SECURING DATA STORAGE ON AMAZON AWS S3

[1]Amanpreet Sharma, [2]Rajnish Kansal
[1]Computer Science and Engineering, Asra College of Engineering and Technology, Punjab, India

Abstract: Cloud Computing is a phenomenon technology that has enabled many other technologies like Internet of Things, Artificial Intelligence as Cloud has solved many problems like storage over the internet and sensors connectivity with the applications which can be controlled from anywhere. This paper describes the introduction to cloud computing, its types and benefits. Companies like Amazon and Microsoft are providing the best cloud services with plethora of functions with their products like AWS and Azure. One of the biggest problem related with Cloud is Security as data is stored in third party infrastructure. In the recent times, information has become the biggest asset of the organizations and everyone wants to store their data in secure manner. This paper includes the method to secure the data using AES 256 bit encryption on the AWS using Boto3 library and Python code. It can be used to provide encryption at server side and make storage much secure than before.

Indexterms:Cloud, SaaS, IaaS, PaaS, Cloud Security, Confidentiality, Cloud Hacks, Amazon AWS.

I. INTRODUCTION

A cloud is a special type of network that relies on shared bunch of resources rather than having local servers or personal contrivances to deliver computing services. The concept of cloud computing has been escalating since it first invented in early 1970s. IT Companies offers these computing services are called cloud providers with Amazon, Microsoft, IBM, Cisco etc leading the way and charge for their computing services based on factors such as usage, cost, elasticity, storage and so on. Cloud computing provides myriad benefits to end users which are as follows:

Elasticity: Cloud platform gives elasticity to its user which eliminates the need for gigantic investments in local infrastructure.

Self-service provision: A user can compute resources without any interfering with administrator or any other regulatory body.

Migration flexibility: Organizations can move certain workloads [12] to or from the cloud server or to different platforms for better cost savings or to use new services as they emerge.

Reliability: Cloud computing makes data secure, provides backup of data, disaster recovery and allows business continuity, as data can be stored at multiple redundant sites on the cloud network.

Performance: The biggest advantage of cloud computing services is performance. The data over the network is fast and efficient, because it is placed on several datacenters. Hence, latency for application can be improved.

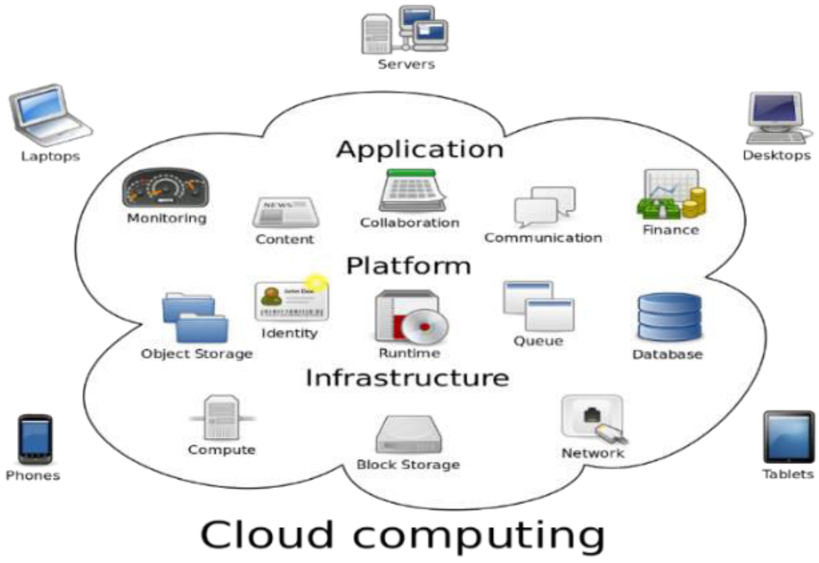


Figure 1.1 – Cloud Computing Architecture[16]

II. CLOUD COMPUTING TYPES

Infrastructure as a Service (IaaS): Infrastructure as a Service, can be abbreviated as IaaS. The main purpose of IaaS is to develop and deployment of PaaS, SaaS, and web-scale applications.

