Secure Data Backup: An Enhanced Data Privacy and Security with Encryption

Phani Sridhar A [1*], Kishore T[2]
[1] Department of Computer Science, Aditya Engineering College (A), Andhra Pradesh, India
[2] Department of Computer Science, Aditya Engineering College (A), Andhra Pradesh, India

Abstract- "Secure Backup" is a web application that provides a simple to use (Graphical User Interface) GUI, which facilitates users to store data of any format in a secured manner. This system is designed in a way that users can encrypt the selected files and upload them in to a local system, a database. Security has problems like multi tenancy, information loss and leakage, simple accessibility of cloud, identity management, internal threats etc. The proposed application provides security to the data in terms of authentication and confidentiality. This application verifies the authentication of the user by traditional mechanism called password verification, One Time Password generation and CAPTCHA verification. User need to be validated by providing details like One Time Password within 10 seconds, else the session will be expired. Other authentication mechanism like CAPTCHA that differentiates humans and machine inputs. The files that are being uploaded will be stored in the local system for backup. Whenever the user needs a backup the data is processed and can be downloaded by the user. The files uploaded can be encrypted using triple DES algorithm, a symmetric block cipher. On the basis of security requirements, a security key is added during the time of encryption and the same key given while decrypting the respective file. The prime conviction is to assure security to the data by storing the data in an encrypted format in to a database.

Keywords- Backup, Captha, OTP, DES, Cipher.

I.INTRODUCTION

Data security refers to the protection of data from unauthorized access, use, change, disclosure and destruction and includes network security, physical security, and file security Data storage refers to holding your data files in a secure location that you can readily and easily access. Data backup, in contrast, refers to saving additional copies of your data in separate physical or virtual locations from data files in storage. Your data is the basis of your research. If you lose your data, recovery could be slow, costly, or impossible. It is important that you secure, store, and backup your data on a regular basis. Securing your data will help to prevent:

- Accidental or malicious damage/modification to data
- Theft of valuable data
- Breach of confidentiality agreements and privacy laws
- Premature release of data, which can void intellectual property claims
- Release before data have been checked for accuracy and authenticity

Keeping reliable backups is an important part of data management. Regular backups protect against the risk of damage or loss due to hardware failure, software or media faults, viruses or hacking, power failure, or even human errors. Security of Data is the most important task in today's world. Over the years various encryption schemes have been developed in order to protect the database from various attacks by the intruders. This paper discuss the importance of database encryption and makes an in depth review of various database encryption techniques and compare them on basis of their merits and demerits.

Database security using encryption

Security needs to be considered for all copies of your data, including your working data set, backup copies and archived copies.

- Network security
  - Keep confidential data off the Internet
  - Put sensitive materials on computers not connected to the internet
- Physical Security
  - Restrict access to buildings and rooms where computers or media are kept
  - Only let trusted individuals troubleshoot computer problems
- Computer Systems & Files
  - Keep virus protection up to date

- Don't sent confidential data via e-mail or FTP - use encryption, if you must send data
- Use good passwords on files and computers

One of the most important data management tasks is keeping backups of your data. There is a real risk of losing data through hard drive failure or accidental deletion.

- Remember to use the Backup 3-2-1 Rule
  - 3 copies of your data - 2 copies are not enough
  - 2 different formats - i.e. hard drive tape backup or DVD (short term)+flash drive
  - 1 off-site backup - have 2 physical backups and one in the cloud
- Backup options
  - Hard drives - personal or work computer
  - Departmental or institution server
  - External hard drives
  - Tape backups
  - Discipline-specific repositories
  - University archives
  - Cloud storage

The existing systems that we use for storage purpose provided websites may provide storage of files and documents but they do not provide proper security for the files and data. There may be a chance of intrusion, hacking and corruption. Though the storage is provided efficiently, there could be a one concrete systemized approach, which provides authentication and security to serve this purpose. Our purposed system strives at this very solution.

Moreover, there have been rarely any attempts done to provide proper security. That is, there is no proper authentication to access the account, anyone who knows can access the account which is a threat to the data. Furthermore, there is no encryption provide to store the files that are uploaded securely.

Some of the Disadvantages are as follow:

- High chance of missing out on stored data.
- There's no proper security.
- There is a chance of hacking, loss of data, data modification.
- No proper validation of users credentials.
- There is no provision for the files to be stored securely.

