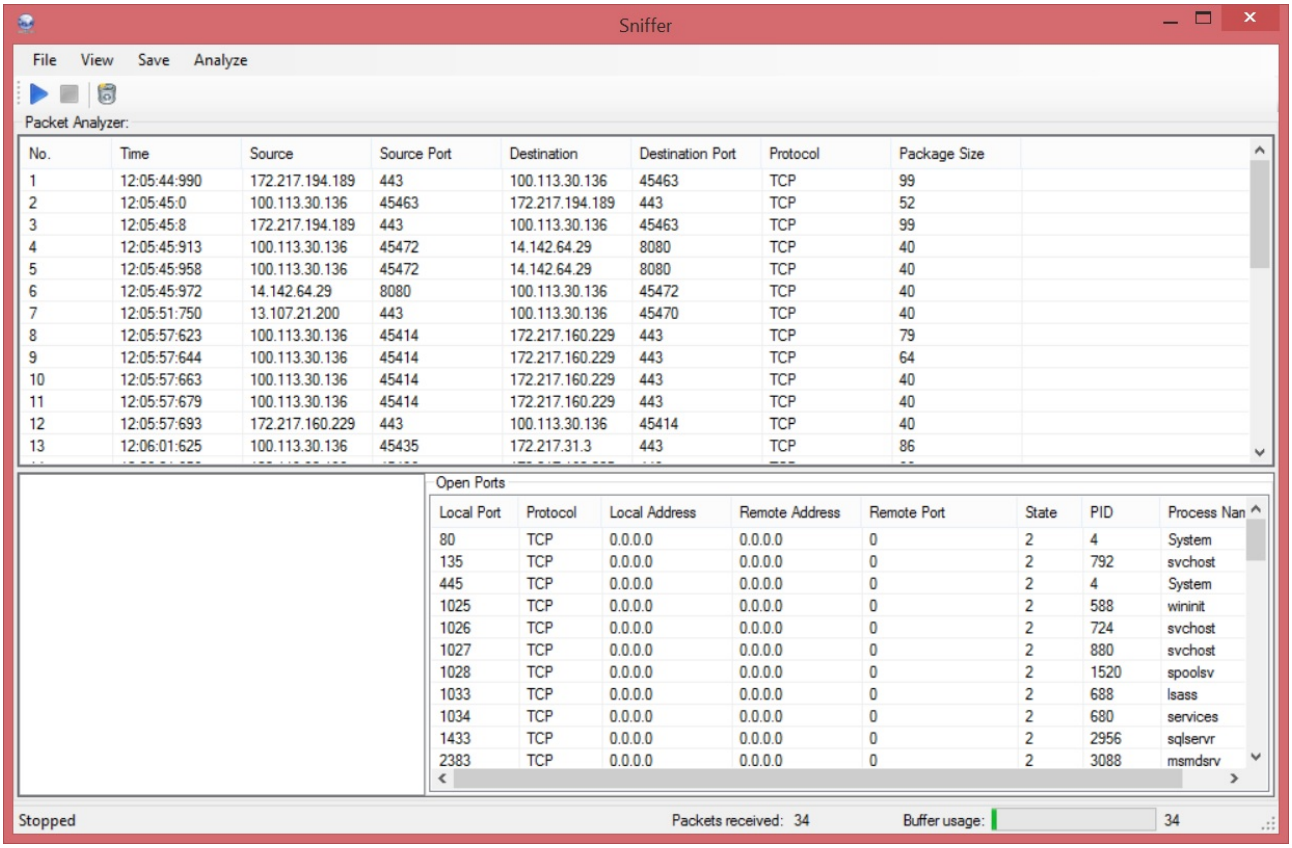

Fig 4.18 Real Time data packet tracker 3

Fig 4.19 Image Steganography

4.2 Testing
4.2.1Testing Plan

Type of Test	Will it be performed?	EXPLANATIONS	Software Component
Requirement Testing	Yes	It is testing the requirements and making sure if they are feasible or they are Not. Because any project or assignment depends on various factors such as time, resources, budget etc. Before we start working on a project it's important to make sure these requirements are met.	Manual work, need to make a plan about all the software requirements, And the time needed to Develop them and the technology that is going to be used.
Unit	Yes	Testing the individual units of the software to make sure that they work as intended	All major functions such as calculating entropy for decision tree, probability for naïve Bayes and hybrid algorithm.
Integration	Yes	Combining the individual software modules and making sure that they are working as intended	Combining and compiling all the classes and testing them as a single code unit
Performance	Yes	Testing to check weather optimal results are achieved	By selecting the non optimal/ non dominant inputs
Stress	Yes	Testing under excessive load i.e.	By selecting the non optimal/ non dominant

		beyond the normal load that it will face.	inputs and testing boundary conditions
Compliance	No	Not Needed	NA
Security	Yes	To determine weather the software protects the privacy and integrity of the data while performing as expected	By using the UNSW-NB15 Dataset and testing against attacks.
Load	Yes	Testing to make sure that the system outputs the expected results under normal load conditions	By selecting large files and rapidly sending them.
Volume	No	Not Needed	NA