Platform as a Service (PaaS): Platforms as a service eliminate the need for organizations to manage the infrastructure (i.e. hardware and operating systems) and allow user to focus on the deployment and management of applications.

Software as a Service (SaaS): Software as a Service provides a completed solution that is run and managed by the service provider. SaaS applications can be run directly from browsers without any installations.

Based on a cloud location, we can classify cloud in 3 different section:

- Public
- Private
- Hybrid

Public cloud is a third-party cloud service provider which delivers the cloud service over the internet. Public cloud services are provided according to user's need or usage. In this model, the end customers only pay for the bandwidth or service they consume.

Private cloud refers to usage of a cloud network solely by single customer or organization. It is not shared with other users, although it is remotely located. By using private cloud, it easier for an organization to customize it according to meet specific IT requirements. Private clouds are often used by mid- to large-size organizations or government sectors where flexibility and security are at top priority.

Hybrid Cloud is a combination of both public and private cloud. It gives greater flexibility and more data deployment features to user. Figure 1.2 below depicts hybrid cloud:

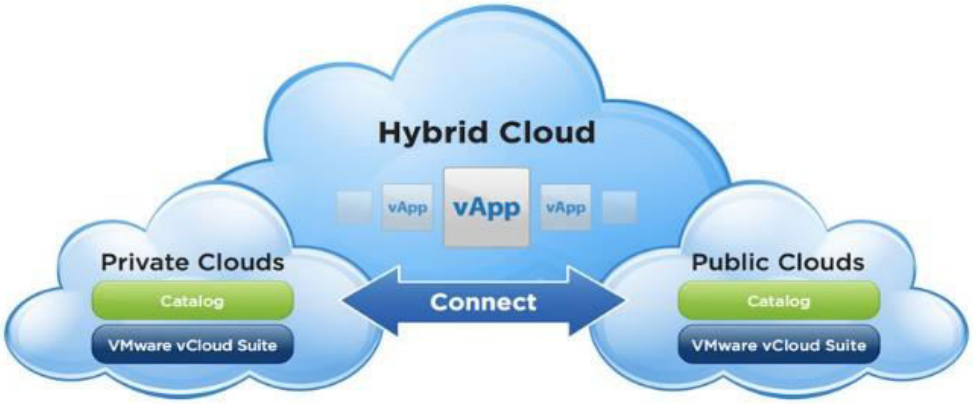


Figure 1.2 – Private/Public/Hybrid Cloud[17]

III. CLOUD COMPUTING CHALLENGES

Apart from many benefits that AWS Cloud provides, there are some security issues that it faces from day one. Three major challenges with Cloud Computing in order to build a secure and trustworthy cloud are explained below:

Outsourcing – Outsourcing is one of the reason cloud customers are using it because it reduces capital and operation expenditure of cloud customers. But it also means, that customers do not retain the total control on hardware[3], software and data. Amazon makes sure that Client's data remain secure and confidential on AWS, but there is no such thing as 100 percent security and most of the times configuration issues at the client end makes cloud vulnerable. Therefore to overcome this challenge, a cloud has to be trustworthy and should provide security services like confidentiality and integrity.

Multi-tenancy – Almost all Cloud computing vendors share their cloud among multiple customers. Virtualization and Containerization is used heavily by Cloud Vendors to optimize the resource allocation and to manage the resources in a much better manner. There is a pretty common but also risky situation as different customer's data is stored in same physical machine. Adversaries can exploit this type of vulnerability and can launch different type of attacks like flooding attacks[1] etc and if the best security practices are not used, then it can result in fatal damage for the cloud vendor and customer. Different vendors like Amazon, Microsoft, Google provides virtual web based firewall services to secure the cloud. For example Amazon AWS provides Shield application which can be configured and tuned for DDOS protection. Even if one can have Shield protection from Amazon AWS, another problem lies in configuring or integration of Cloud with the Shield as it requires special kind of knowledge in security domain to accurately configure Shield and most of the end users using Cloud without any security expert can create problem.