Cloud computing is a flexible, cost- effective and proven delivery platform for providing business or consumer IT services over the Internet. Cloud computing supports distributed service oriented architecture, multi-users and multi-domain administrative infrastructure, it is more prone to security threats and vulnerabilities. At present, a major concern in cloud adoption is its security and Privacy. Intrusion prospects within cloud environment are many and with high gains. Security and Privacy issues are of more concern to cloud service providers who are actually hosting the services. In most cases, the provider must guarantee that their infrastructure is secure and clients' data and applications are safe by implementing security policies and mechanisms. While the cloud customer must ensure that provider has taken proper security measures to protect their information. The issues are organized into several general categories: trust, architecture, identity management, software isolation, data protection, availability Reliability, Ownership, Data Backup, Data Portability and Conversion, Multiplatform Support and Intellectual Property.

II. RELATED WORK

For protecting the data in cloud includes some challenges, such as the data is highly broadcasted in network and increasing the complexity of the management by introducing the vulnerability status. In decentralized system it creates the storage system is shared by the users in various security domains with different policies. In the extended time gives more windows for attackers. This raises the less compatibility issues and the data migration process including encryption algorithm moving. Data at a primary storage system is encrypted and remote copied to a secondary storage system. A Remote Copy Configuration Information (RCCI) is created that identifies the encryption mechanism, keys, data source volume, and target volume for the remote copy. The RCCI is backed up on a trusted computer system. In one embodiment, the secondary storage system is an off-site data storage system managed by a third party. Upon detection of a failure in the primary storage system, the encrypted data and RCCI are transferred to a tertiary server, which is optionally created upon detection of the failure, and operations of the failed primary server are resumed by the tertiary server. In one embodiment, the failure is detected by loss of a heart beat signal transmitted from the primary storage system to a management server that initiates the transfers to the tertiary server.

Other work proposes a novel model of secure location based query search implementation with Path server to identity the route and spatial system or location based server identifies the spatial query results. Even though various models proposed by various researchers from the years of research, every model has its own advantages and disadvantages. Initially user makes the spatial query request, in turn it forwards to location based server and followed by path or route server. It computes the path and forwards to LBS and gets the respective spatial results and forwards to requested user. Our proposed model gives efficient results than traditional models.

We propose a novel model of secure location based query search implementation with Path server to identity the route and spatial system or location based server identifies the spatial query results. Even though various models proposed by various researchers from the years of research, every model has its own advantages and disadvantages. Initially user makes the spatial query request, in turn it forwards to location based server and followed by path or route server. It computes the path and forwards to LBS and gets the respective spatial results and forwards to requested user. Our proposed model gives efficient results than traditional models [1].

Grid Computing represents the latest and most exciting technology to evolve from the familiar realm of parallel, peer-to-peer and client-server models. However, there has been limited investigation into the impact of this emerging technology in medical imaging and informatics. In particular, PACS technology, an established clinical image repository system, while having matured significantly during the past ten years, still remains weak in the area of clinical image data backup. Current solutions are expensive or time consuming and the technology is far from fool proof. Many large-scale PACS archive systems still encounter downtime for hours or days, which has the critical effect of crippling daily clinical operations. In this paper, a review of current backup solutions will be presented along with a brief introduction to grid technology. Finally, research and development utilizing the grid architecture for the recovery of clinical image data, in particular, PACS image data, will be presented.

The focus of this paper is centric on applying a grid computing architecture to a DICOM environment since DICOM has become the standard for clinical image data and PACS utilizes this standard. A federation of PACS can be created allowing a failed PACS archive to recover its image data from others in the federation in a seamless fashion. The design reflects the five-layer architecture of grid computing: Fabric, Resource, Connectivity, Collective, and Application Layers. The test bed Data Grid is composed of one research laboratory and two clinical sites. The Globus 3.0 Toolkit (Co-developed by the Argonne National Laboratory and Information Sciences Institute, USC) for developing the core and user level middleware is utilized to achieve grid connectivity.

Thesuccessful implementation and evaluation of utilizing data grid architecture for clinical PACS

data backup and recovery will provide an understanding of the methodology for using Data Grid in clinical

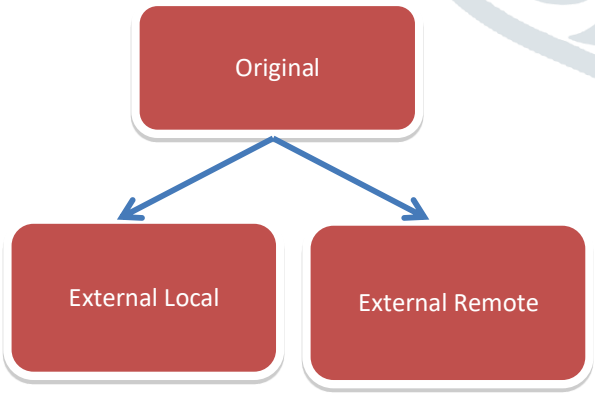

image data backup for PACS, as well as establishment of benchmarks for performance from future grid technology improvements. In addition, the test bed can serve as a road map for expanded research into large enterprise and federation level data grids to guarantee CA (Continuous Availability, 99.999% up time) in a variety of medical data archiving, retrieval, and distribution scenarios [2].