Table 11:Testing Plan

4.2.2Test Team Details

Role	Name	Responsibility
Testing Probabilistic values in Naive Bayes, entropy in decision Tree and fitness function in hybrid approach.	Saumitra Vikram Singh	Making sure that the Results matches the expectation and optimal output
Testing the performance of individual modules and the code as a whole.	Gaurav Sood	Produces optimized results

Table 12:Test Team Details

4.2.3 Test Environment

Software Items	Description
Microsoft Visual Studio	Integrated development environment (IDE) on which project was built
Dot Net	.NET Framework
MYSQL	Language for relational database
GoDaddy cloud server	Cloud hosting service
HARDWARE ITEMS	**DESCRIPTION**
Laptop with windows 8	The Software Successfully runs
laptop with windows 7	The Software Successfully runs

Table 13:Test Environment

4.2.4 Component Decomposition and Identification of test cases

S.NO	List of various components (modules) that require testing	Type of testing required	Technique for writing test cases
1.	Entropy Values	Unit	Black Box
2.	Naive bayes probabilistic values	Unit	Black Box
3.	hybrid fitness function values	Performance	Black Box

Table 14: Component test cases

4.2.5 Limitation of Solution

The solution suggested in the project, although sound in its implementation, still has some limitations

1. The solution only uses symmetric encryption techniques(AES, DES) which poses an increased problem of key management. Possible usage of Asymmetric encryption is a solution to this problem.
2. As a user makes use of more and more techniques provided (encryption, compression, folder locks) for sharing of files, the complexity and time efficiency decreases.
3. For intrusion detection the machine learning algorithms used (Decision tree, Naive Bayes, Association rule and hybrid) produce similar results but there is still a problem of false intrusion detection which can cause the system admin to unnecessary probe the files and the users.
4. The solution like the cloud is heavily dependent on Internet facilities so sub-par speeds can cause uploading and downloading of large files to be a tedious and time consuming task.
5. The availability of necessary computers for all employees is a challenge for budding and small companies.

Chapter-5 Findings and Conclusion

5.1 Findings

we have found out the following results when we have applied machine learning algorithms to the database of attacks o our system, we have shown the results for three algorithms, which are Decision tree, Naive Bayes and the hybrid algorithm proposed by us which is a hybrid between Decision tree and Naive bayes. we have further classified the attacks based on eleven attributes which are source IP, Source Port number, Destination IP, Destination port number, the protocol being used, the state, the duration, Dbyte, sttl, dttl, destination port.