Massive data and intensive computation – As Cloud Vendors have all the infrastructure that make a data center capable of intensive computation, therefore these massive resources need a much better security than traditional security mechanisms and have newer and different security requirements which are needed to be fulfilled in order to make a trustworthy cloud

IV. RECENT AWS AND AZURE SECURITY FLAWS

Accenture AWS Data Breach - In one of the recent data breaches, Accenture accidently configured four S3 buckets in AWS as public[1][14]. Any user who is able to get the URL got the access to download data in those buckets. Those S3 buckets contained hundreds of GBs of data, that also had some passwords and private signing keys.

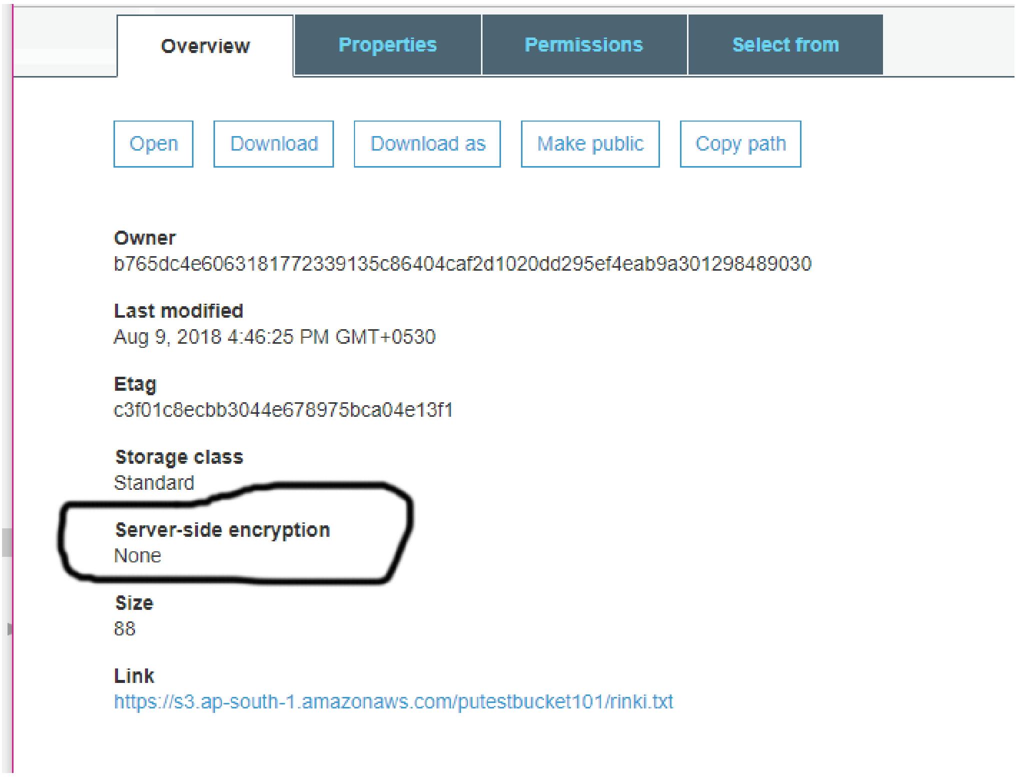
Time Warner Data Breach - Another data breach which was highlighted a lot was a misconfiguration by Time Warner Cable[14] in their Amazon S3 buckets which made them public and along with that around 4 million Time Warner Cable customers have their personal information exposed to the public internet.

Uber Security Breach –Uber[14] was also affected with the AWS data breach. It was infamous as Uber did not notify its 57 million hacked customers and drivers that their information is compromised and bribed hackers with a payment of $100,000 to keep incident quiet. The problem occurred when two hackers were able to gain access of Uber's private Github account and from where they were also able to get into company's AWS credentials.

