SECURE BIG DATA COMPUTING

Synopsis Report submitted in partial fulfillment of the requirement

for the degree of

Master of Technology

in

Computer Science & Engineering

under the Supervision of

Dr. Pardeep Kumar

By

Richa Verma (152202)

Jaypee University of Information Technology

Waknaghat, Solan – 173234, Himachal Pradesh

May, 2017

Certificate

This is to certify that synopsis report entitled “SECURE BIG DATA COMPUTING”, submitted by Richa Verma in partial fulfillment for the award of degree of Master of Technology in Computer Science & Engineering to Jaypee University of Information Technology, Waknaghat; Solan has been made under my supervision. This synopsis has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Date: **Dr. Pardeep Kumar**

Assistant Professor (Senior Grade)

Acknowledgement

I wish to express my heart felt thanks and a sense of gratitude to my" **Dr. Pardeep Kumar "**Assistant Professor, Computer Science & Engineering Department, for his valuable guidance and constant inspiration in preparing this report. My frequent interactions with him in all aspects of the report writing have been a great learning experience for me. I shall always cherish his support and encouragement.

Last but not the least, I heartily appreciate all those people who have helped me directly or indirectly in making these task a success. In this context, I would like to thank all the other staff members, both teaching and non-teaching. My friends were always very helpful to me. They were always there for me whenever I needed their support.

Date: Signature:

Richa Verma

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

Abstract

Amid the most recent decades, information security has turned into a noteworthy subject. Encoding of data has as of late been generally explored and created in light of the fact that there is an interest for a more grounded encryption. Encryption has been done to secure transmitted information from unapproved individual, in encryption the transmitted information is broken into mystery figure and perusing data or changing over the genuine data into various data. Encryption procedure will be connected in sound record. As the utilization of web expands step by step in this way, it end up noticeably important to secure the transmitted data from the unapproved individual. Cryptography assumes an imperative part inside the field of system security. There are numerous encryption strategies that have been available in the market to secure the data. Amid this exploration we'll have a framework of encoding systems. This exploration principally concentrates on encryption methods for sound and for content information. In the examination cross breed encryption system alongside neural system has been utilized. Many-sided quality discoverer was utilized to discover the rightness of the proposed work and parameters like exactness, review, and measure and precision will be figured.

CHAPTER 1: **INTRODUCTION**

1.1 BIG DATA

IBM has expressed that, gadgets which are handheld, machine-to-machine correspondences, and on the web/portable informal organizations they make 2.5 quintillion bytes of information every day. It is difficult to catch, store, oversee, share, examine, and imagine with existing information and handling apparatuses as it cause tremendous issue. The idea of enormous information has been advanced hence [1].

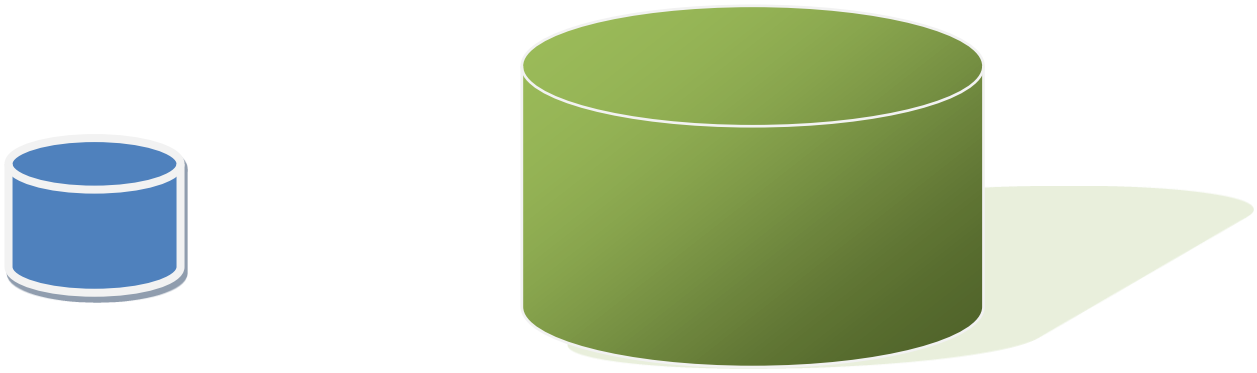

Fig.1.1: Amount of data volume

We have to make our capability for data generation more powerful and enormous from creation of the information technology to advanced technology impact. As another illustration, the primary presidential verbal confrontation between President Barack Obama and Governor Mitt Romney happened on 4 October 2012, inside 2 hours more than 10 million tweets has been set off, the particular minutes that created the most exchanges among every one of these tweets really uncovered the interests of people in general, for example, the talks about medicare and vouchers. Expanding interest has been happened as of late in Big Data. Be that as it may, the term Big Data stays dubious [2]. In Wikipedia, Big Data is an extensive term for any social occasion of educational accumulations so considerable and complex that it winds up clearly difficult to handle using standard data gets ready applications. A broadly perceived definition has a place with IDC: "colossal data developments depict another time of advances and structures, expected to fiscally remove a motivator from incomprehensible volumes of a wide grouping of data, by enabling the fast catch, disclosure, and also examination" exploring and using the

extraordinary estimation of Big Data must grow threats of security and insurance. For instance, "Our shopping inclinations has been monitors by Amazon and habit of browsing is learnt by Google, while what's goanna on our minds is known by Twitter All the information catches by Facebook too, along with our social relationships. Whom we talk to, and who is nearby all known by mobile operators. Who analyze Big Data for them it is an important insight all points seems to be pointed out further like gathering, storing, and reusing our personal data. Big Data seems to be in danger when security and privacy threatened the Internet age.

On the off chance that the age of the Internet risk to security and protection, then the period of enormous information will imperil them. Before moving ahead for what big data is, a moment is required to look at the below diagram by Hewlett-Packard [3].

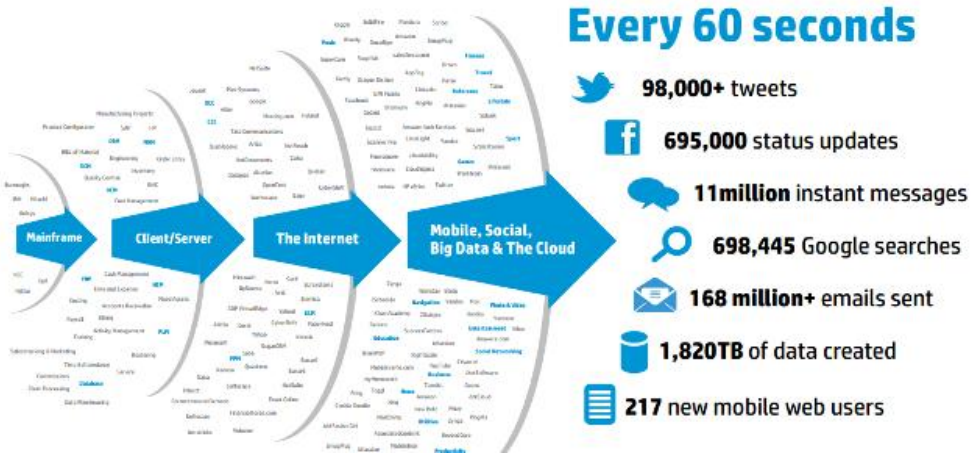

Fig. 1.2: Big data of social sites

Source: http://www.bigdataplanet.info/p/what-is-big-data.html

The above figure depicted the quantity of the data being generated of each minute, in each second, of each hour, of every month, of every day, of every year. So, it being true that more than 90% of data has been generated in the last two years itself. In the "3Vs" model, Species speaks to different sorts of information including organized, semi-organized and unstructured information; information volume speaks to the measure of information; speed implies all procedures extensive information must be quick and in a convenient way, with a specific end goal to expand the estimation of enormous information. The data processing has large number of these feature data, using various types of data has never been used before with the unstructured data attributes, with the distinction between data mining and large data [4].

IDC has characterized huge information in 2011, as 4 Vs. that is Velocity, Variety, Volume with Value, in which the ramifications of Velocity, Variety and Volume is indistinguishable with the 3Vs model correspondingly and Value characterizes enormous information with extraordinary social esteem

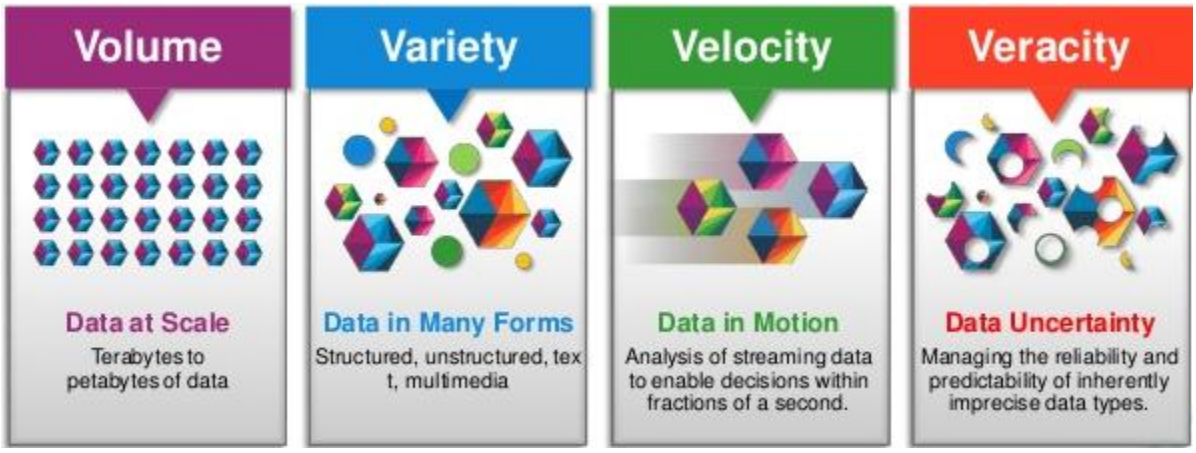

Fig. 1.3: Varieties of big data

Source: https://www.slideshare.net/ibmsverige/building-confidence-in-big-data

4Vs model is generally perceived, in light of the fact that it demonstrates that the most vital question that how to discover awesome, and rapidly create different sorts of information from a substantial arrangement of information qualities. Variability: with Velocity with the data stream may be highly inconsistent with periodic peaks [5].

Many organizations around the world have taken a lot of data operations, because they realize enormous social and economic value of the big data exploration and use.

NASA is produced or will create a lot of data, the Large Hadron Collider experiment square kilometer array telescope and the Sloan Digital Sky Survey and the large astronomical telescopes such as scientific research. Many organizations, for example, Amazon, eBay, Walmart, Facebook, FICO, have effectively settled various expansive scale information distribution centers to store their business information.

"It gauges that the volume of business information around the world, over all organizations, pairs at regular intervals" There are many advantages of enormous information illustrations. A striking case is Google, it found that the spread of this season's flu virus related hunt questions time and place with generally extraordinary frequencies and access to opportune data to foresee and find

influenza flare-up by totaling countless inquiries amid the 2009 H1N1 emergency. Another illustration is Farecast, a US-based innovation new business, aircraft tickets acquired by Microsoft motor guaging framework to foresee slants in carrier ticket costs and increment or reduction has been joined into a Bing look in 2008. "By 2012 the framework was making the right call 75 percent of the time and sparing voyagers, by and large, $ 50 for every ticket" . These two cases delineate the considerable logical and social information and enormous information can be a wellspring of financial esteem. The most celebrated case of the tremendous capability of enormous information might be the Internet. Huge Data applications in the field of Internet web based business are web based publicizing, online news, web search tools, travel booking, texting, long range informal communication, blog, microblogging, online video, online music and web based amusements. Web monsters, for example, Google, Facebook, Microsoft, Alibaba, Tencent and other enormous financial potential, on the grounds that these organizations have amassed a formerly inconceivable number of client information. For instance, 'Facebook has more than 900 million clients transfer more than 250 million photographs, and tap the' like 'catch more than 2.5 billion times each day.' Now, these information as of now has a substantial client of Internet organizations are looking for better approaches to use their client information. The advantages of a developing number of associations gather the data about online clients by building enormous information stage to share vast information. [6].

Another illustration is Flicker, an open photograph sharing site, got an everyday normal of 1.8 million photographs from February 2012 to March. Accepting that the extent of every photograph is 2 megabytes (MB), each photo needs 3.6 TB of capacity in a solitary day. Truth be told, as a familiar axiom puts it: 'a photo worth a thousand words', blazing billions of pictures that we find treasure trove of human culture, get-togethers, open undertakings, fiasco, and so forth., just when we can utilize a substantial number of The information.

These illustrations have demonstrated incredible ascent of Big Data application improvement in information accumulation, past the normally utilized programming apparatuses inside 'middle of the road time' catch, administration and preparing capacities. The major test is to investigate enormous information applications for a lot of information and concentrate valuable data or learning for future activity [7]

1.2 BIG DATA SECURITY

Big data security is divided into various levels which includes; communications, processing, authentication, and storage level. Present technology was not built for big data security. Data is stored in plain text, wherein important information can be easily stolen by hackers. Logging to critical data is not logged in which means any abused of data cannot be recognized. This technology is used by many businesses to store and examine their own data and their customer's data which makes privacy and security very essential in gaining the confidence of customers. Hence, it is require for investing, studying, and understanding the challenges and providing better solutions to secure big data [8].

Big Data security is usually achieved using big data to increase the security of distributed systems, reliability and security solutions. Enormous information security insurance concentrated on huge information from unapproved utilize and pointless deduction. As we as a whole know, on the premise of a strong and precise security arrangements in view of enormous information is a significant wellspring of data. In any case, huge information regularly contains delicate data that should be ensured to forestall unapproved get to and discharge. Clearly, on the off chance that we don't extricate an incentive from huge information, then there are no security and protection challenges. Thusly, the guideline of enormous information security and protection must be adjusted with extra social estimation of huge information [9].

"Through late exposure, the National Security Administration routinely gathers and investigates monstrous measures of individual information gotten from heterogeneous information sources, for example, broadcast communications, the Internet, and the client databases of substantial organizations, including Microsoft, Yahoo, Google, Facebook, PalTalk, YouTube, Skype, AOL, and Apple". Numerous certainties demonstrate that if huge information is not accurately deal with the huge information security, than such a large amount of information will harm the client's protection, it must be scrambled, recording, cell recognition innovation. In numerous associations, conveying enormous information for extortion location is alluring and valuable [10].

Safety structured data can be considered through the use of existing security theology or standard SQL queries. In this manner, the intrigue is to give security to unstructured information, including content, XML, picture, video, sound, et cetera. By taking into account existing safety standards or algorithms, a large unstructured data secure is developed. The method uses data

analysis techniques described for analysing unstructured data, to build a database of data nodes, which contain different types of text, XML, email, images, video, audio and other data, the next step is to build the security to provide security suites protection.

In order to ensure that a malicious third-party data security, encryption is a common choice, Fully homomorphic encryption key management can saved all the trouble, because even after performing any type of operation, encrypted data set may still be decrypted using the key for encryption.

1.3 NEED OF SECURITY IN BIG DATA

For marketing and research, many companies use large amounts of data, but may not have the basic assets, particularly from a security point of view. If big data security breaches will lead to more serious consequences, than the current legal and reputational losses.

In this new era, many companies utilize pet-byte data for their companies, enterprises and customers to store and analyse. Therefore, the information classification becomes more important. In order to secure, large data need to be encrypted and recorded by honey detection technique. In numerous associations, sending enormous information for extortion location is exceptionally alluring and valuable. Propelled danger identification and anticipation challenges and malignant gate crashers must be utilized to understand enormous information examination style. These systems encourage the utilization of more advanced example examination, and investigation of various information sources to identify early dangers. [11].

Not only security, data privacy also poses a challenge to existing industry and federal organizations. With the expanding utilization of a lot of information in the business, many organizations are protecting security issues. Information protection is an obligation, so the organization must be founded on security grounds of safeguard. Be that as it may, with various security, protection ought to be considered as resources, consequently turning into an offering point for clients and different partners. There ought to ping Heng between information protection and national security.

With the growing popularity of possible computing environment, safety by adapting this technology is introduced. While cloud computing offers many benefits, but it is vulnerable to attack. Attackers are always looking for loopholes to attack the cloud computing environment. Since these cloud computing deployments, traditional security mechanisms used are

reconsidered and require visualization, control and inspection of the network link and port capacity to ensure security. Therefore, there is need to invest in understanding the challenges and vulnerable components of cloud computing facing and propose a less vulnerable platforms and infrastructure [12].