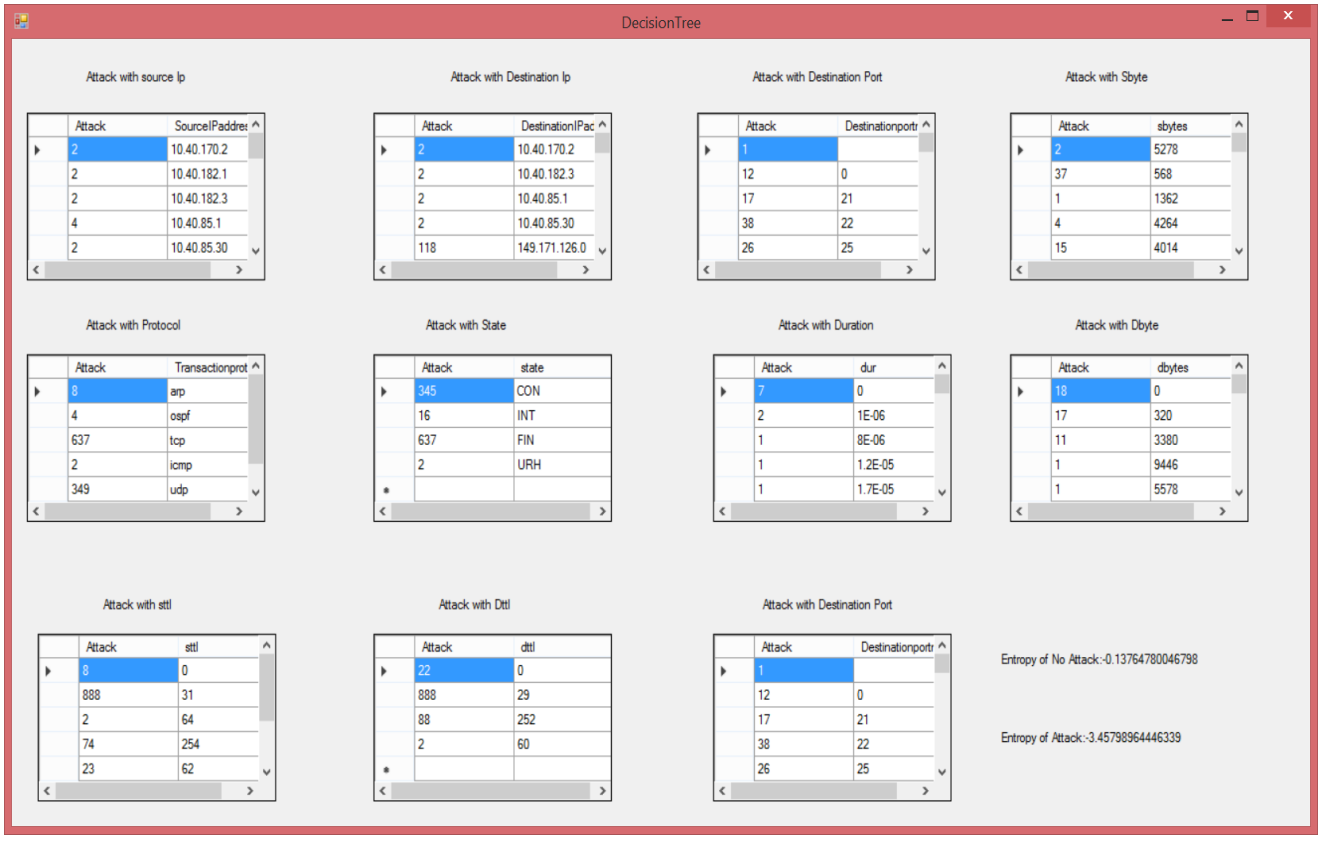

Fig 5.1 Results using Decision tree

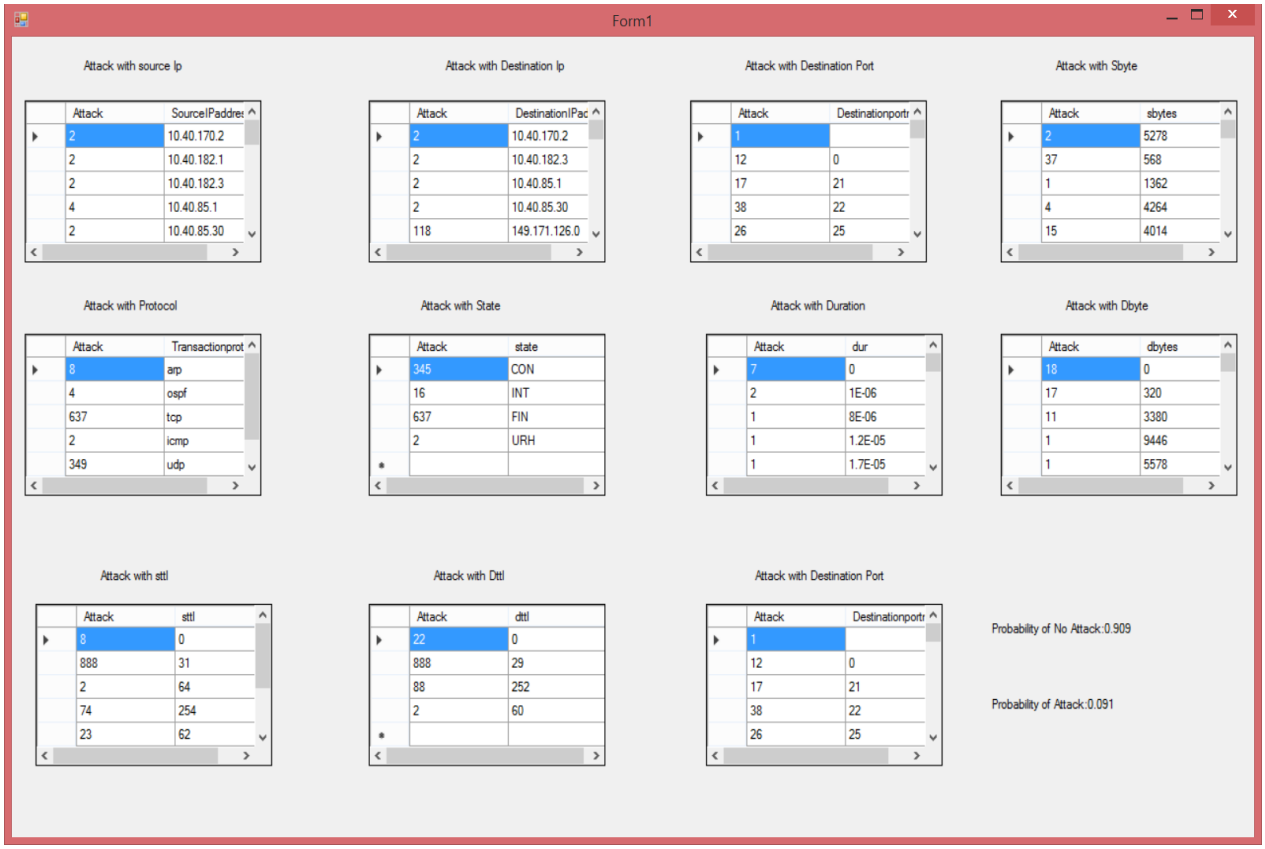

Fig 5.2 Results using Hybrid algorithm

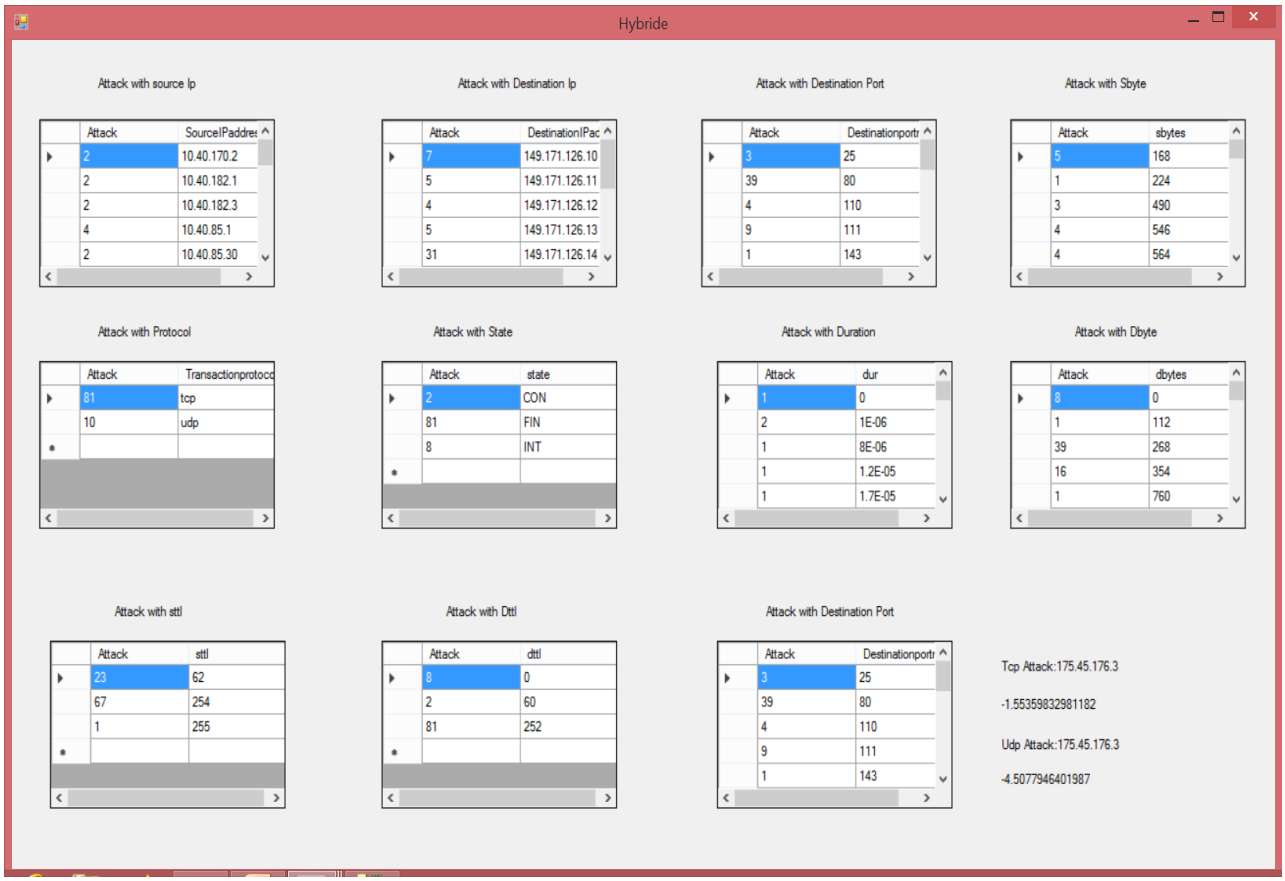

Fig 5.3 Results using naive Bayes

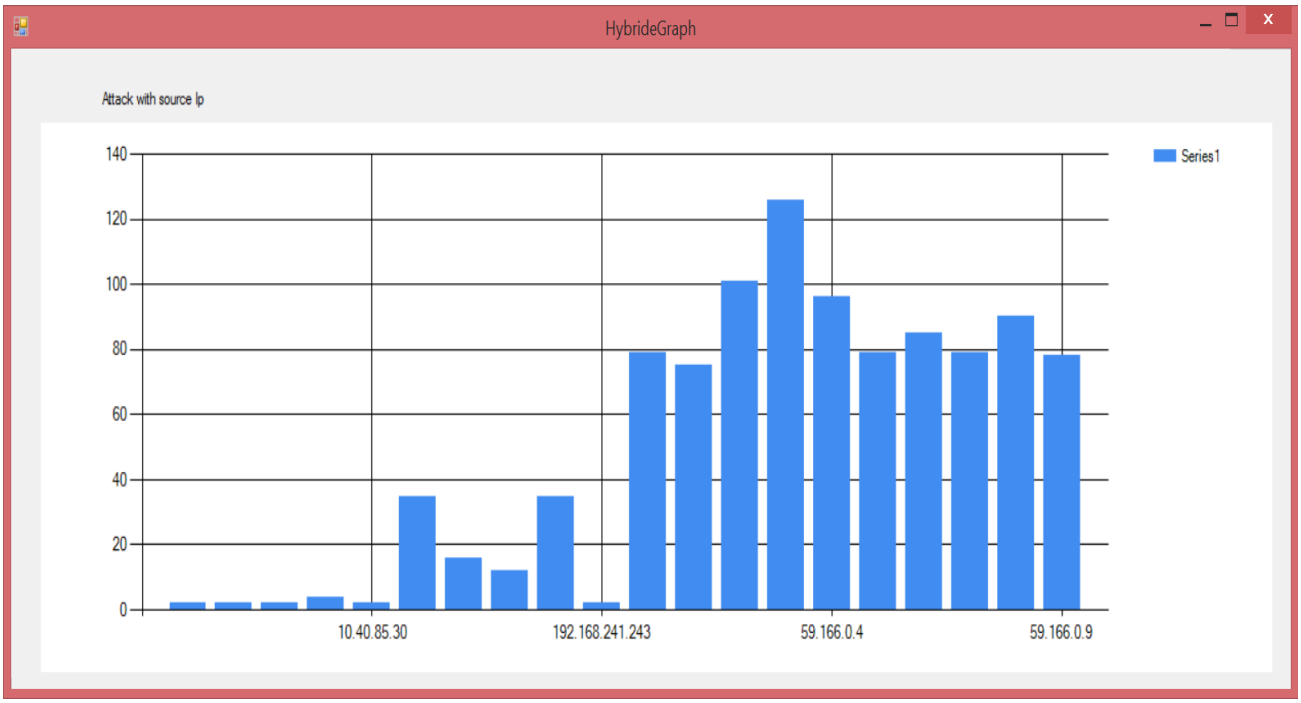

Fig 5.4 Graphical representation of Hybrid algorithm

