PROVIDING SECURITY FOR BIG DATA WITH CRYPTOGRAPHY USING HYBRID CLOUD

G.Sreelatha[1], Dr.A.Kanaka Durga[2]
[1]Asst.Professor, [2]Professor
[1,2]IT Department, SCETW

Abstract

In today's data-centric world, big-data processing and analytics have become critical to most enterprise and government applications. Thus, there is a need for an appropriate big data infrastructure that supports storage and processing on a massive scale Cloud computing, which is a new business model, is considered as one of most attractive solutions for big data, and provides the benefits of low cost through sharing of computing and storage resources. Also, Cloud computing has become the tool of choice for big data processing and analytics due to its reduced cost, broad network access, elasticity, resource pooling, and measured service. However, the increasing concerns of the privacy of data which is stored in public cloud have slowed down the adoption of cloud computing for big data because very sensitive information may be contained among the big data or the data owner themselves do not want any other people to watch and observe their data. Since the data volume is huge and mobile devices are widely used, the traditional cryptographic methods are not suitable for big data.

In this paper we are implementing the modified approach for the image and text encryption. Also we focuses on state-of-the-art provably secure cryptographic techniques for protecting big data applications. We do not focus on more established, and commonly available cryptographic solutions. The goal is to inform practitioners of new techniques to consider as they develop new big data solutions rather than to summarize the current best practice for securing data. With the image encryption technique we have used encryption for text data storage. In this paper we are going to use hybrid cloud mechanism to store the data. We store small amount of information about data on the private cloud as a reference and other data is store on the public cloud.

Keywords: Cloud computing, Hybrid cloud, Encryption, Decryption, Cryptography, Big data.

I. INTRODUCTION

The amount of data is being produced directly proportional with the rapid development of electronic technology and communication, which makes it hard to cost effectively manage and store these big data. We know the cloud computing is getting popular nowadays because it provides flexibility for data storage.

Cloud support ubiquitous computing which means the user can access their data from anywhere, anytime. It also supports large size of storage where user can store their data efficiently.

While providing the services to the user cloud computing proposes two types of cloud which are private cloud and public cloud. Private cloud is considered as secure cloud because the information of user can only access by the user itself and no other person can access it but the problem with private cloud is it is expensive it is not suitable to store large data because of expensiveness on the other hand public cloud is cheaper than the private cloud but public cloud is not as much secure like private cloud. The information which is stored
on the public cloud can be accessible to the cloud service providers. They can access the user's

information and can be used for their benefits like they may use that information for advertising etc.

Today large amount of data is being generated from different sources like social networking sites, online shopping and this data can contain some sensitive information. Private cloud is safe place for storing this sensitive data but the private cloud is having small size and the cost for data storage on the private cloud is more so this is not efficient for the user. There is another option for this data storage which is public cloud but as we know the public cloud does not provide security like private cloud it is also not efficient. So can we use hybrid cloud mechanism for securing our valuable data?

Hybrid cloud is an infrastructure which combines both private and public cloud. It stores the user information on the public cloud by specifying their references on the private cloud. Hybrid cloud stores little information about data on the private cloud and most of the data is store on the public cloud. While accessing the data from the public cloud there is need of communication between private and public cloud.

In our paper we are proposing a solution for security of the data using hybrid cloud which secure our image as well as text data.

II. RELATED WORK

Hybrid Cloud:

The National Institute of Standards and Technology defines hybrid cloud as "a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models." The public and private clouds in a hybrid cloud arrangement are distinct and independent elements. This allows organizations to store protected or privileged data on a private cloud, while retaining the ability to leverage computational resources from the public cloud to run applications that rely on this data. This keeps data exposure to a bare minimum because they're not storing sensitive data long-term on the public cloud component.