1.4 SECURITY CHALLENGES OF BIG DATA ANALYTIC

This segment characterizes diverse enormous information for security and protection challenges. The main data may be large unlike the conventional method, four types, namely: volume, variety, and speed values.

The security challenge are characterized by the speed, and a large amount of diversity to improve the data, such as large cloud frame mode and the data source, the data acquisition cascade nature. These can quickly improve and change supplier management threaten and attacks. Comparison of various security challenges are given below:

Table 1.1 Comparative study of various security challenges [12]

Security Challenges	**Explanation**
Transaction log file and Protected database storage	With the increase of database volume and availability, the scalability needed to automatically run out of big data management. Auto exhaust solutions do not meet the actual storage location of the database, so the need to protect the database storage is there.
Secure computations in distributed frameworks	In order to handle very large data, parallel computing is used for computational and physical storage. Such as the Map Reduce framework. Protecting the data and protecting the mappers in the presence of untrusted mappers are two major anti-attack measures.
compliance monitoring and Real-time	In this method, the numbers of alerts are

security	generated by the privacy device. These alerts guide have false positives, mainly by being abused or simply 'clicked' because humans cannot cope with the amount. It is used to give real-time problem detection, such as scalable privacy analysis.
secure communication and Cryptographically enforced access control	This approach is used to provide fairness, authentication, and negotiation in distributed entities.
Secure information	Information security has become a big data problem. From a security perspective, how to try large data becomes a daunting task. Such as financial services, business websites, health departments, social networking sites, networks, and anomaly detection. So data security is one of the big data problems.

Multimedia data is a set of the following media: content, audio, motion pictures and movies. Multimedia protection involves the prevention of such material. By using multimedia information, the multimedia data can be transferred from a user having a convertible device, such as a variable phone. Multimedia data escapes the issue of material rights and privacy, so the sequential defense of multimedia becomes critical in multimedia-enabled devices [13]. Because a lot of multimedia information has a wide range of sizes, we need a qualified encryption technology to defend multimedia data while meeting the same needs.

1.5 MULTIMEDIA DATA SECURITY

Because of the present advancement in PC arrange innovation, giving out of computerized interactive media satisfied through the web is enormous. However, the enlarged number of

advanced reports, minimized circle preparing instruments, and the universal convenience of Internet get to has made an extremely suitable medium for selective rights misrepresentation and rebellious conveyance of sight and sound substance. A noteworthy condition now is to secure the researcher belonging of interactive media content in smaller circle systems. There are figure of information sorts that can be describe as interactive media information sorts. These are typically the nuts and bolts for the building squares of general media conditions, stage, or coordinate instruments. The basic sort can be depicted as content, pictures, sound, video and Graphic items. Interactive media discovers its motivation in different zones numbering, yet not restricted to, ads, craftsmanship, training, diversion, building, pharmaceutical, arithmetic, business, logical research and spatial fleeting applications. Mostly in Medicine, specialists can get qualified by taking a gander at a virtual surgery or they can repeat how the human body is valuable by maladies reach out by infections and microorganisms and afterward create method to avert it [14]. So with the extension of data message and PC innovation, there has been the advancement of Digital doctor's facility, Telemedicine in system by index of computerized restorative picture.

In this exploration, we prospect the watchful encryption progress for ensuring mixed media information. Computational workload vital for this encryption is less [15]. Encryption all in all way to change the message into code or turned shape, so that anyone who does not have the "key" to unravel the code can't see it. This is as often as possible done by utilizing a 'figure'. A figure is a kind of calculation utilized as a part of encryption that uses certain portray strategy to stir up the information. The figure must be "deciphered" with a 'key'. A key is the genuine portray technique' that was utilized to foul up the information, and thus the key can likewise translate the information When the information is unscramble by the utilization of a key, that is what is known as 'decoding'. It is the clashing of encryption and the depict strategy' of scramble is basically functional in discredit, to interpret it. Thus, the in a state and messy content winds up noticeably decipherable at the end of the day. Without encryption and clarification, there would be no "security" in the system.

1.6 MULTIMEDIA DATA

There is variety of information sorts that could be defined as an interactive media information sorts. It is commonly the components for the building squares of mineral summed up mixed

media situations, stages, or coordinating apparatuses [16]. The essential sorts can be depicted as takes after:

- **Text:** The structure in which the content can be put away can shift enormously. Notwithstanding ASCII based documents, content is commonly put away in processor records, spreadsheets, databases and annotations on more broad media objects. With accessibility and expansion of GUIs, content textual styles the employment of putting away content is getting to be complex permitting enhancements (shading, shades etc).
- **Images:** An awesome difference is there in the quality and size of capacity for motionless pictures. The digitalized pictures be succession of pixels to speaks to an area in the client's graphical showcase. The space overhead for still pictures changes on the premise of determination, size, many-sided quality, and pressure plan used to store picture. The prominent picture arrangements are jpg, png, bmp, and tiff.
- **Audio:** An undeniably well-known information sort being coordinated in the greater part of uses is Audio. It's truly space concentrated. Single moment of sound could take up to 2-3 Mbs for space. A few strategies be utilized for pack it in appropriate arrangement.
- **Video:** The most space devouring sight with the sound information sort is digitalized feature. The digitalized features are put away as arrangement of casings. Contingent on its determination and size a solitary casing can expend upto 1 MB. Additionally to have practical feature playback, the transmission, pressure, and decompression of digitalized oblige consistent exchange rate [17].

1.7 SECURITY THREATS IN MULTIMEDIA

A. Inside assaults

There is plausibility used for phishing and taking of media substance in the representative of administration supplier itself.

B. Legal and theft troubles

The more serious trouble on the account is to place the media content on the web outside the server that is outside the country. Similarly, the transfer of media material rights to different stages and the dissemination of media content for certain extent or to the maximum is also limited

C. Migration

The client may believe to shift all the media substance for some another spot taking into account his adjustment in prerequisites. In any case, now the client does not have the opportunity of doing that [18].

D. Challenges over gauges

At present numerous merchants (individual who offers administrations) creating and propelling their own private cloud situations in light they could call their own conditions and security highlights which prompts issues in interoperability soon.

E. QOS

Clients going for questionable systems without their insight to offer the media substance despite the fact that there are accessibility of all the more encouraging gushing innovation and expanded broadband pace.

1.8 ADVANTAGES OF MULTIMEDIA DATA SECURITY

The media equipment provides a critical amount of compensation for the provider and the user from one side to the other to expand the completion time, organize the good data storage capacity, and less computation cost[19]. It creates a staggering crash in multimedia content distribution, such as editing, storing, encrypting and decrypting, games, streaming, and so on. A number of more recompense is defined below:

- **Cost**

Media registering gives practical military administrations to its specialist organizations by proficiently reusing sight and sound substance, for example, sound, video, and pictures, with a typical foundation, server, advancement, virtualization, portability, and custom preparing. There is no prerequisite to really obtain correspondences or saves in the nearby framework, in this way decreasing expenses [20].

- **Upgradable**

The media is always associated with the service provider and therefore can be upgraded and maintained without any human intervention. Software and security will be updated.

- **Compatibility**

The media allows the media to meet anywhere through any intelligent mechanism [21].

- **Storage**

Media information has many bases for store the media content utilizing the wage. Additionally it is more shielded since the store media satisfied will be copy without manual interruption.

1.9 DISADVANTAGES OF MULTIMEDIA

1. Costly
2. Not generally simple to arrange
3. Requires Special Hardware
4. Not just Compatible [21].

1.10 ENCRYPTION

One of the major challenges of sharing resources in a network data communication is its security. This premise is based on the following facts: once there is a connection established between computers they can share the resources, so security of data become critical.

Today's the use of exchanging digital images becoming very frequent due to the increasing growth in the network technology. Protection of multimedia data, sensitive information (such as credit card, social security number and bank transactions) becomes very important. You can use a lot of encryption to protect confidential data from unauthorized access. Therefore, in order to provide data security, a number of cryptographic techniques such as symmetric and asymmetric techniques are used. A variety of non-symmetrical cryptographic methods, like Rivest Shamir and Adleman (RSA), Diffie-Hellman and Digital Signature Algorithm (DSA) are present. In fact, we use encryption technology every day, most of us cannot appreciate the 'how'/'why' we know the security of data is important, if the encryption help the users, and then the user should on the board. Almost each day, we interact with computing devices use some form of encryption technology. From a smart phone (usually can encrypt their data), tablet, desktop, laptop, and even your trusted Kindle, encryption everywhere.

Tendency in encryption products through the general trend of PC technology is moving inside the direction of miniaturization. Like, late in 1988, AT & T master encryption device has

produced weighing seventeen poundsWith the emergence of the PCMCIA (Personal Computer Memory Card Industry Association) technology, one or more manufacturers generally expect to use credit card-sized modems to publish encryption soon. This process is critical to create a digital signature. Consistent with the increase in electronic communications, business and personal transactions need to write their own signatures. At present, writing own signature requires written information. However, the use of electronic communications is so much that many commercial and personal transactions do not actually see each other's growth, nor the actual signature of both parties; more and more use of digital signatures to provide certification information. Expanded fame of the Internet has started a requirement for security. We trust that email clients' secrecy and sender validation are progressively should have been encoded. Some are utilizing PGP. Others started to utilize the Privacy Enhancement Mail (PEM), which is an Internet encryption instrument presented by the Defense Advanced Research Projects Agency-financed Reliable Information System as a business item organization. It utilizes DES encryption calculation, RSA calculation for sender validation and key administration. The protection upgrade message likewise offers help for non-denial; this enables the outsider to forward the message beneficiary to validate the originator (the message not just the repeater) to distinguish and check whether any progressions to the first content are made. While PEM is not yet accessible, numerous sellers ought to be utilized as a part of conjunction with or in conjunction with business email applications, and the European Community has embraced its PASSWORD PEM extend, which is a piece of an endeavor to build up a pilot arrange security framework. [21].

1.11 TEXT MESSAGES

It will send a short message toward a device like PDA, cellular phone, or pager. SMS messages for no more than a few hundred characters messages. The term is typically for applying messaging in order to transfer messages between mobile devices. Safety is a major issue in modern data communications. There's a lot of cyber-crime has emerged with the development of technology. Encryption came from cryptography. In this work, encryption is based on symmetric key encryption algorithm, where encryption and decryption keys are used.

1.12 TEXT ENCRYPTION

Data security implies shielding data and data frameworks from unapproved get to, utilize, exposure, devastation, adjustment, or annihilation. An imperative part of data security and hazard administration is the recognizable proof of the estimation of data and the ID of proper methodology and the required data assurance. Not all data is the same, so not all data needs a similar assurance. It is closely related to cryptography and cryptology discipline. The main goal of modern cryptography involves the following four things, namely confidentiality, integrity, authentication and non-repudiation. Hacking information is widely regarded as one of the security system for any potential attack. Therefore, we must protect the system from a text-based message encryption type of vulnerability [22].

Encryption is a new form of encryption for allowing the users to hide information from other people. The individuals who utilize refined encryption calculation known as the watchword for the standardized information like plaintext in an arrangement for apparently irregular characters like cipher text, the characters don't have an uncommon key is utilized to turn decode it is not a clear. The individuals who have the key can decode the information to see the plain content rather than the cipher text.

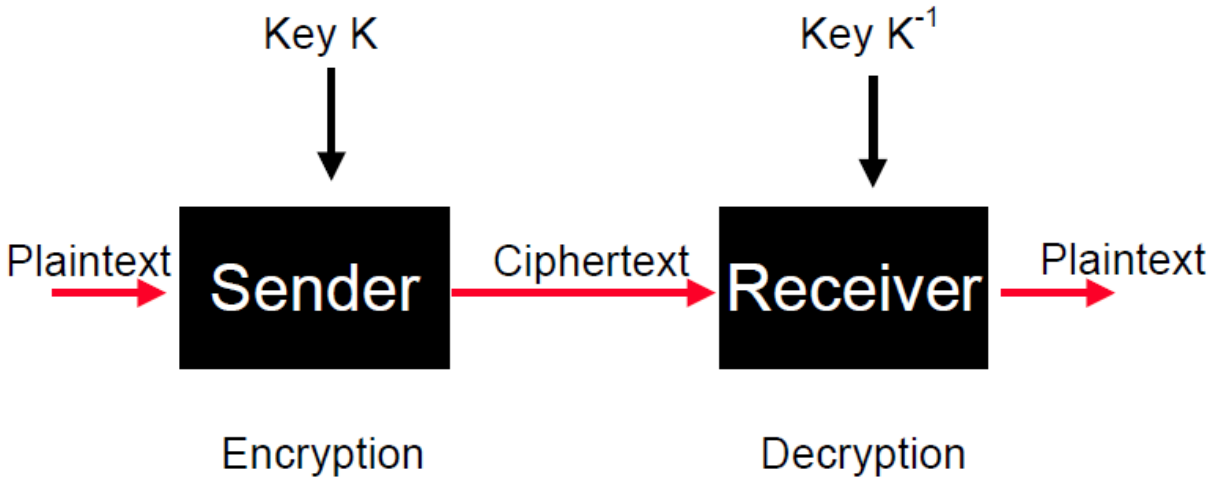

Fig. 1.4: Block diagram of encryption and decryption

Source: http://www.pling.org.uk/cs/cry.html

The whole process to encrypt and to decrypt data is known as cryptography. Firstly input which we want to send is given to the sender box which is plain text. In encryption scheme, an encryption algorithm used to encrypt the information or message is expected to be known

plaintext, the cipher text generated decrypted only when they are to read. For technical reasons, the encryption schemes typically use an encryption key of generated in the algorithm. It could decrypt message by not possessing the key principle, but for well-executed encryption scheme, resources with skills needed to calculate large. Authorized recipients can be easily provided by the originator to the recipient but not the authorized user key to decrypt the message.

1.13 TYPES OF KEYS

i. Symmetric key / private key [Update]

In the symmetric key scheme, the encryption and decryption keys are the same. The communicating party must have the same key to ensure communication.

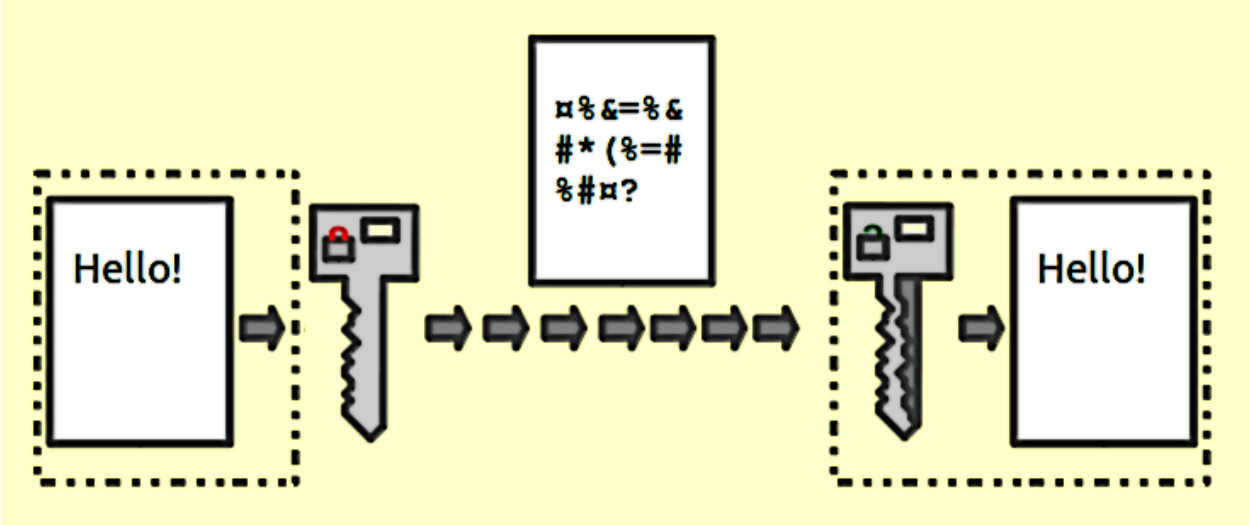

Fig. 1.5: Encryptions used for public key

Source: https://en.wikipedia.org/wiki/Encryption

From the figure it is clear that the message we want to send is “Hello”. Hello is converted into some other language so that the data remains secure and the end i.e. at the receiver we get the same data i.e. “Hello” by using some description algorithms.

ii. Public Key

In the public key encryption method, the encryption key is available to anyone to use and encrypt the message. However, only recipients can access the decryption key to read the message. First, public key cryptography is described in 1973 with the secret documents; in the past, all programs used symmetric key encryption (also known as private).