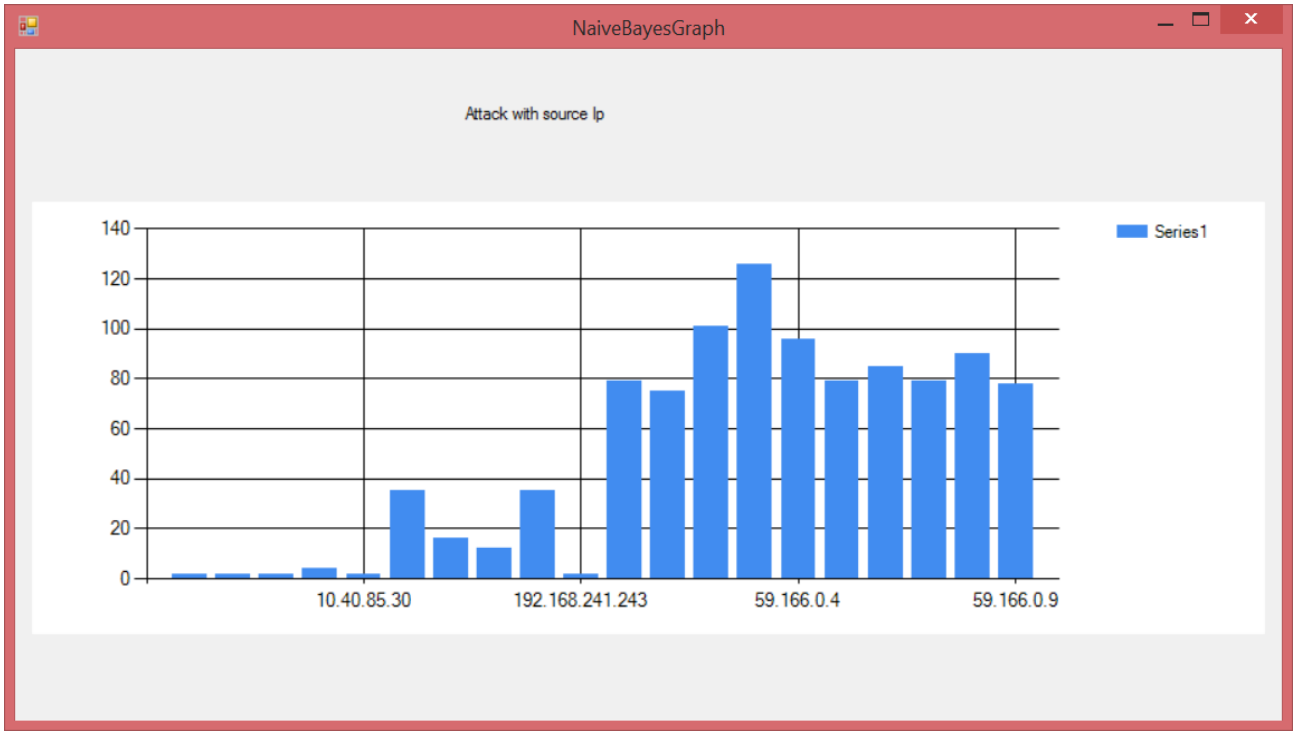

Fig 5.5 Graphical representation of naive bayes

5.2 Conclusion

Security in cloud computing is the need of the hour, our solution provides a safe and reliable method for businesses and organizations to store their data and share it via the cloud, preventing leakage of sensitive information and detecting intruders who try to hack this information. our solution provides a multi layered solution, unlike any other currently available for the cloud. it also effectively shows us which files of ours are the most targeted and the IP addresses and source port of the intruder.