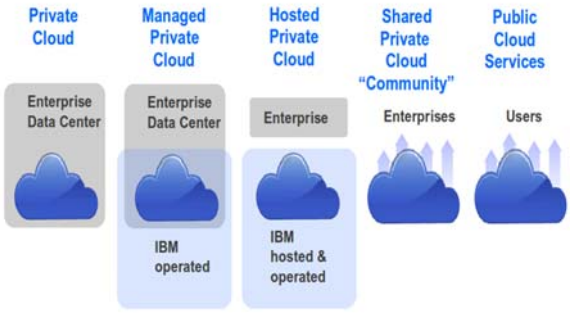

Fig1. Hybrid Cloud

Often, hybrid cloud refers to a combination of a public cloud service and a private cloud on-premises; however, hybrid clouds could also consist of two public clouds provided by different providers or even a combination of a cloud and traditional IT. Any hybrid cloud setup has some challenges that need to be considered during the planning and design phase:

The most obvious challenge is network connectivity, especially if remote cloud services like a public cloud or a hosted private cloud are involved. Not only must bandwidth, latency, reliability and associated cost considerations be taken into account, but also the logical network topology must be carefully designed (networks, routing, firewalls). Another huge challenge is the manageability of different cloud services.

When different cloud services are used, every service provider will have its own management and provisioning environment. Those environments can be considered completely independent from each other. By having instances in different cloud services, there is no complete picture available showing the number of totally deployed instances and their statuses. An orchestration layer can be a possible solution for this problem. This layer provides a single interface for all cloud-related tasks. The orchestration layer itself communicates with the different cloud services through application programming interfaces (APIs). The big advantage of an orchestration layer is the ability to track and control activities on a central point to maintain the big picture.

Hybrid cloud adoption can be an effective strategy for a wide variety of businesses that have a tighter focus on security or unique physical presence demands. Although there is greatly minimized risk in a hybrid cloud model, allowing access from a public cloud has the *remote potential* of being insecure, or being

the conduit through which data can be harvested. This, however, is true of almost any public network communication. And while the upfront cost of server hardware for the private component of the hybrid cloud is high, the control that IT departments can wield over hardware selection and system design for the private component offers an invaluable way of properly tailoring resources to the needs of the organization. Assembling a private cloud to handle a standard workload, with burst compute offloaded to the public cloud, can be a long-term budget-friendly arrangement.

Ultimately, hybrid cloud allows organizations to leverage the capabilities of public cloud platform providers without offloading the entirety of their data to a third-party data center. This provides a great deal of flexibility in computing tasks, while keeping the most vital components within the company firewall.

Security Goals:

Having defined the adversary we want to protect against, we need to describe the security goals. The three most fundamental security goals are confidentiality, integrity, and availability, collectively known as the CIA triad. ***Confidentiality***: Confidentiality is the goal of keeping all sensitive data secret from an adversary. More formally, traditional definitions of confidentiality guarantee that an adversary should learn no information about the sensitive data, other than its length. Confidentiality is critical in big data applications to guarantee that sensitive data is not revealed to the wrong parties. ***Integrity***: Integrity is the goal that any unauthorized modification of data should be detectable. That is, a malicious adversary should not be able to modify such data without leaving a trace. This is very important to help guarantee the veracity of data collected in big data applications.

Availability: Availability is the goal of always being able to access one's data and computing resources. In particular, an adversary should not be able to disable access to critical data or resources. This is a very important security goal in big data processing, as the sheer volume and velocity of the data make guaranteeing constant access a difficult task. However, in today's big data systems, availability is typically guaranteed via non-cryptographic means such as replication, and we will not discuss it further in this chapter.

Cryptography in Big Data:

We now give a very brief overview of the basic cryptographic tools used to ensure the aforementioned security goals in simple applications such as data in transit.

Encryption: The main tool for guaranteeing confidentiality of data is data encryption. Encryption takes a piece of data, commonly called the plaintext, together with a cryptographic key and produces a scrambled version of the data called the cipher text. Using the key it is possible to decrypt the data to recover the plaintext, but without the key the cipher text hides all information about the original data, other than its length. This security property, commonly known as semantic security guarantees that, without the key, an adversary cannot learn any (potentially sensitive) property of the underlying data even if he has a lot of insight as to what the data may be.