1.14 HISTORY OF ENCRYPTION

Greeks were the first to use cryptography social order with their enemies and the public eye hide sensitive data in writing. They use a very primitive encryption method relies on the usage of scytale as a tool for creating a permutation encryption that is answer key to decoding the encrypted message. Scytale be a cylinder, which is utilized for wrapping the parchment to deciphering the code. When both sides use the same thickness of the cylinder communication, write from left to right to read the information. When the parchment expanded, it looks like a thin parchment, long, seemingly random numbers with the letters. Therefore, as it might seem nonsense rolling competition, when we scroll scytale it looks like this :

Fig. 1.6: Message rolled on Scylate

Source: http://www.makeuseof.com/tag/encryption-care/

Other examples are:

- **Polybius Square:** another encrypted breakthrough lies in ancient Greece in the upper left corner of the letter 'A' and the lower right corner of the 'Z' 5 x 5 grid start (“'I' and 'J' share a square”). Numbers 1-5 emerge at the top with at the top of the left-most

horizontal and vertical line character. Code resides to give a numeral, and after that put it on the grid.Like, 'Ball' is 12,11,31,31.

- **Enigma:** This machine is an electrical and mechanical rotor cipher machine known as the WWII technology. It looks like an oversized typewriter that provides the operator towards input plaintext and encrypted message machine, and sends it to another unit. The receiver stores received after mechanical write encrypted random string of letters, and then setting the original mode from the sender's machine to crack the code.

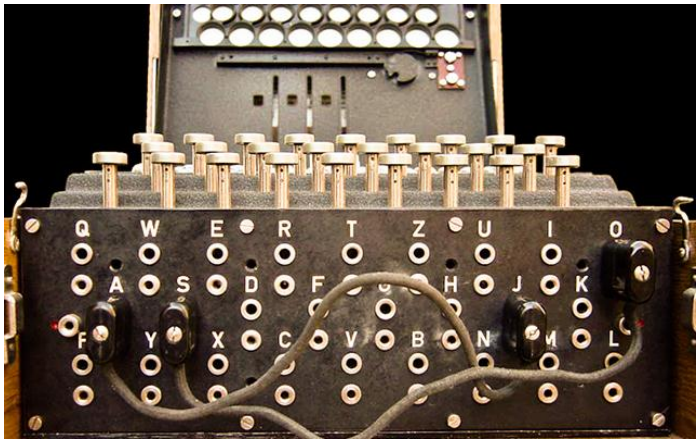

Fig.1.7: Crypanalysis of Enigma machine

Source: http://www.makeuseof.com/tag/encryption-care/

- **Data Encryption Standard:** Data Encryption Standard (DES) encryption of digital data is a modern first symmetric key algorithm. In the 1970s, IBM developed DES in 1977 in the US Federal Information Processing Standards and became the foundation of modern encryption technology.

1.15 MODERN ENCRYPTION TECHNOLOGY

Modern encryption technology uses more complex algorithms and larger key sizes in order to better hide encrypted data. The larger the key size, brute force attack must be run to find success more possible combinations to decrypt the ciphertext. With the continuous improvement of key size, use strong encryption of the length of time the attack began. For instance, in spite of the fact that the 56-bit key and 64 key is by all accounts really close, however the 64-bit key is really 256 times higher than the 56-bit key. Most current encryption utilizes at least 128-piece keys, various 256-piece key or bigger. To delineate this point, break the 128-piece key would require a solid assault to attempt 339,000,000,000,000,000,000 conceivable key mixes. In case you're interested, it really requires a million years to figure the right key to utilize savage constrain assault, which is the most effective supercomputers some time recently. To put it plainly, hypothetically outlandish that anybody would even endeavor to break a 128-piece or higher encryption innovation.

3DES

The encryption standard has been far since it’s initially utilization of DES in 1977. Truth be told, another DES innovation called Triple DES (3DES) is very normal; it depends on the advanced rendition of the first DES calculation. In spite of the fact that the first DES innovation is exceptionally constrained, there is just a single 56-bit key size, the present size of 168 3DES keys, making it essentially more troublesome and tedious.

AES

Propelled encryption gauges depend on the present US government standard symmetric secret word. AES has been embraced all inclusive as a successor to the obsolete DES standard in 1977, despite the fact that the declaration is quicker than the solid assault, however the intense AES innovation is as yet viewed as difficult to ascertain. Also, AES offers an assortment of equipment and stable execution, and gives rapid and low RAM necessities, settling on it the favored decision for generally applications. On the off chance that you are utilizing a mainstream device for FileVault Mac, encryption utilizes AES in one of numerous applications.

RSA

RSA is one of the main uneven encryption frameworks generally utilized as a part of information transmission. In the first place depicted in 1977, and depended on open qualities in light of two expansive prime numbers and helper values in encoded messages. Anybody can utilize general society key to encode the message, however just the individuals who realize that individuals can attempt to unravel the message can utilize the key. RSA has opened various cryptographic conventions, for example, advanced marks and scrambled voting techniques. It is additionally an open source innovation behind a few calculations, for example, PGP, which enables you to encode computerized correspondences.

ECC

Elliptic bend encryption is a standout amongst the most intense and minimum comprehended types of encryption. Supporters ECC innovation alludes to a similar security level, quicker run time, mostly because of a similar security, while utilizing a littler key size. Superior benchmarks are because of the general effectiveness of the elliptic bend, which makes them perfect for an implanted framework, (for example, a savvy card).National Security Agency is the largest supporter of technology; described above as an alternative to the RSA approach

1.16 LIST OF ALGORITHMS USED FOR ENCRYPTION

i. RSA
ii. Blowfish
iii. Two fish
iv. AES

1.16.1 RSA

In modern computer RSA algorithm is used to encrypt and decrypt messages . It is an asymmetric encryption algorithm. Asymmetry means that two different keys. It is also known as public key cryptography, because one of them will be given to each person. Another key must be kept private. It is based on the fact that: find an integer factor is difficult (factorization problem). RSA stands for Ron Rivest, Adi Shamir and Leonard Adleman, who first publicly described it in

1978.RSA users to create and publish product and assist value of two large prime numbers as their public key. The main factors should be kept secret. Anyone can use the public key to encrypt a message, but using the current release, if the public key is large enough, the only person who can determine the strength factor able to decode the message.

i. RSA Operations

In RSA two keys Public Key and Private Key are used. Everyone knows the public key, which is used to encrypt messages. The messages which are encrypted with the Public key could only be decrypted with the Private Key. The Keys for the RSA algorithms will be generated in the following ways:

Step 1: Select two big prime numbers

n,m

Step 2: Compute

$P=n\times m$

P is the modulus for the public key and private key

Step 3: Compute the totient

$u=(n-1)\times(m-1)$

Step 4:Slect small odd integer ‘r’ relatively prime to u

Gcd(r,u)=1

Step 4: Compute q such that

$(q\times r)\%u=(r\times q)\%\ u=1$

Step 5: Public key is (r, p)

Step 6: Private key is (q, p)

For example:

When we put the values in the above steps we get private and public keys

If n=11

M=29

Then $P=n\times m$

P=29*11=319

U=280 therefore r=3

Q=187

Then we get the value of Public key which is (3,319)

Private key value is (187,319).

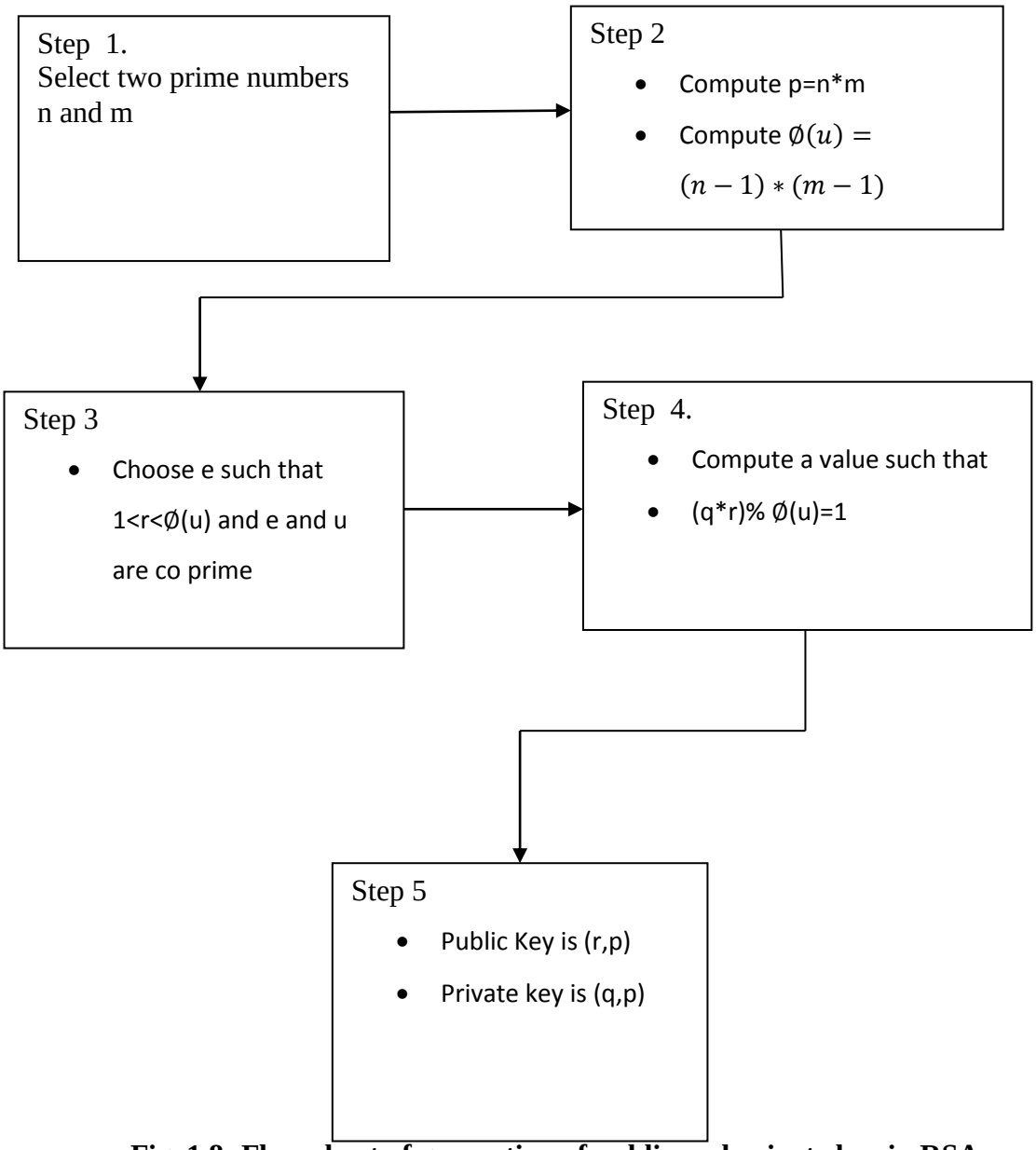

Fig. 1.8: Flow chart of generation of public and private key in RSA

Source:http://www.c-sharpcorner.com/uploadfile/75a48f/rsa-algorithm-with-c-sharp2/

The most commonly used algorithm is RSA algorithm which uses public key encryption algorithm.

ii. Encryption

Calculate cipher text S from plaintext message Z such that

Z=S^e mod p

Public key encryption can also be stated as asymmetric cryptography. In this two keys that are Public and Private keys are used. Public key is used for encryption and private key is used for decryption of text, also known as secret key.RSA algorithm is shown in figure below:

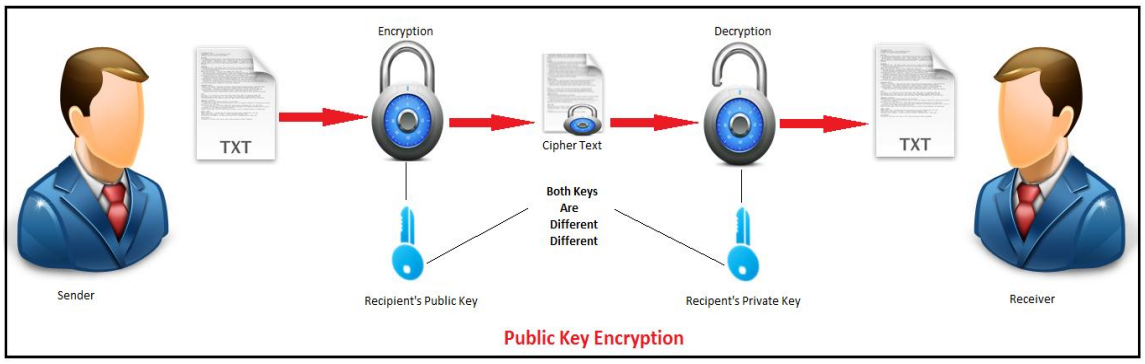

Fig. 1.9: Public key encryption

Source:http://www.c-sharpcorner.com/uploadfile/75a48f/rsa-algorithm-with-c-sharp2/

Figure 1.9 depicts how a public key encryption works. In Figure 1.8, users are shown: First of all, the sender (sends information like text or image through public key of the recipient).Second: the receiver (using the private key in order to receive something from the sender) Thus, in public key cryptography, the sender uses the receiver's public key to encrypt the data, and by using an 'encryption algorithm' can also be determined by the receiver, and the receiver transfers only the public key and encryption algorithm. However, by using the public key, the encrypted data cannot be decrypted, and the data at the receiver could only be decrypted by the private key. Therefore, no user could hack the data. In other words:

Public Key: Share by the public which needs to send the data.

Private Key: Confidential secret, so that when someone transfers through our public key to encrypt data with us, we can use the private key to decrypt the data.

1.16.2BLOWFISH ALGORITHM

The Blowfish is other algorithm given by Bruce Schneier in late 1993 and designed for replacing DES. The symmetric cryptographic breaks message in 64-bit blocks with encrypting them individually. It could be found in the software category, as of e-commerce platform to payment protection to protect password management tools.It is undoubtedly one of the most flexible

encryption methods.Their great speed and overall efficiency are well known, as many people claim that they have never been defeated.At the same time, providers have made the most of their free availability in the public domain.

It is a symmetric block encryption algorithm planned with,

- **Fast:** It encrypts the data on huge 32-bit microprocessors with a rate of 26 clock cycles/ byte.
- **Compact:** It could run in less than 5K for memory.
- **Simple:** It utilizes XOR, addition, lookup table having 32-bit operands.
- **Secure:** The key length become variable and it can be ranges from 32 to 448 bits. By default, 128 bits is considered as the key length.
- It is applicable for applications in which the keys cannot change fast like an automatic file encrypted or communication link.

i. Structure of Blowfish algorithm

It has a 64-bit block size with anywhere from 32-448 key length. It is a 16-round encrypted Feistel and uses large S-boxes relative to the key. It is similar to the use of fixed S-boxes in structure to CAST-128.

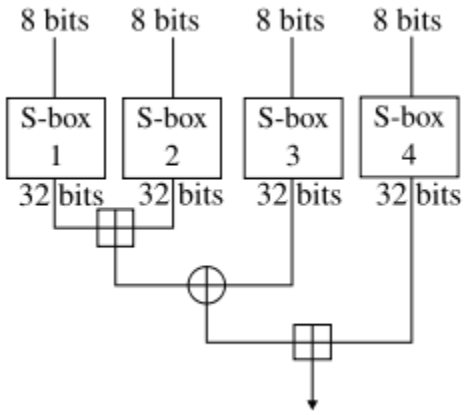

Fig. 1.10: Block diagram of BlowFish

Source: http://www.tips2secure.com/2016/02/blowfish-encryption.html

As shown in the figure above 8 bit data input is given to each of the S-Box which produces 32-bit outut.Each line here represents 32 bits.This algorithms has two subkeys which consists of '18-entry P-array' and 'Four 256- entry S-boxes'. In every round one entry of the P-arry is used, Each half of the data block is XORed with one of the two remaining unused P items. The F function of the Blowfish is shown on the right side of the graph. This function divides the 32-bit

input into four octets and uses these quarters as inputs to the S-box. The output is added to 232 module and exclusive OR to get the final 32 bit output

Since Blowfish is a Feistel network, it can simply be reversed by reversing P17 and P18, and then using the P entry in the reverse order.

This algorithm is divided into two parts.

1. Key-expansion
2. Data Encryption

ii. Key Expansion

Key expansion converts 448 bits into many sub key arrays of 4168 bytes. These keys are generated before the encryption and decryption of data.