Further building an effective intrusion detection system using machine learning methods has received much attention for network security. Data set always contain a huge number of features where most of it are redundant or irrelevant. Employing feature reduction method is an essential to reduce the computational cost and increase the classifier performance. Feature selection and feature extraction are having advantages and disadvantages, which make it hard to choose a single method to implement. It's recommended to use feature ex-traction followed by feature selection as a hybrid approach to increase the accuracy of intrusion detection.

The IDS (Intrusion Detection System) is for the oversee procedure can be additionally formed into a different, robotized framework with the accompanying upgrades:

•Help document can be incorporated. The framework, starting at now, does not bolster any help office for the clients of the framework. A help menu can be given a unique capacity key and help order in the fundamental page itself. Help can be either presented as a different window, a reference to a printed manual or as maybe a couple line proposal created in a settled screen area.

•The framework can utilize wrote orders, as they were at one time the most well-known method of correspondence with the framework. The wrote order can be given through control arrangement or capacity keys or wrote word.

•A preparing module can be incorporated into the framework. This module can be utilized to prepare the clients of the framework about the frameworks utilization.

5.3 Future Work

In Machine Learning (ML), the instance which has a relating mark is known as regulated learning, and the occurrence which has no name is known as unsupervised learning. In a circumstance where a portion of the cases are named and some are not, this is commonly known as Semi-regulated learning .

A. Administered learning

The procedure of direct characterization is to assemble a model that can separate between no less than two classes in view of in terms of number highlights with insignificant blunders for the new inconspicuous before tests. For model assembling, the classifier has to be in a need of marked preparing dataset that has all the typical and assaults tests. There have been reported many advtanges for superior rate of identification on the fact that regulated learning provides furnished classifier with indepth knowledge than semi-directed strategies and unsupervised strategies as well. Be that as it may, managed learning endures a few issues as follows: (I) Unavailability of a given datasets in the preparation time that can cover every authentic angle. (ii) Accurate marks are not generally ensured. (iii) High false caution rate if the preparation dataset contains clamor.

B. Semi-administered learning

The semi-administered strategy is somewhat in the middle of regulated and unsupervised techniques. In the use of constant oddity interruption discovery, the semi-directed strategy is more useful since it requires just the named information of the typical class, yet such technique isn't holistically utilized on the basis that there are marks for accessibility for each conceivable inconsistency in the preparation time.

C. Unsupervised learning

Utilization of marked information does not include unsupervised learning, it utilizes factual models to parcel information into ordinary and inconsistencies with no earlier information, in light of two suppositions mentioned as below:

(I) It presumes that measure of information is ordinary and the peculiarities speak to a little measure of it.

(ii) Statistically, the ordinary and oddities information are not quite the same as each other.

Information REDUCTION TECHNIQUES USED IN IDS

Data reduction is a strategy that is known to offer a mechanism or device or a tool for examination which makes it conceivable to get a helpful data from an immense dataset so it can be utilized as a part of more investigation. The greater part of the data mining and machine learning techniques couldn't work well in IDS because of the tremendous size of system information, and that can cause a long computational time. One explanation behind that is the idea of gathered system information which contains a major number of extricated highlights that IDS should process. The quantity of factors in information is fundamental for a decent order as though the highlights are too high, that will make lost speculation, and in the event that it is too little, that additionally could debase characterization quality. Highlight decrease is a decent technique that causes individuals to comprehend the quality and significance of highlights and how they are identified with each other. Also, observational outcomes in

IDS demonstrate that when utilizing an element diminishment method, the classifier execution regarding precision and computational cost is moved forward. Strategies for highlight and dimensionality decrease, for example, grouping, include extraction, and choice, has been utilized as of late by numerous scientists in IDS as a pre processing advance to enhance the exactness and diminish the computational multifaceted nature.

References

[1] Almorsy, Mohamed, John Grundy, and Ingo Müller. "An analysis of the cloud computing security problem." arXiv preprint arXiv:1609.01107 (2016).

[2] Gonzales, Dan, et al. "Cloud-trust—A security assessment model for infrastructure as a service (IaaS) clouds." IEEE Transactions on Cloud Computing 5.3 (2017): 523-536.

[3] Müller, André, André Ludwig, and Bogdan Franczyk. "Data security in decentralized cloud systems–system comparison, requirements analysis and organizational levels." Journal of Cloud Computing 6.1 (2017): 15.

[4] Salman, Tara, et al. "Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments." Cyber Security and Cloud Computing (CSCloud), 2017 IEEE 4th International Conference on. IEEE, 2017.