III. METHODOLOGY

In traditional working mechanisms and data storages the majority of evidences are stored in local/ remote spaces. But the data storage and back up streams could be much better if the huge storage is needed for the access of localized or terminal applications. Consider your data storage and backup strategy before you start collecting and creating your data. Your strategy should be able to accommodate the amount of data that you

anticipate collecting and be stable for the length of time that you anticipate keeping your data. Let us assume the fraudulent detection mechanism for credit card based system approach uses the model of security rules that are relevant for enhancing the measures of security. Here are two types of options for encryption of backup. One is host-based and the other is appliance-based [3].

Cyber-attacks are constantly evolving, so security specialists must stay busy in the lab concocting new schemes to keep them at bay. Expert observers are hopeful that a new method called Honey Encryption will deter hackers by serving up fake data for every incorrect guess of the key code. This unique approach not only slows attackers down, but potentially buries the correct key in a haystack of false hopes. Then there are emerging methods like quantum key distribution, which shares keys embedded in photons over fiber optic, that might have viability now and many years into the future as well. Whether it's protecting your email communications or stored data, some type of encryption should be included in your line up of security tools. Successful attacks on victims like Target show that it's not 100 percent bullet proof, but without it, you're offering up convenient access to your data [4].

Algorithm

This program utilizes:

- a user inputs password
- a user inputs secret message (preferably a U.S. state)
- a hardcoded dictionary of secret messages (in this program, these messages are U.S. states)
- a dictionary containing manipulated passwords (Sweet words) in addition to the real password
- a dictionary containing seeds (Seed generator)
    - seeds are simply pointers that point to the secret message
- the encryption algorithm: c = sk ^ sm
    - the cipher_text = seed value of the key XOR seed value of the message
- the decrpytion algorithm: m = sk ^ c
    - the message = seed value of the key XOR cipher_text
- a try/catch block to search for passwords that do not exist in the dictionary of sweetwords
- a query to prompt the user for another attempt.

In host-based encryption of backup data, encryption takes place on the host itself. For this type of cryption, most enterprises won't need to buy an additional solution because most backup software solutions support encryption (including EMC Networker, EMC Avamar, Symantec Netbackup, IBM TSM, and Commvault Simpana. You can opt for encryption on the backup client side. In this option, you encrypt data on the backup client, and send the encrypted data on the network and then to the backup device.

In appliance-based encryption, the encryption of backup is handled by an appliance (which sits in the storage network). In other words, it's directly in the data path, and encrypts data at wire speed—without the clients and backup server being aware of this appliance. For example, NetApp Data Fort is considered an industry-trusted appliance for such encryption of backup data. This type of encryption of backup doesn't cause any overheads, since the encryption takes place at wire speed and has robust key management features. Enterprises can distribute these keys to various trusted employees in the form of smart cards, and all the smart cards will be needed for key regeneration. In addition to encrypting backup data, these appliances also give you the advantage of writing encrypted data to the SAN. The hard part is that these appliance-based encryption solutions are much costlier than the backup software solution which a company may have already invested in.

To wrap up, both types of backup encryption have their pros and cons, so you need to choose which type of encryption suits you best—backup software applied or hardware appliance applied[5].

Some works have the goal is to build up a repository to facilitate the data integration and sharing across cloud along with preservation of data confidentiality. For this it is necessary to be using an encryption technique to provide data security on data storage.

1. To develop a system that will Provide Security and Privacy to Cloud Storage.
2. To Establish an Encryption Based System for protecting Sensitive data on the cloud and Structure how owner and storage Service Provider to operate on encrypted Data.
3. To Create a System where the user store its data on the cloud the data is sent and stored on the cloud in encrypted form As in normal cases in cloud computing when a user login to the cloud and they store data on cloud storage device the data stored on the server cloud is not much secure as it can be readable to anyone which have permission to access and Leaving data vulnerable.

4. To Develop a retrieval System in which the data is retrieved by the user in encrypted form and is decrypted by the user at its own site using a public and private key encryption both the keys working at the user level.