The Q – array consists of 18, 32- bit subkeys:

Q1,q2,q3…….q18

Four 32-bit S-Boxes comprised of 256 entries with each:

'r1,0, r1,1,……. r1,255'

'r2,0, r2,1,…….. r2,255'

'r3,0, r3,1,…….. r3,255'

'r4,0, r4,1,.....r4,255'

The sub-keys are computed by using the Blowfish algorithm as defined below:

1. P-arry is initialized first and then the four S- boxes,with a fixed string.This string consist of values in hexadecimal values of Pi. Q1 = 0x243f6a88, q2 = 0x85a308d3, q3 = 0x13198a2e, q4 = 0x03707344, etc.
2. Q1 is XORed with the first 32 bits of the key, XORed with the second 32 bits of the key, and so on. This loop is repeated until the entire q-arry has been XORed with the key bits. There is a fairly long key for each short key. For example, if A is a 64-bit key, AA, AAA, etc. are equivalent keys.)
3. Using the subkeys described in steps (1) and (2) all-zero string with the Blowfish algorithm get encrypted .
4. Replace q1 and q2 by the output of step (3).
5. Output of step (3) is Encrypted by utilizing the Blowfish algorithm with the adapted sub-keys.
6. Replace q3 and q4 with the output of step (5).

7. The process will be uninterrupted by replacing all the entries of the q array with the next four S boxes in the order of the output of the ever-changing Blowfish algorithm.
8. A total of 521 iterations are required to generate all subkeys. The application can store the child, rather than executing the derived process multiple times.

iii. Data Encryption

It has a function that has repeated 16 times in networks. Each round includes replacement and key-dependent permutation. All operations are 32-bit word XOR and addition. The only additional operations are four indexed array of data per cycle look-up table.

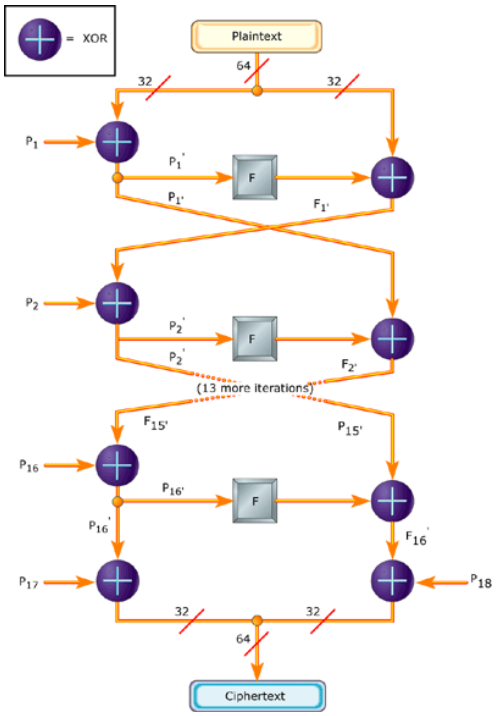

Fig. 1.11: Data encryption algorithm

Source: http://www.tips2secure.com/2016/02/blowfish-encryption.html

iv. Algorithm of Blowfish Encryption

Split M in two 32-bit halves: Ml,Mr

For i=1 to 16:

Ml= Ml XOR Ui

Mr=D(Ml) XOR Mr

Swap Ml and Mr

Swap Ml amd Mr(Undo the last swap)

Mr=Mr XOR U17

Ml=Ml XOR U17

Ml=Ml XOR U18

Recombine Ml and Mr

v. Practical examples blowfish algorithm

Without private RSA key, it has never been a radio wave, an eavesdropper cannot obtain Blowfish key, it cannot decrypt the messages transferred among the two machines.
Like

Suppose if you want to build embedded system security through laptop session data exchange, possibly through wireless media. At the beginning of the meeting, embedded systems and laptops provides Blowfish key private and public and private RSA keys. Embedded systems and laptops exchange public RSA keys and use them to exchange their private and Blowfish encryption keys. Encrypt the rest of the communication using two machines encrypted with Blowfish. When the communication time ends, all the keys are discarded.

1.17 NEED OF HYBRIDISATION

The concept of hybridisation comes into place due to following reasons:

i. Hybrid encryption

- Mixture of more than single cryptographic algorithm
- It integrates a mixture of asymmetric and symmetric encryption
- It has more security.
- Hybrid encryption is taken as highly secure kind of encryption for public and private keys being fully secure.

ii. Need of Hybrid Encryption Technology

- System of conventional encryption utilizes only one key.
- If key is disclosed then encryption and decryption process is completed and becomes susceptible.

- Problem of Key distribution.
- To give greater security for avoiding explicit transfer of secret key than hybrid encryption system is required.

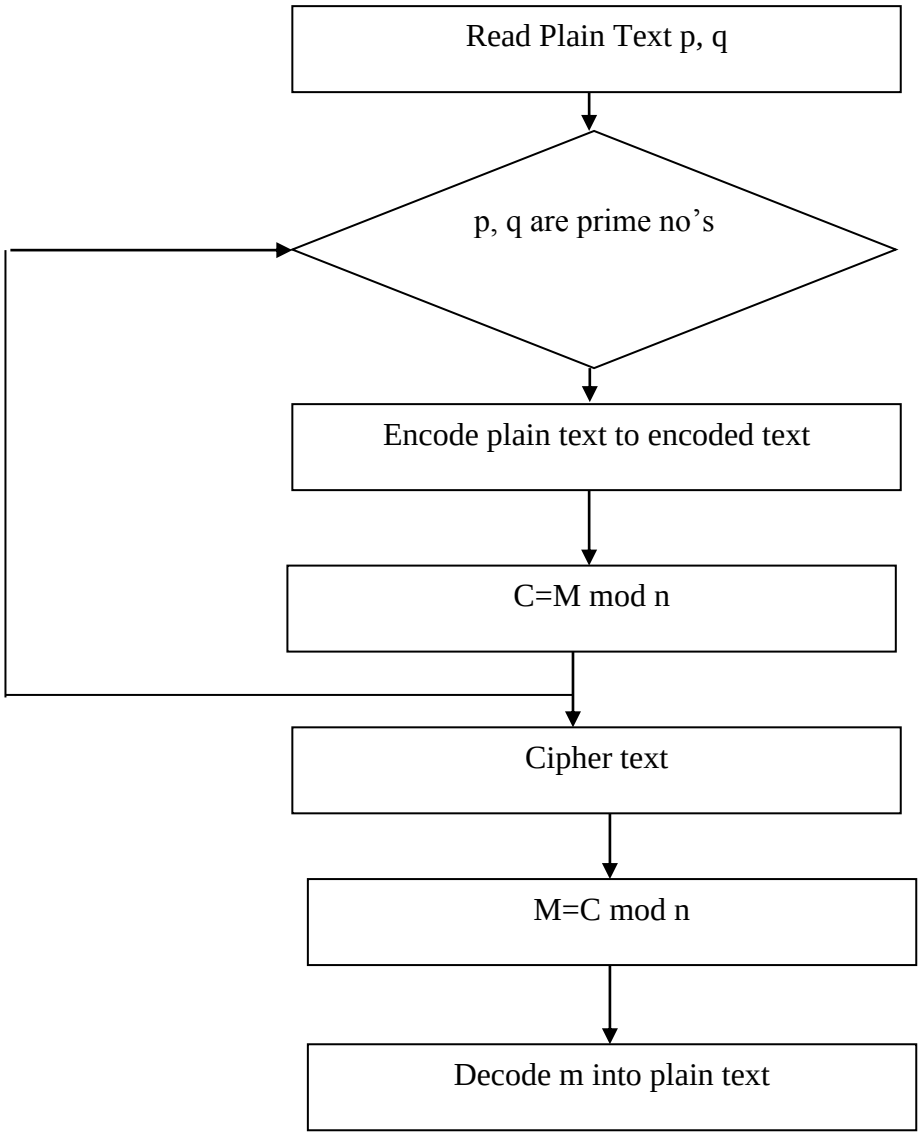

Fig. 1.12: RSA flowchart

1.18 MD-5 ALGORITHM (Message-Digest algorithm 5)

In cryptography MD5 has been used widely with 128-bit hash value. MD5 has been used mainly in security applications. MD5 has mainly '32-character hexadecimal number'. MD5 has been executed by 'Ronald Rivest' for replacing MD4 hash function. Due to presence of some flaws

researchers start using SHA algorithm irrespective of this algorithm. As Message-Digest-5 make use of only 1 pass so two prefixes with same function are generated that leads to collision.

1.18.1 Applications of HASH function

- It makes sure that file has been arriving intact.
- A person creating a file can make other file of same checksum value.
- MD5 can also be used to store passwords.

Algorithm

- Variable length message processing into 128-bits.
- Input message broken into 512 blocks.
- Message padding takes place.
- Remaining bits are filled with 64-bit integer. The master algorithm operates on each 512-bit message block, and each block modifies the state. The processing of the message block consists of four similar phases, called rounds; each round consists of 16 similar operations based on the non-linear function F, the module addition and the left-hand.

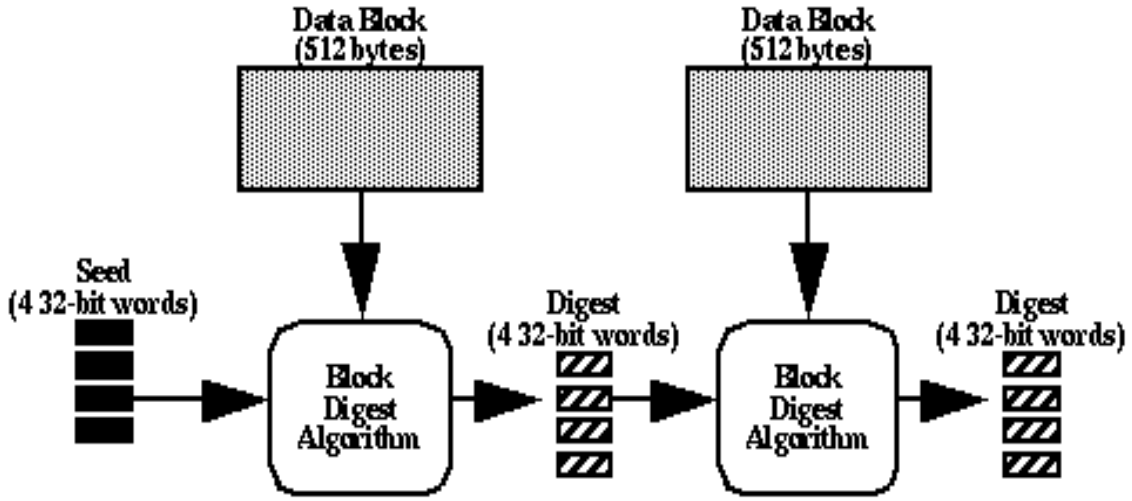

Fig.1.13: MD5

1.19 ARTIFICIAL NEURAL NETWORK

The simplest neural network is ANN (Artificial Neural Network). ANN is designed to be closely related to the neural structure of the brain. The neural network is mainly composed of layers. Layer consists of nodes. The nervous system has three types of layers: the input layer, the hidden layer and the output layer. On the basis of the weight of the output layer, the output corresponding to the input layer is assigned. Most ANNs contain learning rules that help to maintain weights based on the output layer. There are many types of learning rules in a neural network, but incremental rules are conventional rules that are now used.

A number of helpful properties of ANN are defined below:

1. 'Adaptability'
2. 'Non-linearity'
3. 'Uniformity of analysis and project'
4. 'Mapping between Input-Output'
5. 'Fault-tolerance'

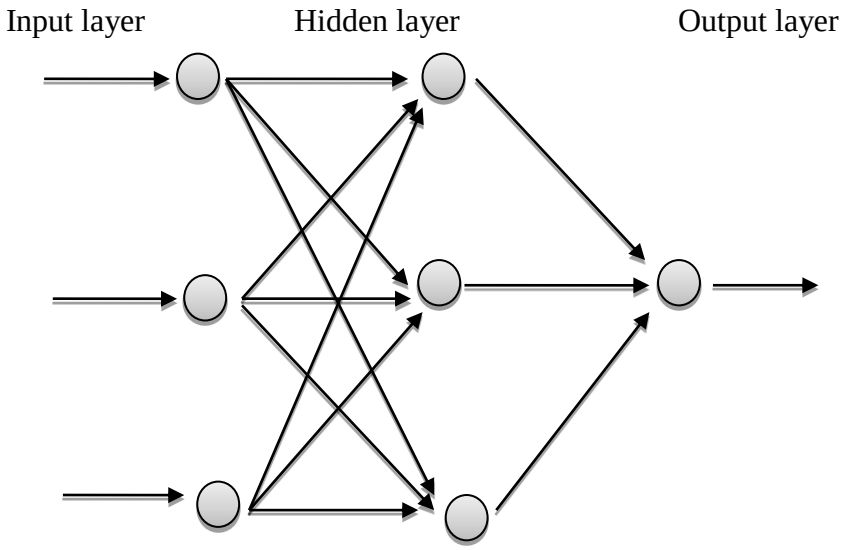

Fig.1.14: Neural network architecture

CHAPTER 2: Literature survey

2.1 RELATED TO BIG DATA

Jungang Xu et al. 2016 proposed, a 5-layer performance evaluation model for big data system , that was a reliable basis for performance analysis, and at the same time. Author also proposed a performance optimization model for big data system , which could help to perform bottleneck location and bottleneck analysis, and also measure performance prameters. An event-based performance tool to profile performance data was implemented based on these two performance models. Simulation results showed that these two performance models are effective for performance evaluation and optimization of big data system, an average running time of big data system will be improved by 19%.

Abdullah Al-Shomrani et al. 2017 focused on the privacy, security and how one could make sure that big data is secured. Managing security policy is a challenge that this proposed framework will handle for big data. Privacy policy require to be integrated, flexible, context-aware and customizable. Author has built a framework to receive data from customer and then examine data received, extract privacy policy and then recognize the sensitive data. Author also presented the techniques for privacy policy which will be created to be used in the framework.

Sahel Alouneh et al., 2016 proposed a method to defend the large data during the analysis can classify the data before any action, such as moving, copying, or processing. Based on large data classification, the security program will be activated according to the key level of the data.

Nitin Naik et al. 2016 presented an sensitive and economical the Big Data Security Analysis Method uses a computer intelligence technology for Windows desktop users, where the combination of Windows batch programming is simulated on a real data set with more than 10 million interpretations collected in the Windows Firewall log to show that desktop users can gain insight into Its rich and complete data, and access to useful information to prevent its systems from being exposed to current and future security threats.

Neetu Chaudhari and Dr. Satyajee Srivastava, 2016 described all the possible challenges and problems associated with large data, while proposing solutions. For privacy protection, different data mining techniques have been implemented, such as anonymization, randomization, and so on. The classification, volume, speed, and variety of large data are very rapidly increasing in order to simultaneously address these challenges, have been implemented to modify, and need to develop more reliable and flexible system algorithms. Additional analysis techniques must be provided for reliance on anomalies and attack detection of common shared datasets.

2.2MULTIMEDIA SECURITY

Guillermo A. et al., 2009 presented situation of image processing technology and the related work of multimedia information security being applied. The main problem with protecting multimedia information is that a large amount of data can be used. Therefore, the common theme found in the literature is to try to limit the perceived effort by identifying certain attributes of the image that should be protected.

George Anastasios Spanost and Tracy Bradley Maples, proposed an Aegis mechanism. This method is used to restrict the amount of data to be encrypted or decrypted by using video compression to reduce the size of the video image sent. Through a wide range of simulations, the performance parameters of Aegis using MPEG video compression through ATM network are explored. The authors has used three types of video traffic: CATV, STU&O TV and video conferencing.

2.3 ENCRYPTION

D.I.George Amalarethinam et al, 2015 introduced a security level by using a separate magic rectangle. It increases the calculated values in the ciphertext. In this article, duplicating of the same password value and the ciphertext is formatted as an image. When this or even the same value is repeated, it assigns a different value for each occurrence. The author has successfully encrypted or decrypted images with different secret keys and algorithms. The cipher text will be completely different from the original image file and will be used for secure transmission over the Internet. This model provides an additional level of security for public key algorithms and uses memory correctly.

Manoj patil et al.,2014 proposed a method of transferring SMS text by using compression and encryption combination. The ended data is first encrypted by utilizing ‘Elliptic curve Cryptographic technique’, But the length of encrypted data is so large so we have to compress

the data so that data can be send in short duration, Compress technique that is implemented here will compress the data up to 99.9%, hence the large bandwidth get saved. This Compression-Encryption of SMS text message could be implemented for the system of android Operating. In this paper hybrid compression Encryption technique is executed for variety of O.S like Symbian and Android and results will be analyzed.