[5] Feng, Wangyan, et al. "Wavelet transform and unsupervised machine learning to detect insider threat on cloud file-sharing." Intelligence and Security Informatics (ISI), 2017 IEEE International Conference on. IEEE, 2017.

[6] Tirodkar, Sagar, et al. "Improved 3-dimensional security in cloud computing." arXiv preprint arXiv:1404.1836 (2014).

[7] **Ahmed Patel, Qais Qassim, Christopher Wills**. A survey of intrusion detection and prevention systems, Information Management & Computer Security Journal (2010).

[8] **Oludele Awodele, Sunday Idowu, Omotola Anjorin, and Vincent J. Joshua**, A Multi-Layered Approach to the Design of Intelligent Intrusion Detection and Prevention System (IIDPS), Babcock University, (Volume 6, 2013).

[9] Host Intrusion Prevention Systems and Beyond, **SANS Institute (2012).**

[10] Intrusion Detection and Prevention In-sourced or Out-sourced, **SANS Institute (2012).**

[11] **Mario Guimaraes, Meg Murray**. Overview of Intrusion Detection and Intrusion Prevention, Information security curriculum development Conference by ACM (2011).

[12] **Muhammad Awais Shibli, Sead Muftic.** Intrusion Detection and Prevention System using Secure Mobile Agents, IEEE International Conference on Security & Cryptography (2013).

[13] **David Wagner, Paolo Soto**. Mimicry Attacks on Host Based Intrusion Detection Systems, 9th ACM Conference on Computer and Communications Security (2014).

[14] **C. Guo, Y. Ping, N. Liu, and S.-S. Luo**, "A two-level hybrid approach for intrusion detection," Neurocomputing, 2016.

[15] **C. Guo, Y.-J. Zhou, Y. Ping, S.-S. Luo, Y.-P. Lai, and Z.-K. Zhang,** "Efficient intrusion detection using representative instances," Computers & Security, vol. 39, pp. 255–267,

2013.

[16] Almorsy, Mohamed, John Grundy, and Ingo Müller. "An analysis of the cloud computing security problem." arXiv preprint arXiv:1609.01107 (2016).

[17] Gonzales, Dan, et al. "Cloud-trust—A security assessment model for infrastructure as a service (IaaS) clouds." IEEE Transactions on Cloud Computing 5.3 (2017): 523-536.

[18] Müller, André, André Ludwig, and Bogdan Franczyk. "Data security in decentralized cloud systems–system comparison, requirements analysis and organizational levels." Journal of Cloud Computing 6.1 (2017): 15.

[19] Salvi, Sanket, et al. "An encryption, compression and key (ECK) management based data security framework for infrastructure as a service in Cloud." *Advance Computing Conference (IACC), 2015 IEEE International*. IEEE, 2015.

[20] Ertam, Fatih, and Orhan Yaman. "Intrusion detection in computer networks via machine learning algorithms." *Artificial Intelligence and Data Processing Symposium (IDAP), 2017 International*. IEEE, 2017.

[21] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153-1176.

[22]Almseidin, Mohammad, et al. "Evaluation of machine learning algorithms for intrusion detection system." *Intelligent Systems and Informatics (SISY), 2017 IEEE 15th International Symposium on*. IEEE, 2017.

Saumitra Vikram Singh

To work in a challenging environment, in a reputed concern where I can further develop my skills, enhance my knowledge base, and a genuine enthusiasm would allow me to progress. Exceedingly ambitious, hungry for success and loyal to a collective goal.

saumitra.singh96@gmail.com

+91-9654432908

A-105, Arunima Palace, Sector-4, Vasundhara, Ghaziabad, Uttar Pradesh, India

02 September, 1996

EDUCATION

B.Tech (Computer Science and Engg.)
Jaypee Institute of Information Technology, Sec-62, Noida
2014 – 2018 *6.0 CGPA (7 Sem)*

XIIth Standard(Senior Secondary) CBSE Board
Amity International School, Vasundhara
2014 *76.75%*

Xth Standard(Secondary) CBSE
Amity International School, Vasundhara
2012 *8.6 CGPA*

PERSONAL PROJECTS

Corporate Cloud Security Solution with Intrusion Detection (07/2017 – Present)
- A multi-layer approach to providing a complete security solution for confidential files and documents using encryption and user identification. Added failty of Intrusion detection and identification using concepts of machine learning.

Open Source Project management tool (05/2017 – 07/2017)
- A project management tool for scheduling and assigning various tasks to be carried for the completion of a project.

Competitive analysis of products based on customer reviews (02/2017 – 05/2017)
- Uses the concepts of sentiment analysis and opinion analysis coupled with web and data mining to evaluate different Internet Service Providers (ISPs) based on customer reviews extracted from the internet.

