International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Secure and Selective Cloud Data Auditing using Deep Machine Learning

S.Radharani, V.B.Narasimha

Abstract: ***The tradition of moving applications, data to be consumed by the applications and the data generated by the applications is increasing and the increase is due to the advantages of cloud computing. The advantages of cloud computing are catered to the application owners, application consumers and at the same time to the cloud datacentre owners or the cloud service providers also. Since IT tasks are vital for business progression, it for the most part incorporates repetitive or reinforcement segments and framework for power supply, data correspondences associations, natural controls and different security gadgets. An extensive data centre is a mechanical scale task utilizing as much power as a community. The primary advantage of pushing the applications on the cloud-based data centres are low infrastructure maintenance with significant cost reduction for the application owners and the high profitability for the data centre cloud service providers. During the application migration to the cloud data centres, the data and few components of the application become exposed to certain users. Also, the applications, which are hosted on the cloud data centres must comply with the certain standards for being accepted by various application consumers. In order to achieve the standard certifications, the applications and the data must be audited by various auditing companies. Few of the cases, the auditors are hired by the data centre owners and few of times, the auditors are engaged by application consumers. Nonetheless, in both situations, the auditors are third party and the risk of exposing business logics in the applications and the data always persists. Nevertheless, the auditor being a third-party user, the data exposure is a high risk. Also, in a data centre environment, it is highly difficult to ensure isolation of the data from different auditors, who may not be have the right to audit the data. Significant number of researches have attempted to provide a generic solution to this problem. However, the solutions are highly criticized by the research community for making generic assumptions during the permission verification process. Henceforth, this work produces a novel machine learning based algorithm to assign or grant audit access permissions to specific auditors in a random situation without other approvals based on the characteristics of the virtual machine, in which the application and the data is deployed, and the auditing user entity. The results of the proposed algorithm are highly satisfactory and demonstrates nearly 99% accuracy on data characteristics analysis, nearly 98% accuracy on user characteristics analysis and 100% accuracy on secure auditor selection process***

Keywords : VM Data Characteristics, Auditor Data Characteristics, Change Frequency, Deep Learning, VM Consolidation.

I. INTRODUCTION

education, business, communication, scientific research and social media have increased ups and bounds. Also, the data related to these applications are also increasing in terabytes by everyday of usage. Hence most of the application owners are migrating or directly building the applications on the cloud-based data centres. The data stored on the cloud-based data centres are not prone to the data losses or mis-matching with the standards to be complied. The loss of the data on the data centres can be due to various reasons and security is one of the most prominent reasons. The detailed survey work by K. Ren et al. [1] demonstrates various challenges on the data on the cloud. Hence the data on the cloud computing environment must be audited in regular intervals and the auditing process is made mandatory by most of the application consumers in terms of various certifications of security [Fig – 1].

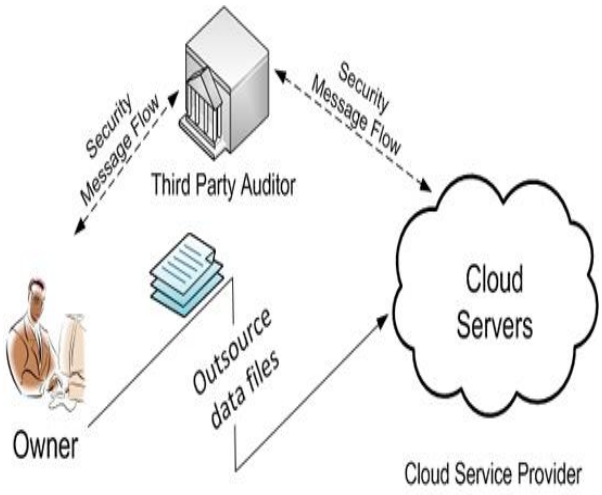

Fig. 1 Process of Certificate-based Auditing reporting on Cloud Data

In the recent years, a good number of researches have tried to build an auditing process for the data on the cloud. The work by G. Ateniese et al. [2] have showcased the validation of the data from the untrusted storage sequences. This work was criticised by the researcher's community for not considering the possibilities of mobility of the cloud computing paradigm as the storage has to match the demand of data mobility and trust management cannot be location specific. Also, the work by A. Juels et al.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
S. Radharani*, Research scholar in Department of CSE , UCE,OU. ramsmcaou@gmail.com
Dr V.B. Narasimha, Assistant professor, Department of CSE,UCE,O.U , vbnarasimha@gmail.com