Yashpalsingh Rajput et al. , 2014 presented an advanced scheme for encrypting the plain text message for security issue. Whole of the conventional encryption techniques are weak and traditional cryptanalysis could be used to early determine the plain text from encrypted text. In this paper a new concept is used to increase the security of the text which is stronger and secure as compare to the earlier concept. Plain text is encrypted in a way that is difficult to decrypt. The proposed system has two phases. In the first phase, a new method of using a multi-letter cipher technique converts plain text messages to the first encrypted text. Encryption is accomplished by using a variable-length key. In the second phase, the mountain cipher technique is applied to the first encrypted text to produce a new encrypted text or ciphertext. At the receiver, if the receiver has the appropriate decryption key, user can generate the same text message as the original message.

Vinay Verma et al ,2014 has proposed An efficient security method based on symmetric key and develops dynamic keys from multimedia files and encrypts multimedia data. In this technique, dynamic keys are randomly selected from multimedia files by using special features. The technology gives 0 loss results in all user data and multimedia content for encryption. This will make encrypted data faster.

Sombir Singh et al. , 2013 proposed a SCTTMR Simple Columnar Transposition Technique with Multiple Rounds. By using Simple Columnar Transposition Technique the plain text is firt converted into cipher text. The different rounds of SCTTMR will depend upon the security in order to provide the message. If we need more security then we add more rounds of the SCTTMR scheme and for normal security then use minimum 1 or 2 rounds. For SCTTMR input is is a plain text message and the output is ciphered text. The output from SCTTMR is then converted into a bit form because the DES algorithm applies its process on bit level as usual. In this paper DES performs same work as the original DES.

Dimple et al. 2013 presented different encryption techniques like (RSA) Rivest Shamir and Adleman, Diffie-Hellman, (Digital Signature Algorithm) DSA are executed. In the Diffie-

Hellman cipher algorithm, it is experienced that the keys are exchanged between two users. In DSA, a digital signature is used to confirm that the received signal is unchanged. Different applications use different technologies. Daily new encryption technology is developed, so fast and secure encryption technology always be implemented with high security.

Nentawe Y. Goshwe. et al., 2013 implemented data encryption and descryption in a network environment . By using this software, data can be travel from one computer to another computer via an unsecured network environment. encryption and decryption is one of the best method to hide the meanings of a message from intruders in a network environment.This paper use RSA algorithm for encryption description of the data. In thi algorithms sender generate a public key to encrypt message and at the receiver end a private key is generated using a secure database. If the generated private key is wrong then it still decrypts the encrypted message but it is different from the original message.

Rohan Rayarikar et al. ,2012 has developed an application on the Android platform that allows users to encrypt and decrypt messages before transmission over the network, using advanced Encryption Standard algorithms for encrypting and decrypting data that can be run on any device that works on the Android platform, which provides secure, fast, and powerful data encryption.

Matthew Wangsadiredja et al., 2011 presented Cipher block encryption algorithm with password feedback (CFB) 8-bit mode. Select this mode because it matches data in 8-bit aligned BlackBerry devices. The encryption algorithm uses a 64-bit block size cipher block to encrypt. The authors have designed an algorithm based on Feistel network and some encryption/decryption algorithms. Then cipher block encryption algorithm is implemented in the software based on Java BlackBerry. In this software development, the Eclipse Eclipse Ganymede has been used by some plugins. The software is also used to process files like pictures and audio by displaying their headers, and to manipulate text encryption by converting ciphertext data into hexadecimal digits. After the experiment, the software has good compatibility with many kinds of BlackBerry equipments and has good security.

CHAPTER 3: PROBLEM STATEMENT AND OBJECTIVES

3.1 PROBLEM STATEMENT

Apart from the clustering and arranging the data on Big data computing environment, there are few more concerns which have to be considered. Security of data is one of the major concerns apart from Clustering.

A lot of encryption algorithms are present in the current scenario but the nature of every algorithm is different. Hence generalizing the encryption algorithm would not be sufficient enough.

The problem of this research work comes from the above written paragraph. The problem is to design a framework which can evaluate that which kind of encryption algorithm will suit which type of data. The problem statement also includes design of a hybrid encryption algorithm to make this successful epic journey more stronger. Classification of the encryption algorithm will also be a part of the problem statement.

The late development of arranged sight and sound frameworks has expanded the requirement for the security of computerized media. This is particularly very important for the security and implementation of protected innovation rights. Computerized media incorporates content, advanced sound, pictures, feature and programming. Various methodologies are available for ensuring advanced information; these incorporate encryption, verification and time stamping. Anyhow, in this proposed work mixed media information can be secured utilizing hybridization of encryption calculations with neural network. Parameters that have been computed for the proposed work are written below:

i. Precision: It is the percentage of the number of related data records retrieved and the total number of unrelated data records. It is defined as

$$Precision = \frac{Relavant\ Data - Retrieved\ data}{Retrieved\ Data}$$

ii. Recall: It is the percentage of the number of related data records retrieved and the total number of related data records. It is defined as

$$Recall = \frac{Relavant\ Data - Retrieved\ data}{Relavant\ Data}$$

iii. Accuracy: It is the indication of the closeness of the proposed work to reality. Accuracy is the description of the errors in proposed work. If error of any proposed system is more than the accuracy of system should be less, so the accuracy of any proposed system should be more.

iv. F-measure: The measure in which the combination of precision and recall takes place is the F-measure.

$$F = 2.\frac{Precision.recall}{Precision + recall}$$

3.2. OBJECTIVE

i. To study the various existing encryption and classification techniques.

ii. To secure the data in proposed work use hybridization of encryption techniques.

iii. To apply the encryption on the data use complexity finder techniques and on the bases of complexity of the encryption technique.

iv. To classify the data correctness use artificial neural network technique as a classifier.

v. To check the efficiency of proposed work we calculate the performance parameters like precision, recall, f-measure and accuracy.

CHAPTER 4: METHODOLOGY

4.1 METHODOLOGY

Following are the different steps for the proposed work:

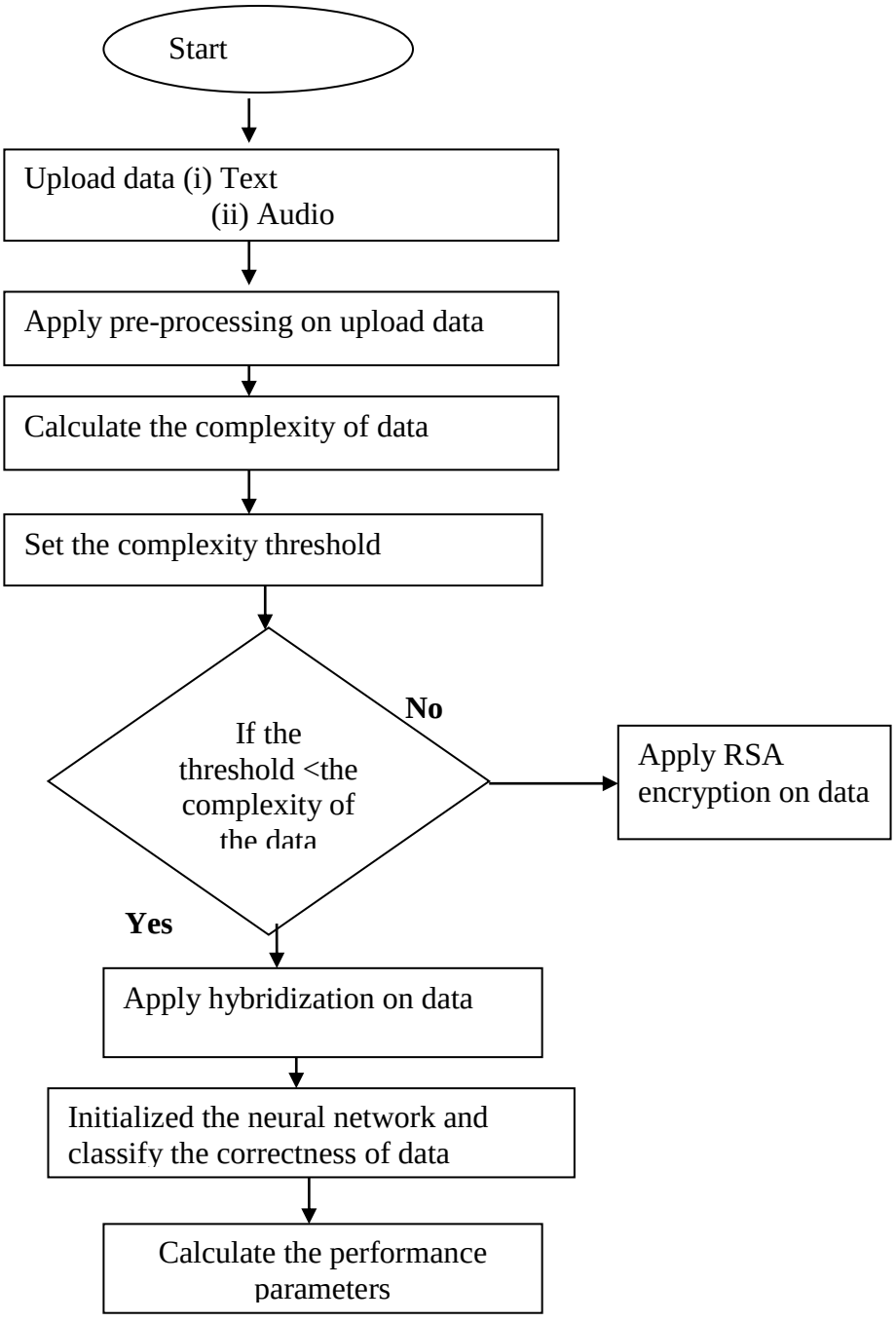

Fig. 4.1 flow chart of proposed work

Step 1: Initially upload the text file and audio file

Step 2: Now apply pre-processing on the uploaded data.

Step 3: Now find the complexity of the uploaded data.

Step 4: Set the threshold value. If the threshold value is less than the complexity finder value than RSA Encryption technique has been applied otherwise apply hybridization encryption technique.

Step 5: Initialize the neural network in order to find the correctness of the data.

Step 6: At last calculate the performance parameters like precision, recall, f-measure and accuracy.

CHAPTER 5: EXPERIMENTAL RESULTS

5.1 SIMULATION ENVIRONMENT

Table Error! No text of specified style in document.**:1 Tools used**

Computer	Core 2 Duo or higher
RAM	3 MB
Platform	Windows 7
Other hardware	Keyboard, mouse
Software	Matlab 7.0.4

The name Matlab represents the Matrix LABORatory. MATLAB is originally designed to conveniently access the matrix software developed by LINPACK (linear system software package) and EISPACK (Eigen system software package) project, and Matlab is the high-performance language of technical computation. It integrates calculations, visualization, and programming environments. In addition, MATLAB is a modernized programming language environment: It has a fine data structure, contains built-in editing and debugging tools, and supports object-oriented programming. These factors make MATLAB a prominent tool for education and research. Compared with the traditional computer language (for example, C,FORTRAN), MATLAB has many advantages to solve technical problems.

MATLAB is an interactive system whose basic data element is an array of no dimensions. The package has been commercially available since 1984 and is now considered a standard tool for most universities and industries in the world. It has a powerful built-in program that can perform various calculations. It also has an easy-to-use graphical command that makes visual results available immediately. Canonical applications are collected in packages called Toolbox. Toolbox for signal processing, symbolic computation, control theory, simulation and optimization

After you log on to the account, you can enter Matlab by double-clicking the MATLAB shortcut icon (7.0.4) on the Windows desktop. When you start the MATLAB, a special window called the MATLAB desktop appears. The desktop is a window containing otherwindows.

The main tool for the desktop or accessible from the desktop is as follows:

- The Command Window

- The Command History
- Workspace
- The Current directory
- Help browser
- Start button

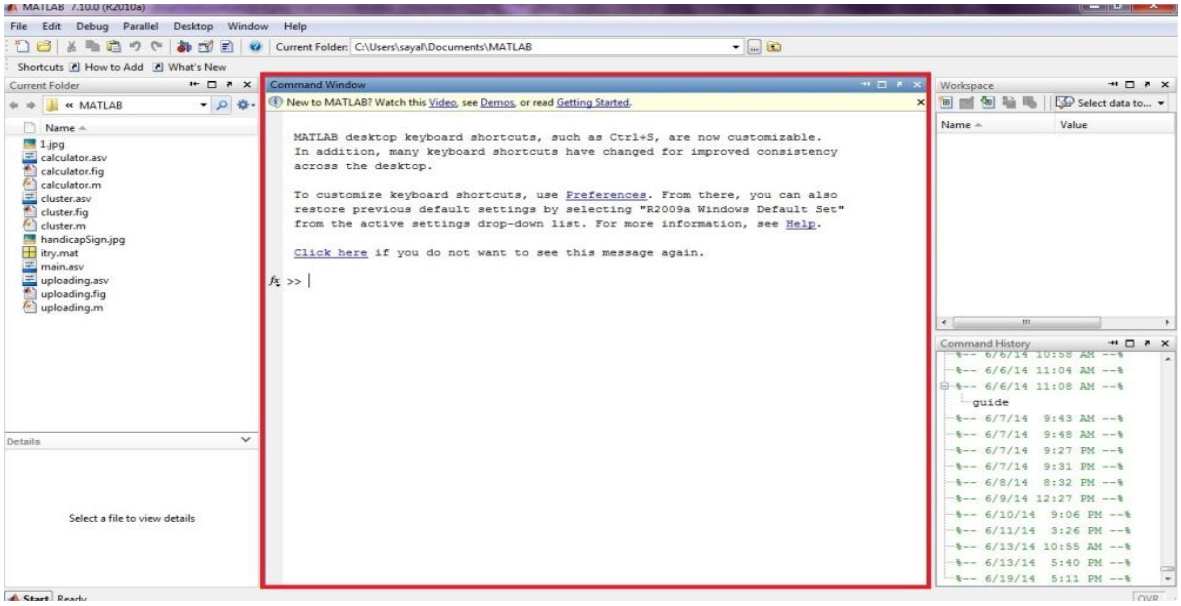

Figure 5.1Command Window

The figure where the commands can be written is command window. Non-graphic output is shown in this window. A '>>' displays that the system that can be used for input. 'Ready' and 'Busy' are displayed at the lower left hand corner of the window when the system is calculating or waiting.

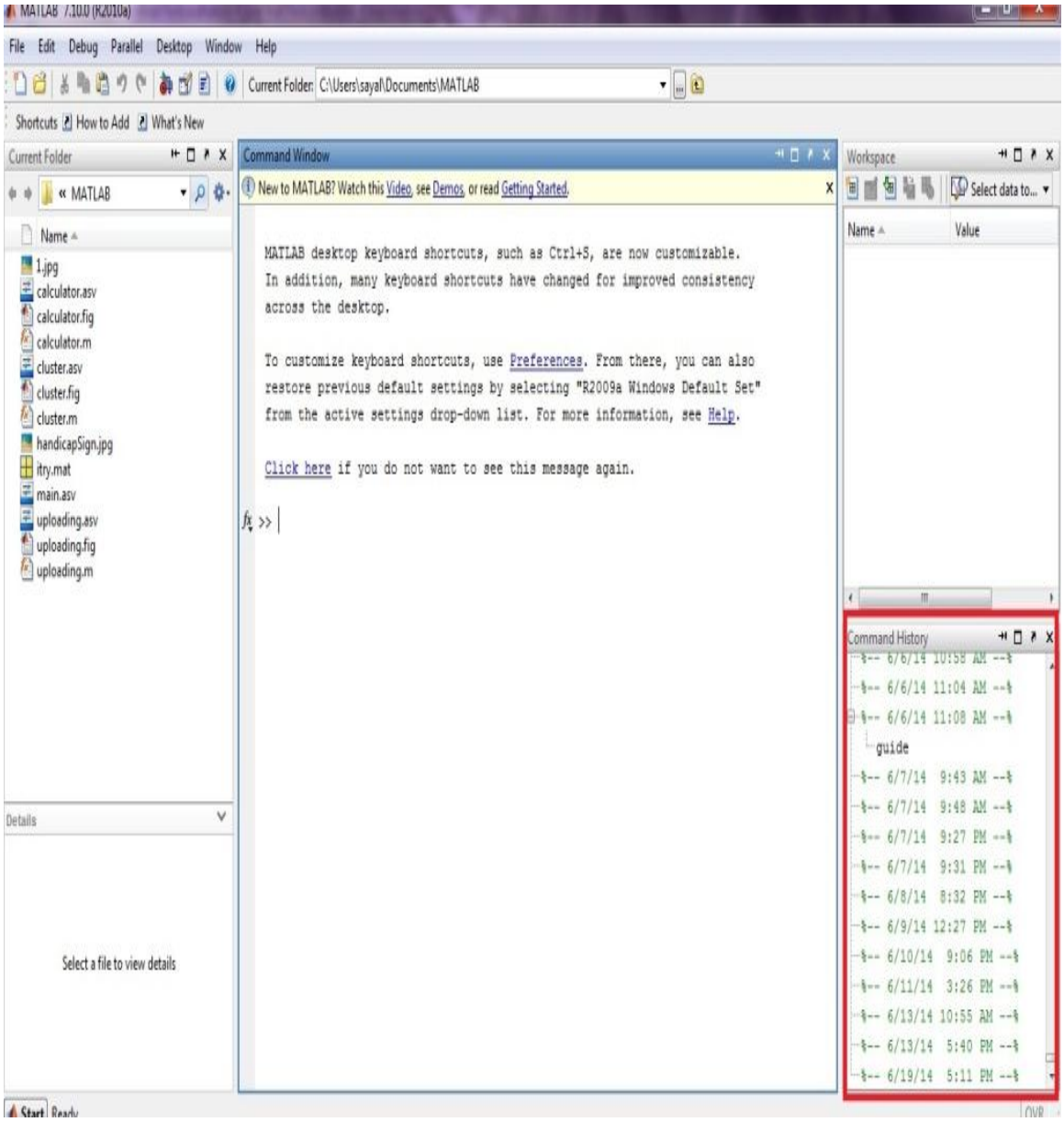

Figure5.2Command History

Above figure is defines the command history that depicts the history of the session and the recent sessions. It can also be used for copy/paste or reference commands.