Deloitte Azure Breach – Attackers were able to access the administrator account of Deloitte[15] email service which was hosted on Microsoft Azure. This account was not protected by two-factor authentication. Hackers also get the access to usernames, passwords, architectural designs for businesses and health information.

V. RESULTS

Amazon S3 is the storage phenomenon of AWS and is used for any frequently accessible data. Almost all the fortune top 100 companies are using Amazon AWS cloud and specifically their S3 storage service. By default, Amazon S3 stores data in plain text format and whenever we upload some data on the bucket, it is stored in plain text format as shown in the below diagram:



Owner
b765dc4e6063181772339135c86404caf2d1020dd295ef4eab9a301298489030

Last modified
Aug 9, 2018 4:46:25 PM GMT+0530

Etag
c3f01c8ecbb3044e678975bca04e13f1

Storage class
Standard

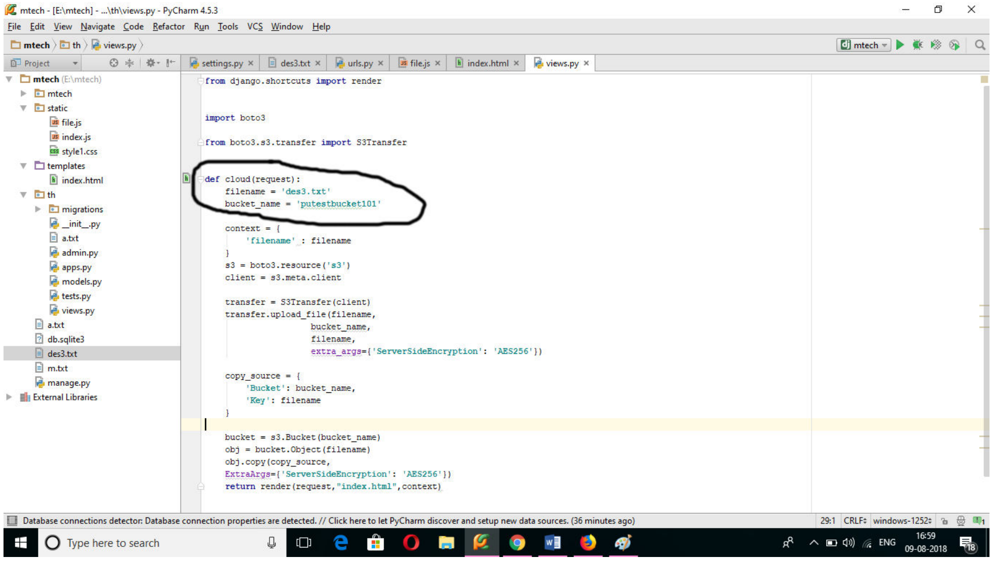
Server-side encryption
None

Size
88

Link
https://s3.ap-south-1.amazonaws.com/putestbucket101/rinki.txt

Figure 5.1 – File uploaded with no Server Side encryption

Data uploaded from local servers or machines towards S3 goes over the internet and needed to be secured by using some kind of encryption. We have created an uploading platform, which can use the AES 128 bit encryption by using the Boto3 package of Amazon AWS and created the code using python. Data can be encrypted with server side encryption using python boto3 package for Amazon AWS as shown below:

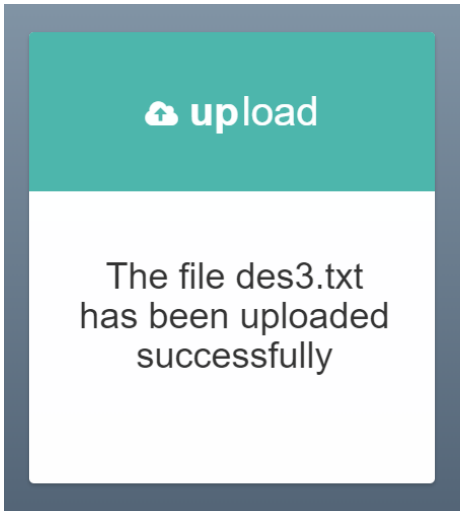


```python
from django.shortcuts import render


import boto3

from boto3.s3.transfer import S3Transfer

def cloud(request):
    filename = 'des3.txt'
    bucket_name = 'putestbucket101'

    context = {
        'filename' : filename
    }
    s3 = boto3.resource('s3')
    client = s3.meta.client

    transfer = S3Transfer(client)
    transfer.upload_file(filename,
                         bucket_name,
                         filename,
                         extra_args={'ServerSideEncryption': 'AES256'})

    copy_source = {
        'Bucket': bucket_name,
        'Key': filename
    }

    bucket = s3.Bucket(bucket_name)
    obj = bucket.Object(filename)
    obj.copy(copy_source,
    ExtraArgs={'ServerSideEncryption': 'AES256'})
    return render(request,"index.html",context)
```

Database connections detector: Database connection properties are detected. // Click here to let PyCharm discover and setup new data sources. (36 minutes ago)

Figure 5.2 – Bucket name and file name which is needed to be uploaded from client end to bucket.

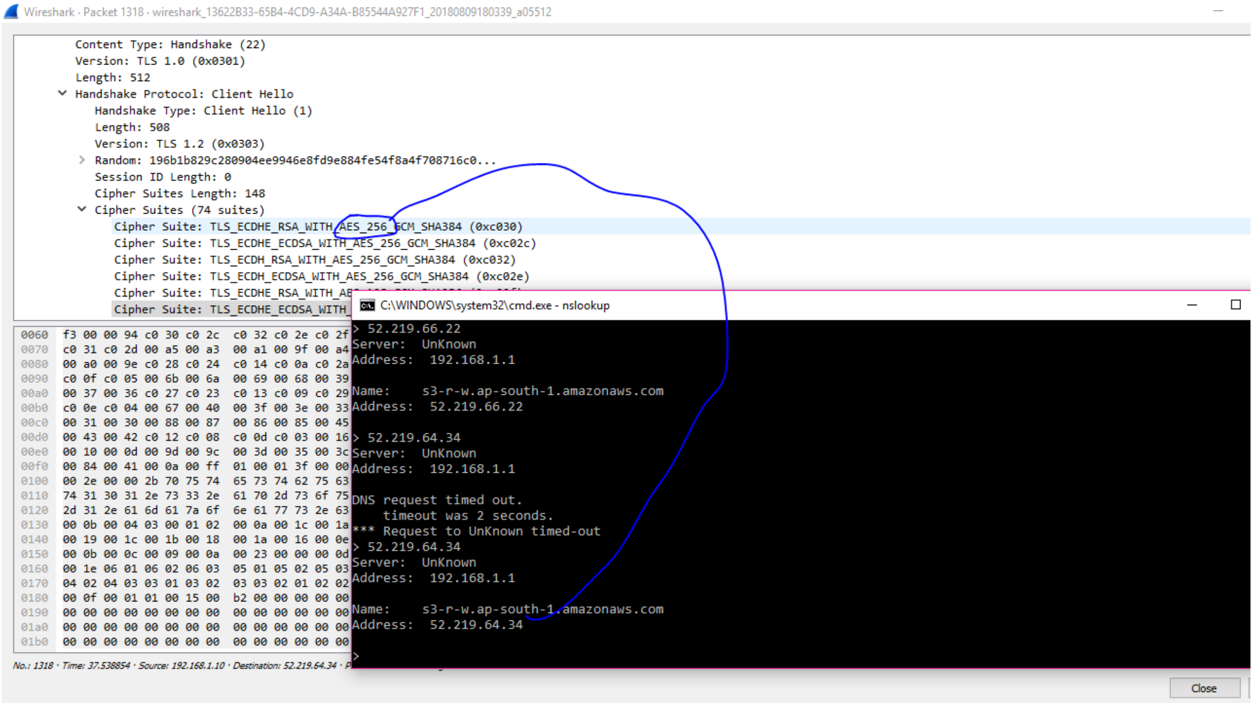
Message comes up after the uploading can be done as shown below:



The file des3.txt has been uploaded successfully

Figure 5.3 – File uploaded successfully

file is uploaded to make sure encryption takes place during the uploading. We took a AWS S3 capture that displays the communication channel is secure with AES encryption with 256 bit symmetric encryption as shown in figures 5.4 showing Wireshark capture and 5.5 below:


Wireshark · Packet 1318 · wireshark_13622B33-65B4-4CD9-A34A-B85544A927F1_20180809180339_a05512

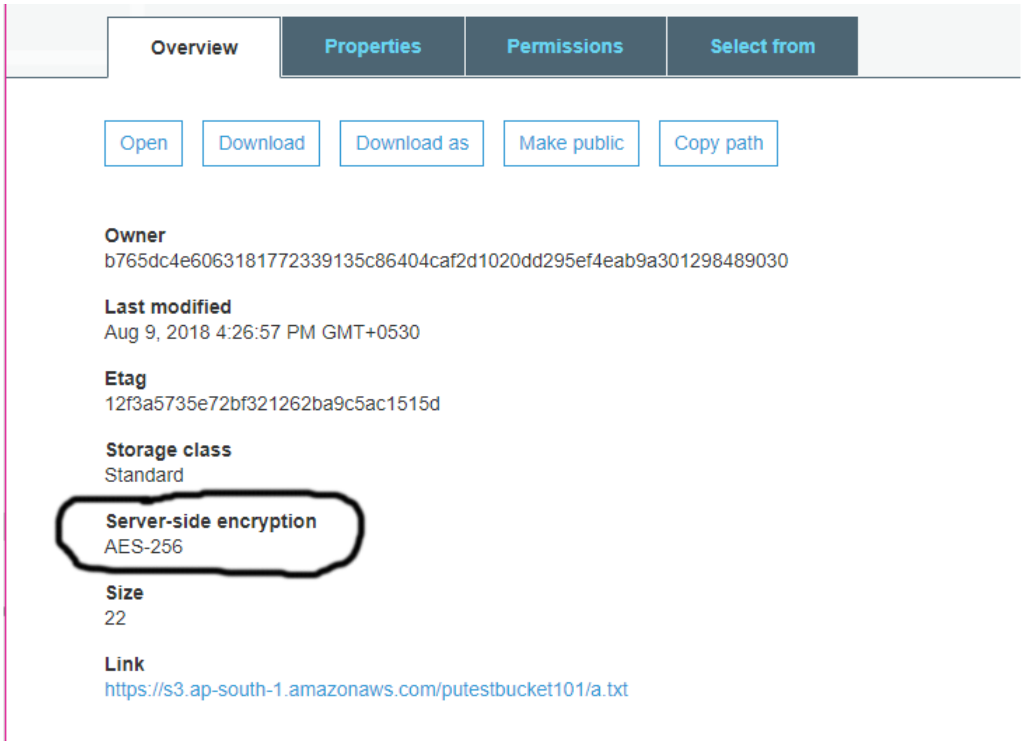
```
        Content Type: Handshake (22)
        Version: TLS 1.0 (0x0301)
        Length: 512
      ⌄ Handshake Protocol: Client Hello
          Handshake Type: Client Hello (1)
          Length: 508
          Version: TLS 1.2 (0x0303)
        > Random: 196b1b829c280904ee9946e8fd9e884fe54f8a4f708716c0...
          Session ID Length: 0
          Cipher Suites Length: 148
        ⌄ Cipher Suites (74 suites)
            Cipher Suite: TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 (0xc030)
            Cipher Suite: TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384 (0xc02c)
            Cipher Suite: TLS_ECDH_RSA_WITH_AES_256_GCM_SHA384 (0xc032)
            Cipher Suite: TLS_ECDH_ECDSA_WITH_AES_256_GCM_SHA384 (0xc02e)
            Cipher Suite: TLS_ECDHE_RSA_WITH_AE
            Cipher Suite: TLS_ECDHE_ECDSA_WITH_
```

```
0060  f3 00 00 94 c0 30 c0 2c  c0 32 c0 2e c0 2f
0070  c0 31 c0 2d 00 a5 00 a3  00 a1 00 9f 00 a4
0080  00 a0 00 9e c0 28 c0 24  c0 14 c0 0a c0 2a
0090  c0 0f c0 05 00 6b 00 6a  00 69 00 68 00 39
00a0  00 37 00 36 c0 27 c0 23  c0 13 c0 09 c0 29
00b0  c0 0e c0 04 00 67 00 40  00 3f 00 3e 00 33
00c0  00 31 00 30 00 88 00 87  00 86 00 85 00 45
00d0  00 43 00 42 c0 12 c0 08  c0 0d c0 03 00 16
00e0  00 10 00 0d 00 9d 00 9c  00 3d 00 35 00 3c
00f0  00 84 00 41 00 0a 00 ff  01 00 01 3f 00 00
0100  00 2e 00 00 2b 70 75 74  65 73 74 62 75 63
0110  74 31 30 31 2e 73 33 2e  61 70 2d 73 6f 75
0120  2d 31 2e 61 6d 61 7a 6f  6e 61 77 73 2e 63
0130  00 0b 00 04 03 00 01 02  00 0a 00 1c 00 1a
0140  00 19 00 1c 00 1b 00 18  00 1a 00 16 00 0e
0150  00 0b 00 0c 00 09 00 0a  00 23 00 00 00 0d
0160  00 1e 06 01 06 02 06 03  05 01 05 02 05 03
0170  04 02 04 03 03 01 03 02  03 03 02 01 02 02
0180  00 0f 00 01 01 00 15 00  b2 00 00 00 00 00
0190  00 00 00 00 00 00 00 00  00 00 00 00 00 00
01a0  00 00 00 00 00 00 00 00  00 00 00 00 00 00
01b0  00 00 00 00 00 00 00 00  00 00 00 00 00 00
```