Proposed technique emphasizes on improving classical encryption techniques by integrating substitution cipher and transposition cipher. Both substitution and transposition techniques have used alphabet for cipher text. In the proposed algorithm, initially the plain text is converted into corresponding ASCII code value of each alphabet. In classical encryption technique, the key value ranges between 1 to 26 or key may be string (combination alphabets). But in proposed algorithm, key value range between 1 to 256. This algorithm is used in order to encrypt the data of the user in the clouds. Since the user has no control over the data after his session is logged out, the encryption key acts as the primary authentication for the user. Proposed algorithm is described below.

Encryption Algorithm

Step 1 Count the No. of character (N) in the plain text without space.

Step 2 Convert the plain text into equivalent ASCII code. And form a square matrix (S X S >=N).

Step 3 Apply the converted ASCII code value from left to right in the matrix. Divide matrix into three part namely upper, diagonal and lower matrix.

Step 4 Read the value from right to left in each matrix.

Step 5 Each matrix use three different key K=K1, K2, K3 for encryption. Do the encryption.

Step 6 Apply the encrypted value into the matrix in the same order of upper, diagonal and lower.

Step 7 Read the message by column by column. Here the order in the columns read from the matrix is the key K4.

Step 8 Convert the ASCII code into character value.

Overview of our approach

Our goal is to build up a repository to facilitate the data integration and sharing across cloud along with preservation of data confidentiality. For this we will be using an encryption technique to provide data security on data storage. The possible modules that are involved in the judgment of data security are mentioned in the following sub contents.

- **Home**
  - It navigates back to the home page.
- **Encrypt**
  - The files can be encrypted using Triple DES and a key will be given or generated by the user itself.
- **Decrypt**
  - The encrypted data will be decrypted back to the normal text by using the same key given at the time of encryption.
- **Upload**
  - The encrypted files can be uploaded into the local system.
- **Download**
  The decrypted files can be downloaded by the user directly.

Delete Files

- According to the user requirements, the unnecessary files can be deleted.
- **View Files**All the files that are uploaded will be displayed.
- **Logout** The user will be logged out of the account

Input Size (bytes)	DES	3DES	AES	BF
20,527	24	72	39	19
36,002	48	123	74	35
45,911	57	158	94	46
59,852	74	202	125	58
69,545	83	243	143	67
137,325	160	461	285	136
158,959	190	543	324	158
166,364	198	569	355	162
191,383	227	655	378	176
232,398	276	799	460	219

Average Time	134	383	228	108
Bytes/sec	835	292	491	1,036

Table 1: Comparison results using Crypto++

Algorithm	Megabytes(2^20 bytes) Processed	Time Taken	MB/Second
Blowfish	256	3.976	64.386
Rijndael (128-bit key)	256	4.196	61.010
Rijndael (192-bit key)	256	4.817	53.145
Rijndael (256-bit key)	256	5.308	48.229
Rijndael (128) CTR	256	4.436	57.710
Rijndael (128) OFB	256	4.837	52.925
Rijndael (128) CFB	256	5.378	47.601
Rijndael (128) CBC	256	4.617	55.447
DES	128	5.998	21.340
(3DES)DES-XEX3	128	6.159	20.783
(3DES)DES-EDE3	64	6.499	9.848

Table 2: Comparative execution times (in seconds) of encryption algorithms

IV. RESULTS AND DISCUSSIONS

This section describes the techniques and execution choices made to evaluate the performance of the data backup .In addition to that, this section will discuss the methodology related parameters like: system parameters, experiment factor(s), and experiment initial settings.

System Parameters

The experiments are conducted using 64bit Mac processor. The simulation program is compiled using the default settings in .Visual studio 2014 Mac applications. The experiments will be performed couple of times to assure that the results are consistent and are valid to compare the different input options

Experiment Factors

In order to evaluate the setup designed, the parameters that the process must be tested for must be determined. Since the security features of each algorithm as their strength against multiple attacks is already known. The chosen factor here to determine the performance is the algorithm's speed to encrypt/decrypt data blocks of various sizes.

Simulation Procedure

By considering different sizes of data blocks (0.5MB to 20MB) the algorithms were evaluated in terms of the time required to encrypt and decrypt the data block. All the implementations were exact to make sure that the results will be relatively fair and accurate. The Simulation program (shown below in Fig. 7) accepts three inputs: Algorithm, Cipher Mode and data block size. After a successful execution, the data generated, encrypted, and decrypted are shown. Notice that most of the characters cannot appear since they do not have character representation. Another comparison is made after the successful encryption/decryption process to make sure that all the data are processed in the right way by comparing the generated data (the original data blocks) and the decrypted data block generated from the process. The complete data security and backup problems are supported and resolved by the Onetime password detail and secure Captha mechanism. The internal authenticate and authorize schedules are run by using the triple DES structures providing and enhancing the better securable , creative and strong replica of master and slave authorities.