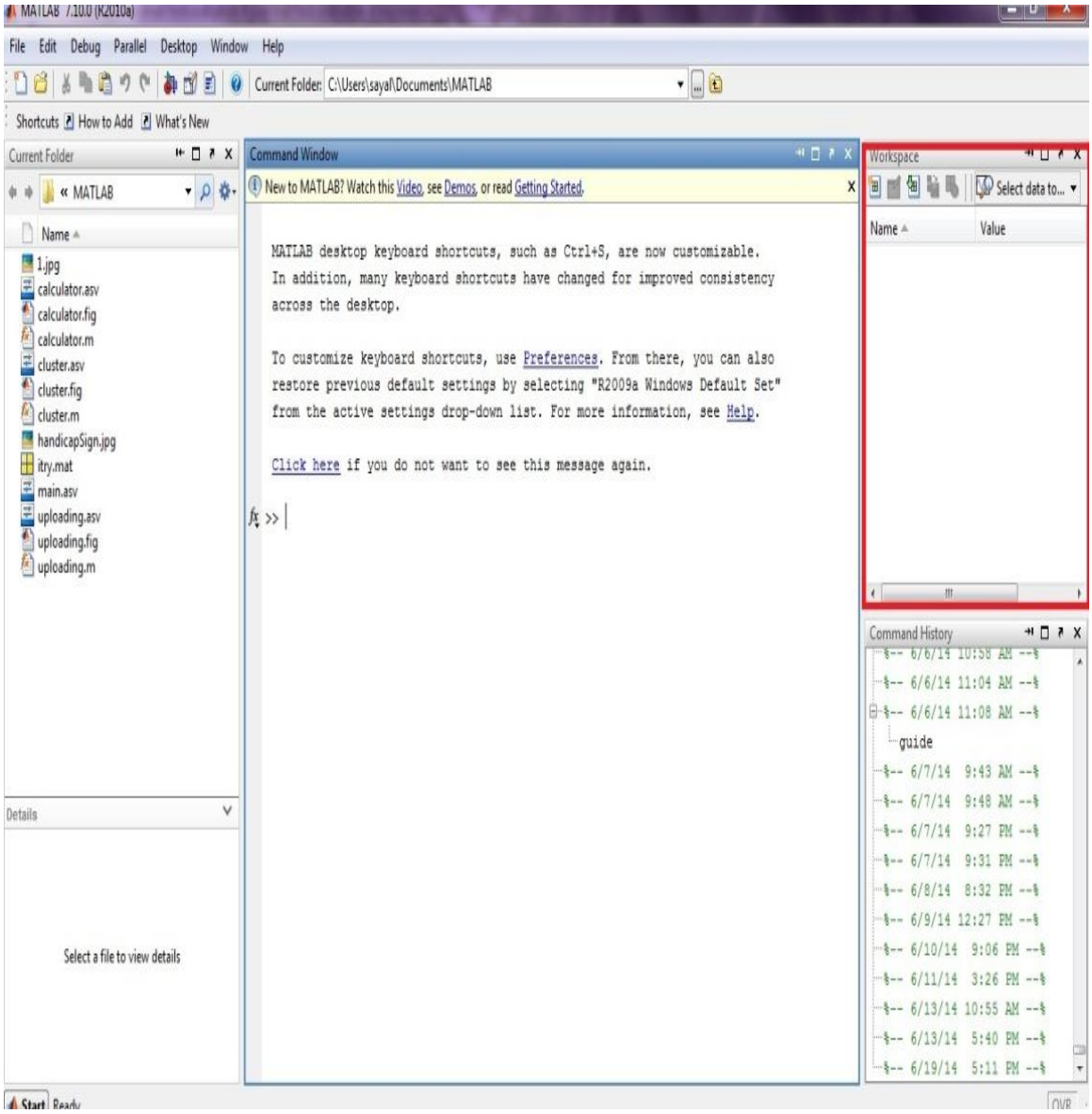

Figure5.3Workspace

The figure displays variables that could currently be defined by the user and the basic data of each variable with its min. And max. Values and the dimensions. The top icons allow the user to create various functions like variables, saving/deleting, creating with plotting etc.

File>Save>workspace command is used for saving the variables from one session to another. '.mat' extension is used for the workspace.

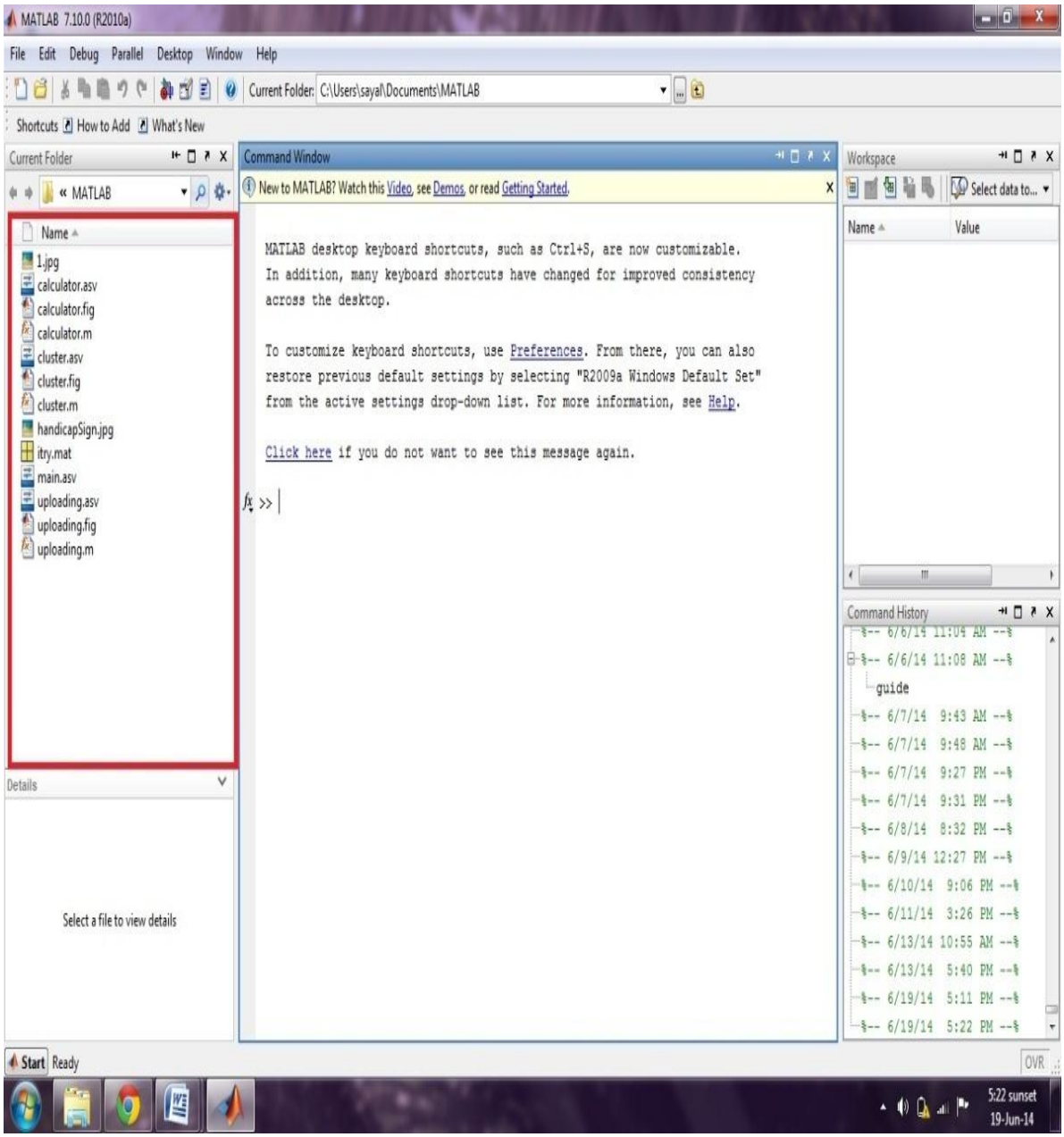

Figure 5.4Current Folder

The current folder is the folder on which the MATLAB is currently working in. Anything can be saving by default. The current folder at the centre top can change the directory. It also displays the list of the files in the current directory.

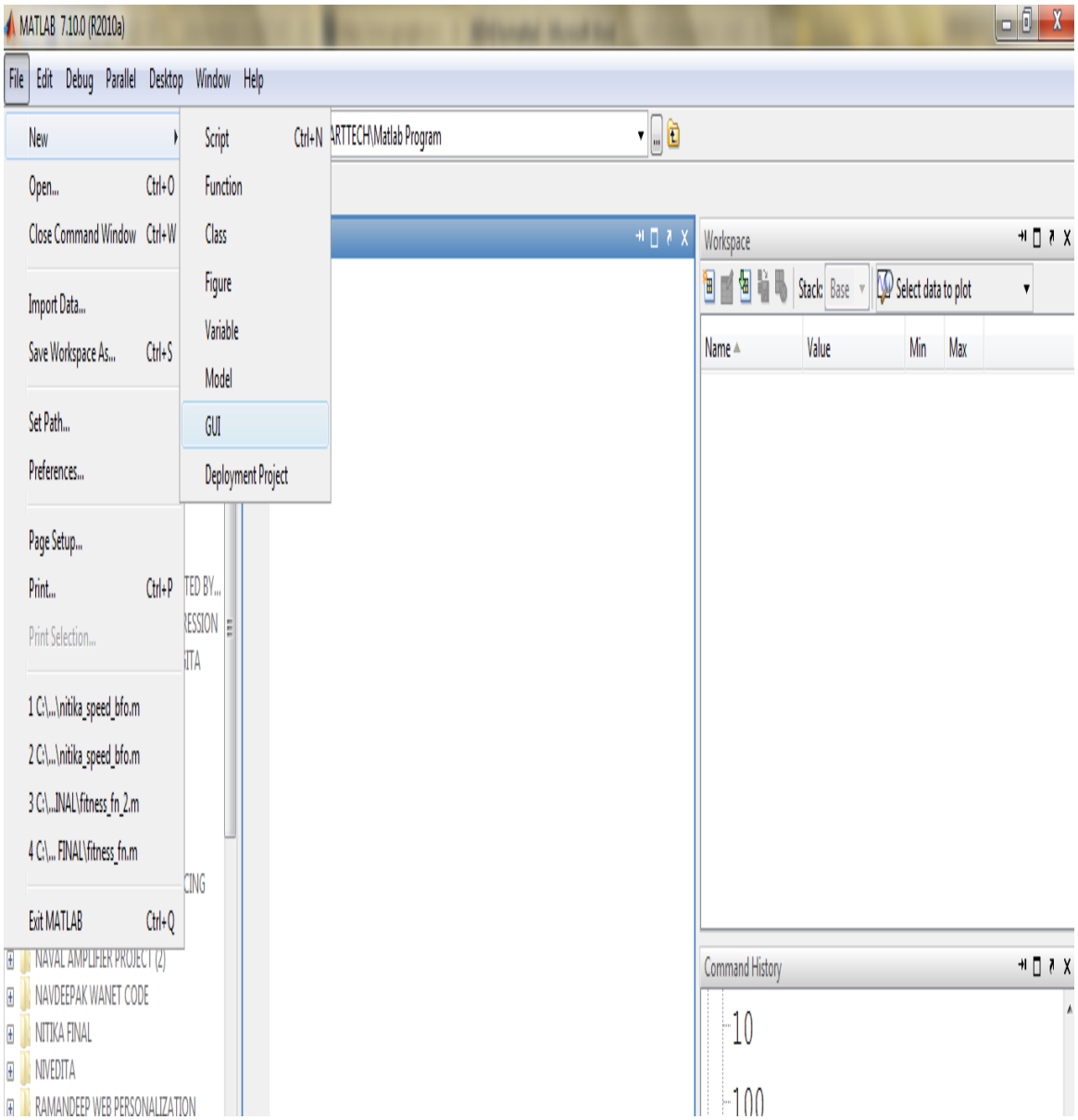

Figure 5.5 Start GUI

The figure shows the Start GUI screenshots. By clicking on the new menu from the top icons, a new GUI can be form.

- Opening a New GUI in the Layout Editor:

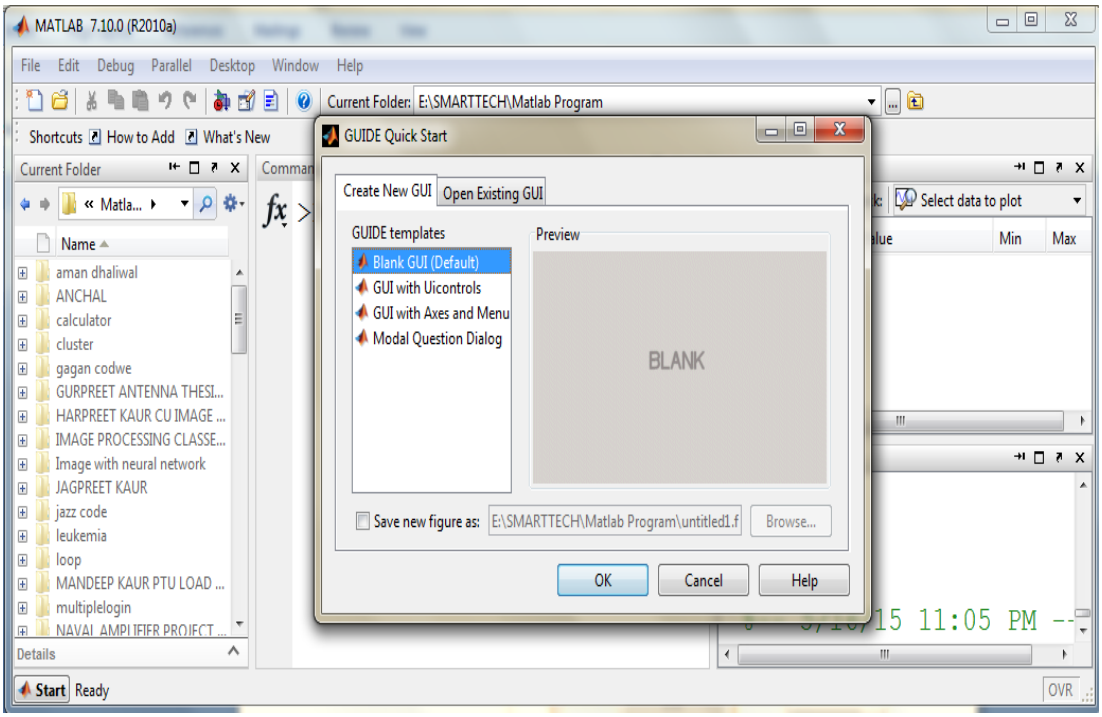

Figure5.6 NEW GUI

The above Figure defines different options for making GUI. Blank GUI's could be develop or GUI'S with some option.