No.: 1318 · Time: 37.538854 · Source: 192.168.1.10 · Destination: 52.219.64.34 · P

C:\WINDOWS\system32\cmd.exe - nslookup

```
> 52.219.66.22
Server:  UnKnown
Address:  192.168.1.1

Name:    s3-r-w.ap-south-1.amazonaws.com
Address:  52.219.66.22

> 52.219.64.34
Server:  UnKnown
Address:  192.168.1.1

DNS request timed out.
    timeout was 2 seconds.
*** Request to UnKnown timed-out
> 52.219.64.34
Server:  UnKnown
Address:  192.168.1.1

Name:    s3-r-w.ap-south-1.amazonaws.com
Address:  52.219.64.34

>
```



Owner
b765dc4e6063181772339135c86404caf2d1020dd295ef4eab9a301298489030

Last modified
Aug 9, 2018 4:26:57 PM GMT+0530

Etag
12f3a5735e72bf321262ba9c5ac1515d

Storage class
Standard

Server-side encryption
AES-256

Size
22

Link
https://s3.ap-south-1.amazonaws.com/putestbucket101/a.txt

Figure 5.4 – Encryption applied after code is applied.

As shown above in the figure, file is now stored with AES 256 strong encryption after using the code to upload the file on AWS S3. It helps in securing the data on the cloud as compared with Amazon AWS S3 plain-text storage.