- KeyGen - a key generation algorithm that generates the necessary cryptographic keys,
- Enc(k, p) = c - an encryption algorithm that uses a key k to scramble the plaintext p into cipher text c,
- Dec(k, c) = p - a decryption algorithm that uses the key k to recover the plaintext p from the cipher text c.

Encryption schemes come in two flavors: secret-key encryption, described above, and public-key encryption. In secret-key encryption the same key is used for encrypting and decrypting data. In public-key encryption, KeyGen produces two keys: a public key and a secret key. The public key is used to encrypt the data, but cannot be used to decrypt the data and thus can be made public. The secret key, which must be kept private, is used to decrypt the data

III. SYSTEM DESIGN

In our paper we propose a novel solution for securing the image and text data by using hybrid cloud infrastructure. Figure: 1 shows the architecture of the hybrid cloud. Hybrid cloud is consisting of both private and public cloud. As we know the private cloud has less space but it is secure cloud and public cloud has large capacity to store data but it has less secure than the private cloud.

Hybrid cloud takes advantage of the both private and public cloud. It stores the sensitive information on the private cloud and non-sensitive information on the public cloud. There is need of communication between the private and public cloud.

In our paper we are propose a system in which we are going to use this hybrid cloud infrastructure for storage of big data. As the large amount of data cannot store on the private cloud because it has less capacity and the cost associated with storage is high therefore it is not efficient solution for data storage. So we store the small amount of data on the private cloud as a reference and the large amount of data is to be store on the public cloud. This reduces the cost of data storage and load of the private as most of the computation on the data is carried out on the public cloud.

Whenever the user wants to store their data on the cloud they will give their data to private cloud. Then the original data will be transformed into encrypted form after that private cloud will store the small amount of information about that data on the private cloud as a reference in the encrypted form and other information will be stored on the public cloud which is also encrypted for security purpose. The private cloud stores the keys to access the data from the public cloud and those keys are also encrypted.

Whenever the user wants to access their data they will send the request to the private cloud. Then using the keys which are stored on the private cloud the users data will be fetch from the public cloud. After that data will be decrypted and then this data will be delivered to the user.

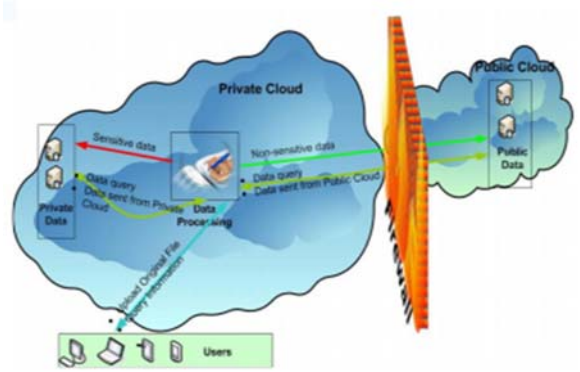

Fig.2.System Design.

IV. ALGORITHMS

To implement the system for providing security to data here we use two algorithm first is DES (Data Encryption Standard) used for text encryption and we propose second algorithm for image encryption and decryption which is based on encryption of block of data. In this algorithm the at first the width and height of image is obtained after that calculate the co-ordinates of each pixel of the image, after that image will be divided into number of blocks, once the image is divided into blocks are shuffled, then the keys which are required to decrypt the image are store on the private cloud of the user and all other blocks get store on the public cloud. Whenever user wants data from public cloud they send request to the private cloud, by using the reference of the data which are stored on the private cloud all the blocks of images and text data are fetched from the public cloud.

A. Image Encryption Algorithm

Input: Original image

Output: Encrypted block of image

Step1: Get the pixels of the original image.

Step2: Obtain the height and width of the image.

Step3: Obtain the encrypted pixel values by splitting image into two or more parts.

Step4: Convert these pixels into final encrypted image.

Step5: Divide the encrypted image into small blocks.

Step6: Shuffle the encrypted image blocks.

Step7: Encrypt the keys and store it on the private cloud.