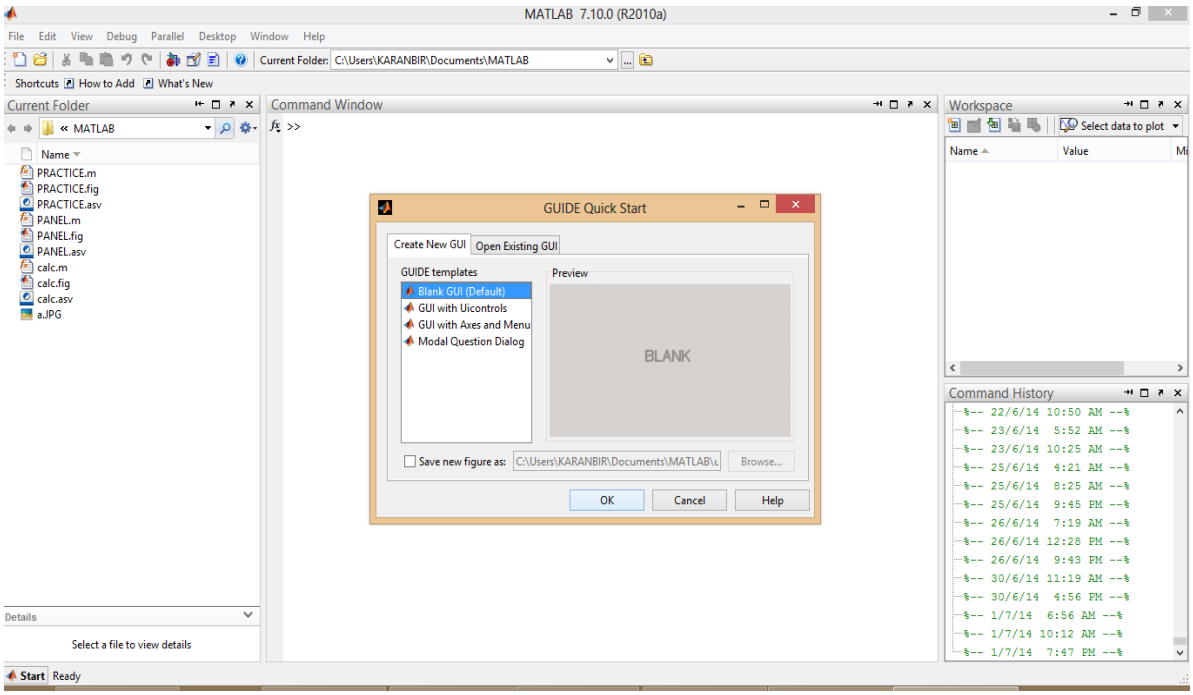

Figure5.7 Other Options for GUI

The figure above shows how to create a new GUI or open an existing GUI. To develop the GUI as needed, we use a blank GUI or the default option. Otherwise, there are many options available to use the predefined GUI.

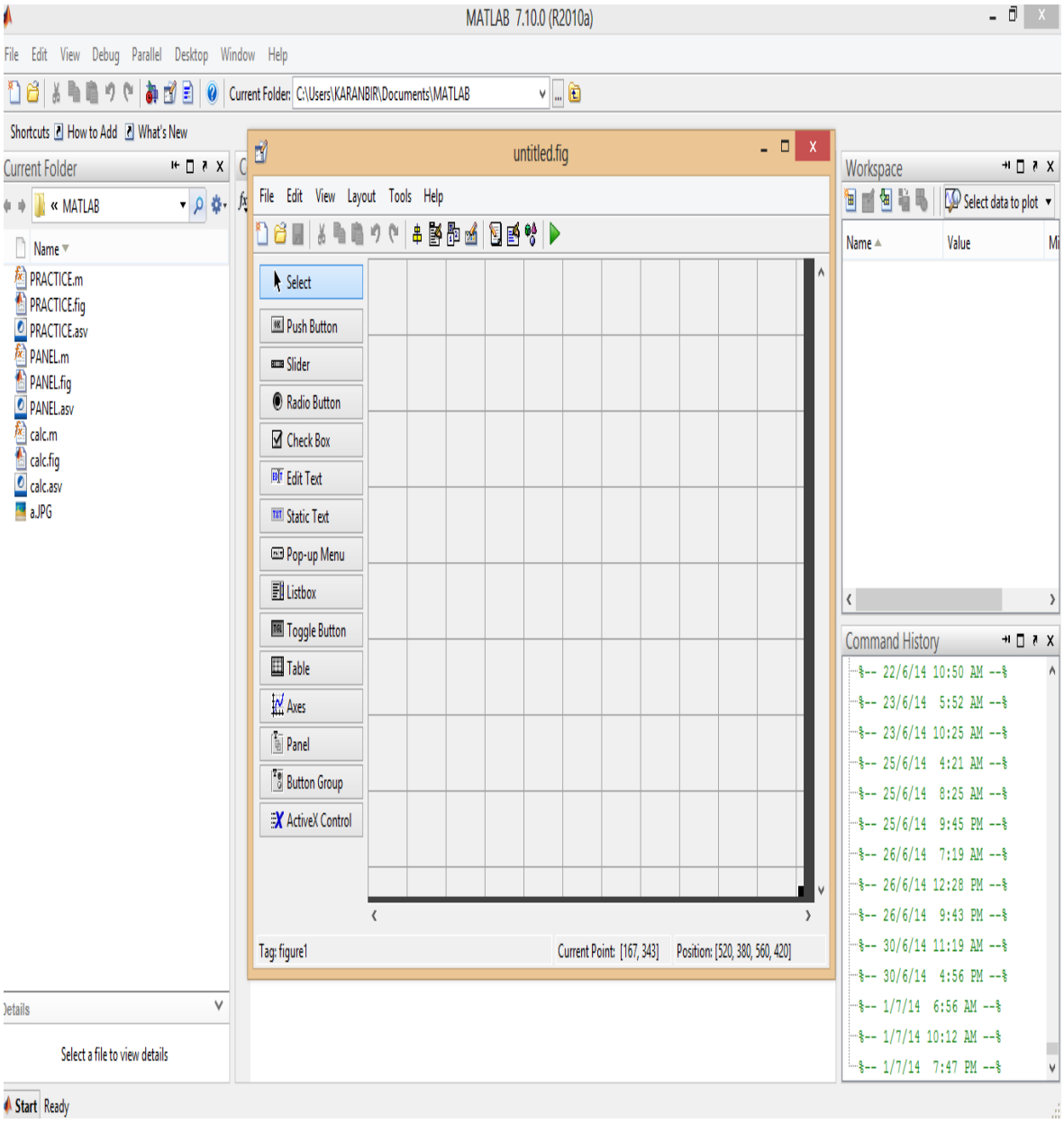

Figure5.8 To Draw a GUI

The above illustration shows a platform for drawing a GUI. Save the GUI by using the extension '.fig'. Many tools can be used to design and control to set the properties of a tool.

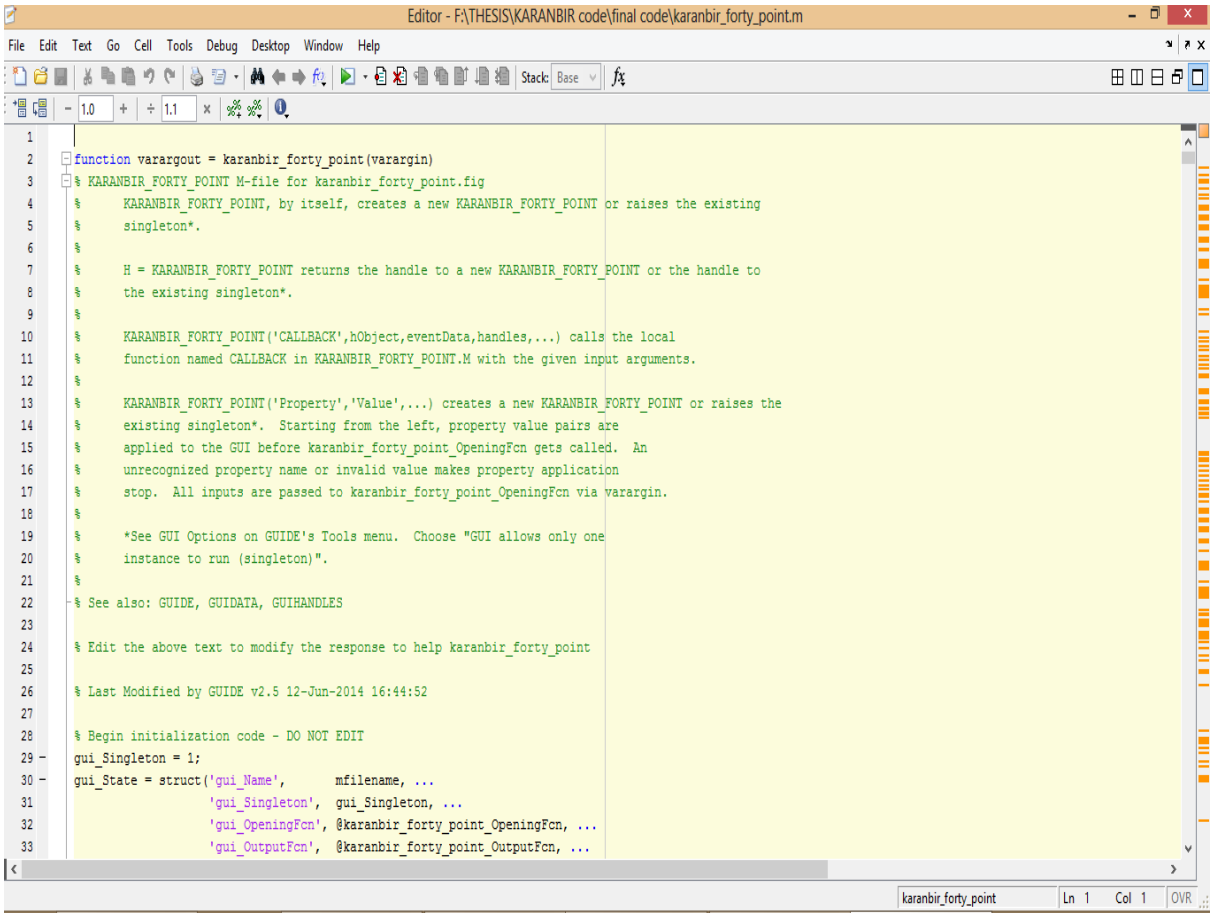

Figure5.9 Editor Window

The code can be edit even after saving it in the editor window. The user can edit .m-files; these are the files that has the scripts with the functions that were defined earlier. By typing 'edit' command in the command window, the above window can be displayed. By 'edit myfile' command, .m file can be open for editing.

5.2 MATLAB CHARACTERISTICS

- Proposed by Cleve Moler in the 1970's
- Taken from FORTRAN subroutines of LINPACK as well as EISPACK with linear and eigen value systems.
- Executed as an interactive system for accessing LINPACK and EISPACK.
- Gained the esteem by authoritatively dispersed.

- It was re-written in C in late 1980's having functionality, including plotting routines.
- The MathWorks Inc. was produced (1984) to marketplace and go on with expansion Of MATLAB.

5.3 STRENGTHS

- It might behave as a calculator or as PL (programming language)
- It integrates adequately calculation with the graphic plotting.
- It is reasonably easy to learn
- It is interpreted/not compiled, errors are easy for fixing.
- It is optimized to be relatively fast while executing matrix operations
- It has few elements of object-orientation.

5.4 RESULT ANALYSIS

The model of the proposed work is shown in the table 5.2.

Table Error! No text of specified style in document.**:2 Result Simulations**

Number of text document	10
Number of audio files	10
Simulation Tool	Matlab
Evaluation Parameter	Precision, Recall, accuracy, F-measure

Here, we formulated the following results in MATLAB with 10 number of text document .The figure 5.10 is for the training of the neural network trained with different text document.

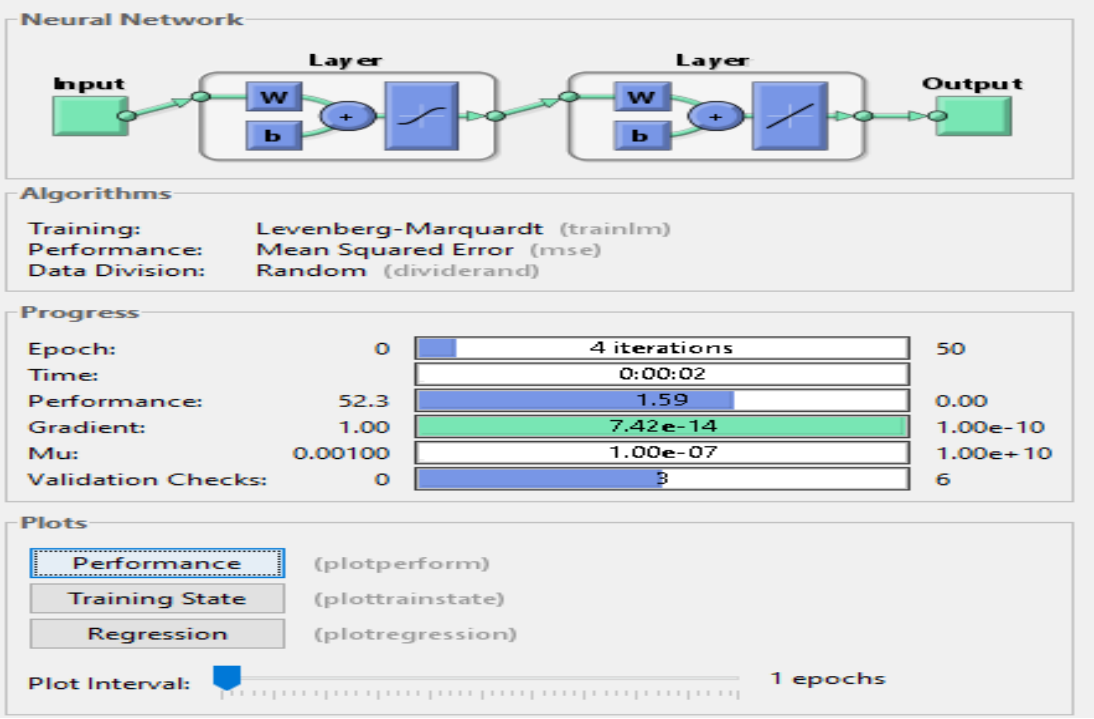

Fig. 5.10 Neural Network training

Artificial neural network training architecture is shown in figure 5.10. As shown in the figure 5.10 neural network has three layers.

i. **Input layer:** The trained data is provided on this layer.

ii. **Hidden Layer:** Processing of the trained data is done in this layer.

iii. **Output Layer:** Classified results are taken from this layer.

5.4.1 EXPERIMENTAL RESULTS FOR TEXT DOCUMENT

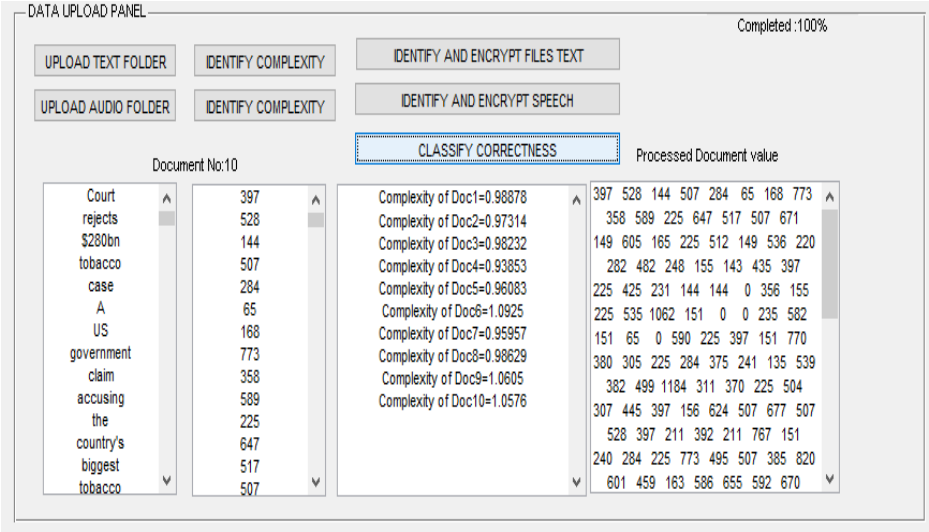

Fig. 5.11 Data Uploaded Panel

Firstly text folder is uploaded by clicking on 'upload text folder'. In the proposed work there are 10 texts that we have used to find the results. After uploading Text1, next step is to find the complexity which is calculated by clicking on identify complexity. After clicking on identify complexity, the complexity for all the uploaded document has been displayed. The next step is to click on identify and encrypt files text. The encrypted document has been displayed in the processed document value panel. Now, the last step is to calculate parameters. When we click on classify correctness all the measure parameters has been displayed in the parameters panel.