CONCLUSION

AWS is the most used and popular cloud in the world. It provides multiple cloud services in the compute, storage, networking, IOT, security, machine learning etc. S3 is the storage application provided by S3 and is used by cloud users to store the data which they use frequently. S3 does not provide any encryption facility to encrypt the data and by default all the data is stored in plain text form. Boto3 is the library package of Amazon AWS which can be used with Python to create scripts to perform functions in automated manner. AES is a symmetric encryption technique mainly comes with 128, 192 and 256 bit standards. We have created a script that uses AES 256 encryption and helps in encrypting the data at Amazon S3. It is always better to secure the data by encrypting files using strong encryption standards and AES 256 is a strong encryption.

REFERENCES

[1]Steffen Müller, Frank Pallas, and Silvia Balaban(2015)," On the Security of Public Cloud Storage",10th Future Security Conference 2015 (Future Security 2015), Fraunhofer.

[2] Prerna and Parul Agarwal(2017)," Cryptography Based Security for Cloud Computing System", International Journal of Advanced Research in Computer Science(IJARCS).

[3] Mosca, P., Zhang, Y.P., Xiao, Z.F. and Wang, Y. (2014),"Cloud Security: Services, Risks, and a Case Study on Amazon Cloud Services". Int. J. Communications, Network and System Sciences, 7, 529-535. http://dx.doi.org/10.4236/ijcns.2014.712053

[4] Anne Canteaut, Sergiu Carpov, Caroline Fontaine, Jacques Fournier, Benjamin Lac, Maria Naya-Plasencia, Renaud Sirdey, and AssiaTria(2017)," End-to-end data security for IoT: from a cloud of encryptions to encryption in the cloud", Cesar Conference(2017)

[5] Ali Abdulridha Taha, Dr. Diaa Salama AbdElminaam, Prof.Dr. Khalid M Hosny(2017)," NHCA: Developing New Hybrid Cryptography Algorithm for Cloud Computing Environment", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 8, No. 11, 2017 .

[6] Mishra, N., Kanchan, K., Ritu, C. and Abhishek, C. (2013)," Technologies of Cloud Computing-Architecture Concepts based on Security and its Challenges.", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2 (3), 1143 – 1149.

[7] Murali, M., Kinnari, S. and Gunda, M. (2013)," Enabling Secure Database as a Service using Fully Homomorphic Encryption: Challenges and Opportunities. Cornell University Computer Science Database." Arxiv: 1302.2654. 1-5.

[8] Kalpana, P. and Sudha, S. (2012)," Data Security in Cloud Computing using RSA Algorithm.", International Journal of Research in Computer and Communication Technology (IJRCCT), 1 (4), 143 – 146.

[9] Google, (2012)," Google Cloud Storage: A Simple Way to Store, Protect, and Share Data.", Google Inc., USA.

[10] Encryption At Rest In Google Cloud Platform, an article available at https://cloud.google.com/security/encryption-at-rest/default-encryption/ , April 2017.

[11] Google, (2012a)," Google's Approach to IT Security: A Google White Paper." ,Google Inc., USA.

[12] Google, (2013)," Just Develop IT Migrates Petabytes of Data to Google Cloud Storage.", Retrieved from http://googlecloudplatform.blogspot.com

[13] Jeff, B. (2011),"New - Amazon S3 Server Side Encryption for Data at Rest."Retrieved from http://aws.amazon.com/blogs/aws/new-amazon-s3-server-side-encryption/.

[14] Brien Posey(2018), "Biggest AWS Security Breaches of 2017." Retrieved from - https://www.sumologic.com/blog/security/aws-security-breaches-2017/

[15] Nick Hopkins(2017), "Deloitte hit by cyber-attack revealing clients' secret emails." Retrieved from - https://www.theguardian.com/business/2017/sep/25/deloitte-hit-by-cyber-attack-revealing-clients-secret-emails

[16] Cloud Computing Image, "Retrieved from https://upload.wikimedia.org/wikipedia/commons/thumb/b/b5/Cloud_computing.svg/1200px-Cloud_computing.svg.png."

[17] Private/Public/Hybrid Cloud, "Retrieved from - http://www.businesscloudnews.com/wp-content/blogs.dir/122/files/2014/07/VMware-hybrid-cloud.jpg."