Step8: Store the encrypted data on the public cloud.

B. Image Decryption Algorithm

Input: Image name (block of encrypted image).

Output: Original Image.

Step1: Obtain the keys from the private cloud.

Step2: Get the encrypted image from small blocks.

Step3: Get the original pixels of the image.

Step4: Get the original image and send it to the user.

V. IMPLEMENTATION

We assume the following image for encryption This clearly shows the image converted to pixels and then to encrypted image and making our data secure.

Fig3. Original Input Image

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	69	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Fig4. Encrypted image into small blocks in pixel form.

1	54	3	4	95	6	7	8	9	10
11	64	13	94	15	16	17	18	19	20
54	74	23	16	17	18	19	28	29	30
49	84	33	26	27	28	29	38	39	40
34	94	43	36	37	38	39	48	49	50
42	54	22	46	47	48	49	27	28	29
87	62	63	56	57	58	59	69	69	70
42	72	73	74	75	96	77	78	79	80
22	82	83	84	85	86	87	88	89	90
99	92	93	14	95	76	97	98	99	100

*Fig5.*Shuffeled Image pixel

Fig5.Encrypted Image

The same way we can gain our decrypted image at user end after applying decryption algorithm.

CONCLUSION

In this system we implement the infrastructure for big data security. In this system we are using hybrid cloud framework for security purpose. This system will help user to secure their data and easy retrieval whenever needed. Future work of this system comprises of more effective techniques for encryption and decryption of data. We can also implement the system with functionalities in which user will get mail notifications automatically whenever they login into their account. We are work on filtering posted audio and video messages. We can also reduce the communication time between the private and public cloud. We can also provide security to audio and video data.

REFERENCES

1. F. C. Lau, C. Wu L, Zhang, C. Guo, ,Z. Li, and "Moving big data to the cloud: An online cost-minimizing approach," JOURNAL ON SELECTED AREA IN COMMUNICATIONS, 2013.
2. D. Chen and H. Zhao, " security and privacy protection issues in cloud computing, "inComputer Scienceand Electronic Engineering(ICCSEE), 2012 International Conference on, 2012.
3. S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in INFOCOM,2010 Proceedings IEEE, 2010.
4. C. Gentry, S. Halevi, C. Peikert, and N. P. Smart, "Field switching in BGV-style homomorphic encryption," Journal of Computer Security, 2013.
5. S. Halevi and V. Shoup. (2014) HElib - an implementation of homomorphic encryption. [Online]. Available: https://github.com/shaih/ HElib "Algorithms in HElib," IACR Cryptology ePrint Archive, 2014.
6. A. Sahai and B. Waters, "Fuzzy identity-based encryption," in EUROCRYPT,2005.
7. D. Boneh, A. Sahai, and B. Waters, "Functional encryption: Definitions and challenges," in TCC, 2011.

8. S. Arora and S. Safra, "Probabilistic checking of proofs: A new characterization of NP," J. ACM, 1998.
9. S. Micali, "CS proofs," 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, 1994.
10. B. Parno, J. Howell, C. Gentry, and M. Raykova, "Pinocchio: Nearly practical verifiable computation," in Proceedings of the 2013 IEEE Symposium on Security and Privacy, 2013.
11. E. Ben-Sasson, A. Chiesa, E. Tromer, and M. Virza, "Succinct noninteractive arguments for a von Neumann architecture," Cryptology ePrint Archive, Report 2013/879, 2013. J. Li, C. Jia, J. Li, and Z. Liu, "Framework for outsourcing and sharing searchable encrypted data on hybrid cloud," in Intelligent Networking and Collaborative Systems , 2012 4th International Conference on. Springer, 2012.
12. T. Jung, X.-Y. Li, Z. M. Wan, and Wan, "Privacy preserving cloud data access with multi-authorities," in IEEE INFOCOM, 2013.
13. K.-W. Wong, Y. Wang, G. Chen, and X. Liao, "A new chaos-based fast image encryption algorithm," Applied soft computing, 2011