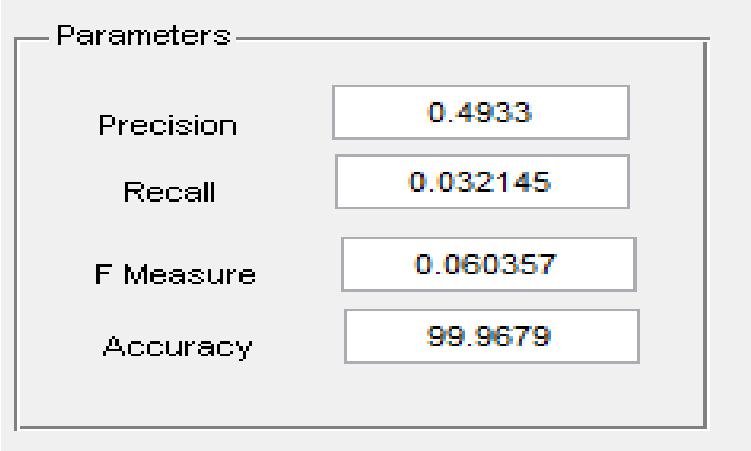

Fig. 5.12 Measured parameters for text document

Above figure illustrated the simulated parameters like precision, recall, F-measure and accuracy obtained for text document are .4933, .032145, .060357 and 99.9679 respectively.

5.4.2 EXPERIMENTAL RESULT FOR MP3 FILE

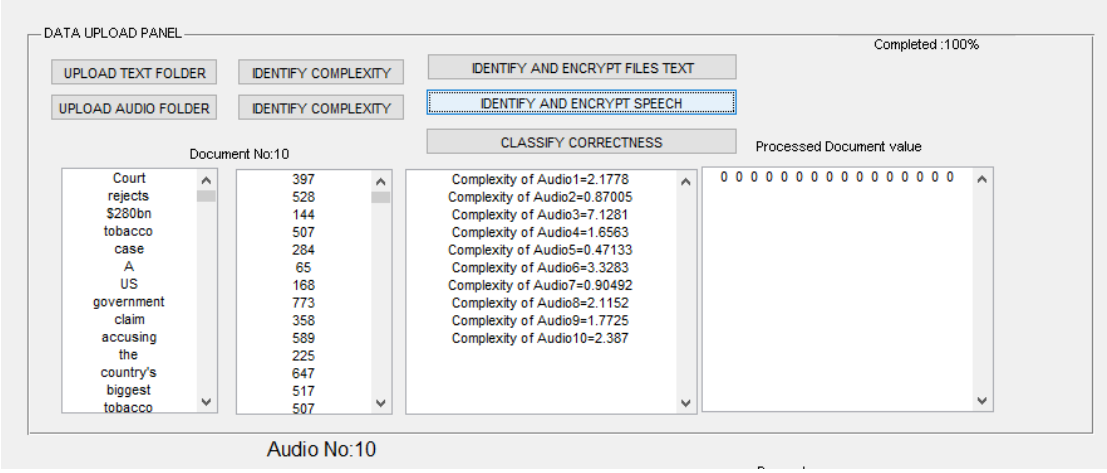

Fig. 5.13 Data uploaded panel for MP3

Firstly music file is uploaded by clicking on 'upload audio folder'. In the proposed work there are 10 music files that we have to be uploaded for the experiment. After uploading music file,

next step is to find the complexity which is calculated by clicking on identify complexity. After clicking on it, the complexity for all the uploaded document has been displayed. The next step is to click on identify and encrypt speech. The encrypted document has been displayed in the processed document value panel. Now, the last step is to calculate parameters. When we click on classify correctness all the measure parameters has been displayed in the parameters panel.

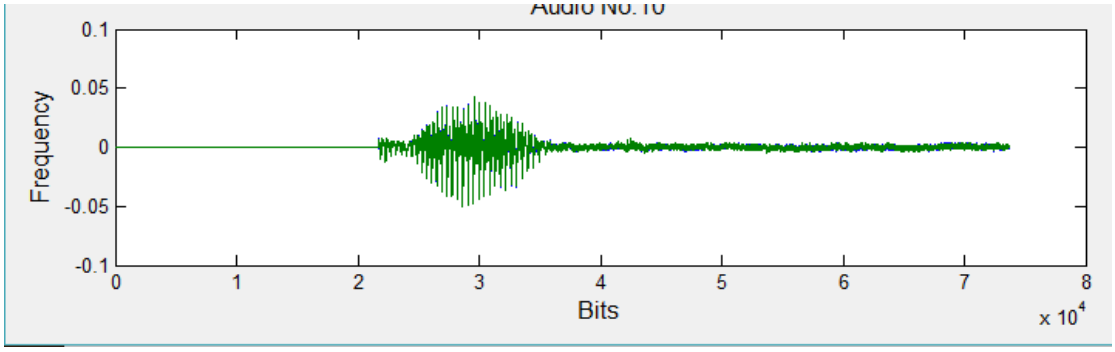

Fig 5.14 Frequency vs bits waveform of speech

The above figure shows the frequency vs number of bits for the uploaded file.

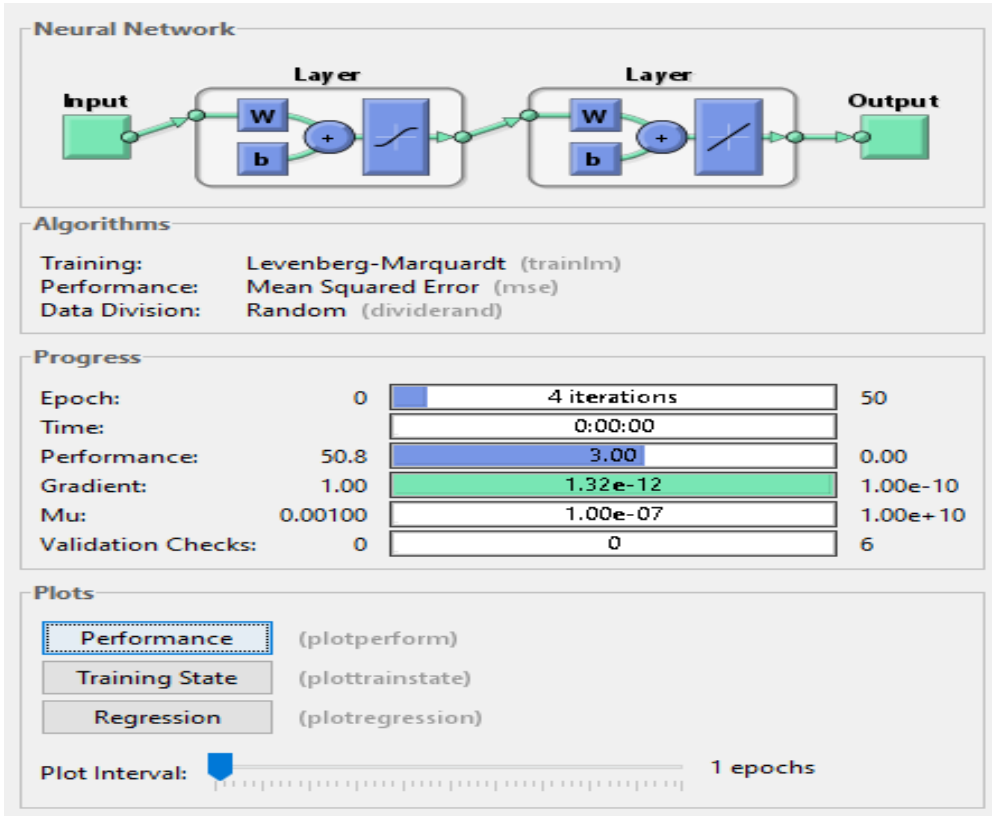

Fig. 5.15 Neural network for speech

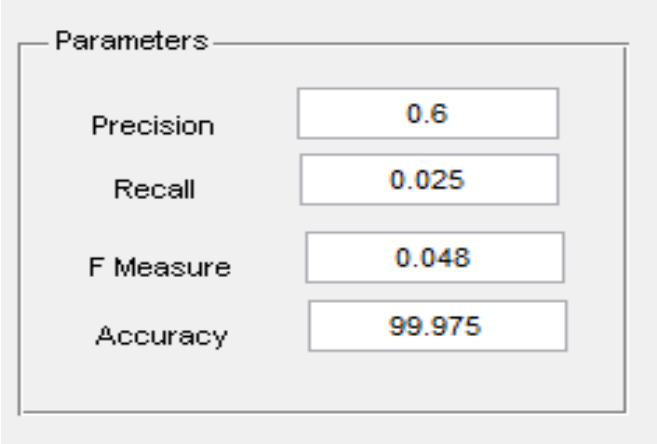

Fig. 5.16 Performance parameters of speech

Above figure illustrated the simulated parameters like precision, recall, F-measure and accuracy obtained for speech file are .6, 0.025, .048 and 99.975 respectively.

CHAPTER 6: CONCLUSION

6.1 CONCLUSION AND FUTURE SCOPE

Because of the late advancements in PC organizing innovation, distribution of computerized interactive media content through the web is vast. Be that as it may, the expanded number of advanced records, interactive media preparing apparatuses, and the overall user-friendliness of Internet access has made an extremely suitable medium for copyright extortion and wild dispersion of sight and sound substance. A real necessity now is to secure the licensed innovation of mixed media content in sight and sound systems. There is a variety of information sorts that could be portrayed as media information sorts and are regularly the components for the building pieces of summed up interactive media situations, stages, or incorporating devices. The essential sorts can be depicted as content, pictures, sound, feature and Graphic articles. Interactive media thinks that its application in different territories including, however not restricted to, notices, workmanship, training, diversion, building, prescription, math, business, logical exploration and Spatial transient applications. In this thesis hybridization of encryption calculations along with neural network has been used to find the complexity of the text and audio file. In the research different parameters like precision, recall, F-measure and accuracy has been calculated for both the files. The simulated parameters like precision, recall, F-measure and accuracy obtained for speech file are .6, 0.025, .048 and 99.975 respectively. For text document the measured parameters are .4933, .032145, .060357 and 99.9679 for precision, recall, F-measure and accuracy respectively.The scheme is based on symmetric key cryptography with the generation of random prime number. There should be a sketch to follow a number of study avenues in future for performing a comparative study of the work with other symmetric key cryptographic techniques such as RC5, RC6. Further investigation must be in new strategies to improve the efficiency of symmetric key encryption towards more efficient security-aware big data streams.

References

1. J. Xu, Get al., "A Novel Performance Evaluation and Optimization Model for Big Data System,"*2016 15th International Symposium on Parallel and Distributed Computing (ISPDC)*, Fuzhou, China, 2016, pp. 121-130.
2. R. A. Abdel, et al., "BIG-BIO: - big data hadoop-based analytic cluster framework for bioinformatics," *2017 International Conference on Informatics, Health & Technology (ICIHT)*, Riyadh, Saudi Arabia, 2017, pp. 1-9.
3. A. Al-Shomrani et al., "Policy enforcement for big data security," *2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*, Abha, Saudi Arabia, 2017, pp. 70-74.
4. K. Wang *et al*., "Wireless Big Data Computing in Smart Grid," in *IEEE Wireless Communications*, vol. 24, no. 2, pp. 58-64, 2017.
5. K. Abouelmehdi et al., "Big data emerging issues: Hadoop security and privacy,"*2016 5th International Conference on Multimedia Computing and Systems (ICMCS)*, Marrakech, Morocco, 2016, pp. 731-736.
6. C. Hu and Y. Huo, "Efficient privacy-preserving dot-product computation for mobile big data," in *IET Communications*, vol. 11, no. 5, pp. 704-712, 3 30 2017.
7. F. Luo; C et al., "Stability of Cloud-based UAV Systems Supporting Big Data Acquisition and Processing," in *IEEE Transactions on Cloud Computing* , vol.PP, no.99, pp.1-1
8. Al-Shomrani, F et al., "Policy enforcement for big data security,"*2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*, Abha, Saudi Arabia, 2017, pp. 70-74.
9. S. Alouneh, et al. "Innovative methodology for elevating big data analysis and security," *2016 2nd International Conference on Open Source Software Computing (OSSCOM)*, Beirut, 2016, pp. 1-5.
10. N. Naik, et al., "Big data security analysis approach using Computational Intelligence techniques in R for desktop users," *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, 2016, pp. 1-8.
11. N. Chaudhari and S. Srivastava, "Big data security issues and challenges," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Noida, 2016, pp. 60-64.

12. A. Dev Mishra and Y. Beer Singh, "Big data analytics for security and privacy challenges," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Noida, 2016, pp. 50-53.
13. G. A. Francia, M. Yang and M. Trifas, "Applied image processing to multimedia information security," *2009 International Conference on Image Analysis and Signal Processing*, Taizhou, 2009, pp. 104-107.
14. B. Furht, D. Socek, and A. M. Eskicioglu, "Fundamentals of Multimedia Encryption Techniques," *CRC Press*, 2004.
15. T. Lookabaugh, et al.,"Security Analysis of Selectively Encrypted MPEG-2Streams," *Multimedia Systems and Applications VI Conference*, Orlando, FL, 2004.
16. B. Furht and D. Kirovski, "Multimedia Encryption and Authentication Techniques and Applications," *Auerbach Publications*pp.91–128, 2006.
17. T. B. Maples and G. A. Spanos, "Performance study of selective encryption scheme for the security of networked real-time video," in *Proceedings of the 4th International Conference on Computer and Communications*, Las Vegas, NV,1995.
18. L. Tang," Methods for Encrypting and Decrypting MPEG Video Data Efficiently," in *Proceedings of the 4th ACM International Multimedia Conference*, Boston, MA, 2006.
19. L. Qiao and K. Nahrstedt, "A New Algorithm for MPEG Video Encryption," in Proceedings *of the 1st International Conference on Imaging Science, Systems and Technology (CISST '97)*, Las Vegas,NV, pp. 21-29,1997.
20. B. Bhargava et al., "MEPG Video Encryption Algorithms", 2002, Available:http://raidlab.cs.purdue.edu/papers/mm.ps
21. C.-P. Wu and C.-C. J. Kuo, "Fast Encryption Methods for Audio visual Data Confidentiality," *SPIE International Symposia on Information Technologies 2000*, Boston, MA, pp. 284-295, 2002.
22. D. I. G. Amalarethinam and J. S. Geetha, "Image encryption and decryption in public key cryptography based on MR," *2015 International Conference on Computing and Communications Technologies (ICCCT)*, Chennai, 2015, pp. 133-138.

23. M. Patil, V. Sahu and A. Jain, "SMS text Compression and Encryption on Android O.S," *2014 International Conference on Computer Communication and Informatics*, Coimbatore, pp. 1-6, 2014.
24. Rajput, et al., "An Improved Cryptographic Technique to Encrypt Text using Double Encryption," *International journal of Computer Applications* vol.6, pp.86-96, 2014.
25. Verma, Vinay, and Rajesh Kumar, "A Unique approach to multimedia based dynamic symmetric key cryptography," *International Journal of Computer Science and Mobile Computing* vol.3, pp.1119-1128, 2014.
26. Singh, Sombir et al., "Enhancing the security of DES algorithm using transposition cryptography techniques," *International Journal of Advanced Research in Computer Science and Software Engineering* vol.3, pp. 464-471, 2013.
27. SHAKTI et al., "Encryption Using Different Techniques: A Review," Vol. 2, No. 1, January-February-2013.
28. Goshwe and Nentawe Y, "Data encryption and decryption using RSA Algorithm in a Network Environment," *International Journal of Computer Science and Network Security (IJCSNS)vol.* 13, 2013.
29. X. Jing, et al., "Text Encryption Algorithm Based on Natural Language Processing,"*2012 Fourth International Conference on Multimedia Information Networking and Security*, Nanjing, pp. 670-672, 2012.
30. Rohan Rayarikar, "SMS Encryption using AES Algorithm on Android", *International Journal of Computer Applications* ,Vol. 50– No.19, pp 12-17, 2012.
31. M. Wangsadiredja and R. Munir, "Text and file encryption application for blackberry using cipher feedback 8-bit mode," *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, Bandung, 2011, pp. 1-6.
32. X. Wu et al., "Data Mining with Big Data," IEEE Trans. Knowledge Data Eng., vol. 26, no. 1, 2014, pp. 97–107.
33. R. Lu et al., "EPPA: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications," IEEE Trans. Parallel Distrib. Sys. vol. 23, no. 9, 2012, pp. 1621–31.