WORK EXPERIENCE

Summer Internship
Bharat Heavy Electricals Ltd, New Delhi
05/2017 – 07/2017
Tasks
- Project Management, IT Department

SKILLS

C/C++ | C# | HTML | PHP | MySQL | Unity 3D | Adobe Photoshop | English Proficiency | Excel | MS Office

ACHIEVEMENTS/ADDITIONAL

Member of core organising committee of Annual Cultural Technical fest of College (2016 – 2017)

Head, School Organization Committee

Star Performer in Instrumental Music (2010 – 2012)

Captain, School Athletics Team (2014)

Member, School BasketBall, Table Tennis, BasketBall Teams (2011 – 2014)

CERTIFICATES

Full Android Developer Course

Introduction to Data analysis
Microsoft

Data Science: using R
HarvardX

LANGUAGES

English ●●●●○

Hindi ●●●●○

INTERESTS

Public Speaking | Debating | New people and Experiences | Music | Sports | Social Work | Administation | Movies | Analytics

Gaurav Sood

+91-9818614622 14103185gaurav@gmail.com A-74 Shivalik, New Delhi, Delhi, 110017

Career objective

To work in a challenging environment, in a reputed concern where I can further develop my skills, enhance my knowledge base, and a genuine enthusiasm would allow me to progress.

Academic qualification

Educational qualification	Year of passing	College/School/Board	Percentage/GPA
Bachelor of Technology (Computer Science)	2018	Jaypee Institute of Information Technology	6.7 (up to 7th semester)
Class XII	2014	Laxman Public School (CBSE)	92.4%
Class X	2012	Laxman Public School (CBSE)	8.0

Area of expertise/interest

- Computer & Wireless Networks
- Data Mining
- Indian Financial System

Projects and training

- '**Big Data and Hadoop**' Training at WebTek Labs. (June – July 2017)
- **Minor Project** on *Competitive Analysis based on Customer Reviews* (January – June 2017)
  -Extraction of customer reviews from mouthshut.com and analysing them to predict which product is superior
- **Major Project** on *File Access Authentication using Intrusion Detection* (August - December 2017)
  -Proposing a solution for organisations to protect their confidential data using a 4 level security solution and intrusion detection.

Scholastic achievements

- **Olympiad rank 19** in 'International Master Mathematics Olympiad' (2010)
- **First Position** in '2nd NHS Inter School Mathematica Talent Hunt' (2009)

Extra-Curricular Activities

- **Social Endeavours**
  - Volunteer at Rotary International (Rotaract Club Of Delhi South, District 3011) (May 2016 – Present)
  - Volunteer at workshop on 'Finance for non finance professionals' (5th September 2017)
  - Volunteer at National Literacy Mission (2009)
- **Sports**
  - **Winner** [U-17 Category] Men's open at Yamuna Wealth Games (2011)
  - **Runner-up** [U-19 Category] Men's open at Yamuna Wealth Games (2012)
  - **Winner** [U-19 Category] Tennis Zonal level, Delhi (2014)
  - **Runner-up** [U-19 Category] Tennis Zonal level, Delhi (2013)
  - **Runner-up** [U-16 Category] Tennis Zonal level, Delhi (2009)

Position of responsibility

- **Director of Club Services** (Rotaract Club Of Delhi South, District 3011) (May 2017 – Present)
- Head of Organising Committee, Inter College debate - Rotaract Club Delhi South (20-21 January 2018)
- Organising Committee, Inter College Debate - Rotaract Club Delhi South (21-22 January 2017)
- **Tennis Team Captain** at LPS (2011-2013)

Interest and Hobbies

- Lawn Tennis
- Graphic Designing

Learning Resource Centre
JIIT NOIDA
Turnitin Report

Dated. 10/5/18

Dear Sir,

Kindly allow me to avail Turnitin software facility.

Name of the Student/Scholar: GAURAV SOOD

Enrolment: 141031 85 Class Ph D/ M Tech /MBA/ B Tech

Phone No. 9818614622 E-Mail ID. gauravsood95@gmail.com

Name of the Supervisor: Dr. Amanpreet Kaur

Title of the Project Report/Thesis/Dissertation/Paper:

File Access Authentication using Intrusion Detection.

(Student's Signature)

For Accounts Department:

Amount deposited Rs. 150 Dated. 10/5/18 Payment Slip No. UNIV1845.000146

(Payment slip is enclosed)

Account Officer

For LRC USE

Thesis/dissertation/Project copy received on 10/05/2018

Report of Turnitin delivered on 11/05/2018

Similarity Index in % 26%

Librarian

Project Coordinator/Research Guide

H. O. D

Dean A&R

File access authentication using intrusion detection

ORIGINALITY REPORT

26%	14%	15%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Abdulla Amin Aburomman, Mamun Bin Ibne Reaz. "Survey of learning methods in intrusion detection systems", 2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES), 2016 Publication	2%
2	documents.mx Internet Source	2%
3	www.ipcsit.com Internet Source	1%
4	arxiv.org Internet Source	1%
5	de.slideshare.net Internet Source	1%
6	toc.proceedings.com Internet Source	1%
7	Submitted to Management Development Institute Of Singapore	1%