V. Conclusion and Future Scope

It's often said that a backup protects your data while an encrypted backup secures and protects your data, your job and your company. All things being equal most would choose to protect and secure backup data. The question is how and the answer is Secure Backup using encryption. Backup encryption can be accomplished in many ways. For example, you could purchase an in line backup appliance(s) which sits in the backup stream purportedly transparent to the backup software or perhaps encryption key software which

is integrated with your tape drives. Both are viable encryption methods, no doubt. However, these offering often come with hefty price tags and add another moving part to your backup infrastructure. Who needs the added cost and complexity? Backup encryption is one of the many OSB enterprise-class features. Secure Backup provides centralized tape backup management for your entire IT environment easily scaling from 10s to hundreds of servers and 1000s of backup tapes. As with all of OSB's advanced functionality, backup encryption is included in the valuable cost per tape drive licensing fee. Backup encryption requirements vary by environment, host or backup. Secure Backup delivers the depth to meet your simplest to most complex backup encryption requirements:

- Two host-based backup encryption options:
  - o OSB native backup encryption
  - o RMAN backup encryption
- Tape drive encryption support for LTO-4, LTO-5, T10000B and T10000C drives
- Policy-based encryption key management
  - o Transparently or passphrase generated encryption keys
  - o Rekey frequency policies for automated key regeneration
- Multi-level backup encryption policies and configuration options:
  - o Global and host encryption settings
  - o Backup encryption defined as part of a recurring schedule
  - o One-off, adhoc, backup encryption for those every once in a while encryption needs

Oracle Secure Backup delivers data protection for the enterprise at over 75% less cost than comparable products. Unprecedented in the backup industry, OSB offers low-cost, single component (per tape drive) licensing making affordable, secure data protection within reach of both small and large IT organizations. With Oracle Secure Backup, you can reduce IT costs without sacrificing functionality.

` Secure data protection from server to tape is crucial for local and offsite storage of mission critical data. The portability and long-shelf life of tape is ideally suited for long-term, offsite backup storage. This advantage has a double-edge in that once backup tapes leave the safety of a secure data center; they are exposed to external variables and potentially 3rd party transportation or storage vendors. Oracle Secure Backup provides policy-based backup encryption securing the backup data on tape whether those tapes are onsite, offsite or lost. What would it cost you and your business if mission-critical backup data fell into the wrong hands? When you consider the financial costs along with potential reputation damage, can you afford not to encrypt important backup data. Avoid unnecessary costs by encrypting, protecting and securing your backup data with Oracle Secure Backup.

VI. References

[1] Gabriel Babutonde Iwaokun, "***Encryption and Tokenization-Based System for Credit Card Information Security",*** Research Gate, August, 2018.

[2] Abhishek Vichare**, "*Data security using authenticated encryption and decryption algorithm for Android phones",*** IEEE publisher, 2018.

[3] Pradnyesh Bhisikar, "***Security in Data Storage and Transmission in Cloud Computing***", IJARCSSE, Vol: **03**, Issue: **03,** 2013.

[4] Brent J Liu, "***Utilizing data grid architecture for the backup and recovery of clinical image data***", Research Gate, March 2013.

[5] Brandford, "***Common Encryption Algorithms and the Unbreakables of the Future",*** Article, March, 2018.

[6] Zaid Kartit, "***Applying Encryption Algorithm for Data Security and Privacy in Cloud Computing***", Research Gate, January, 2016.

Author's Profile

Phani Sridhar Addepalli completed his B.Tech degree from Jawaharlal Nehru Technological University Hyderabad affiliated college, M.Tech degreefrom Gitam (Deemed to be University). Currently pursuing Ph.D in Gitam (Deemed to be University). He has over 10 years of experience in various reputed colleges across the state of Andhra Pradesh. Currently working as Associate Professor in Aditya Engineering College an Autonomous Institution in the department of Computer Science Engineering. His research interests include Network Security, Internet of Things, IoT Gateways.

Kishore Teppala completed his B.Tech degree from Jawaharlal Nehru Technological University Kakinada affiliated college, M.Tech degree from Jawaharlal Nehru Technological University Kakinada affiliated college. He has over 5 years of experience in the veracity of engineering colleges in Andhra Pradesh. At present he is working as Assistant Professor in Aditya Engineering College an Autonomous Institution, CSE department. His research areas of interest include Data security and Privacy, Mining and Fraud detection.