[3] have also highlighted the challenges of the handling the data which is higher in dimensions and in size. The solution to this problem is addressed by H.

Shacham et al. [4]. This solution, providing the mechanism for making the data more compact in nature, make applying security validation compact [Fig – 2].

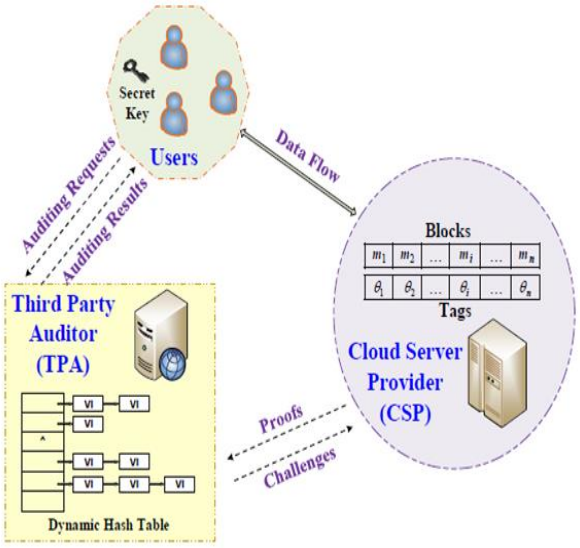

Fig. 2 Process of Data Compacting for Auditing on Cloud Data

In the other hand, the challenges of addressing the data security for public cloud computing environment are also addressed by other parallel researchers. The work by C. Wang et al. [5] have first listed the challenges and aimed to resolve few of the issues. Further, the work by S. G. Worku et al. [6] have formulated the privacy-based security policies for data access. After this research direction, many of the researchers have attempted to secure the data from the auditing process using encryption techniques. The work by C. Guan et al. [7] have demonstrated the benefits of securing the audit data.

Nevertheless, the implications of the cloud-based data auditing are always been a challenge and the survey work by W. Shen et al. [8] have confirmed the list.

Considering the baseline of this research direction, this work identified the following problems to be resolved by the recent researchers:

- Analyse the cloud-based auditing standards and identify the drawbacks
- Propose a novel algorithm for selective access for auditors with characteristics based identify management using machine learning techniques
- Propose a novel algorithm for data schematic analysis and propose a selective auditing for the data generated by applications on cloud
- Propose a novel method to match and assign selective auditing process with data security

Henceforth, this work, proposes a novel secure selective algorithm for validation of data audits on cloud-based data centres and also provides the solutions to the above-mentioned problems.

The rest of the work is organized such as in Section – II, the fundamentals of the cloud data auditing is analysed mathematically, in Section – III, the parallel research attempts are critically analysed for formulating the research problem, in Section – IV, the problem formulation is carried out with the help of the mathematical lemma, in Section – V, the novel algorithm and the components of the algorithms with detailed description are furnished, in the Section – VI, the results obtained from these algorithm components are discussed with the enhancement measures, in Section – VII, the comparative analysis is furnished with the parallel research outcome benefits and finally in the Section – VIII, the work presents the research conclusion.

I. FUNDAMENTALS OF CLOUD DATA AUDITING

In this section of the work, the fundamentals of cloud-based data centre application and data auditing are elaborated using the mathematical modelling.

Firstly, assuming that, the data centre, D, is a collection of physical hosts as denoted by I_x, with total number of instances as n. Hence, the formulation can be presented as,

$$D = \sum_{i=1}^{n} I_i \qquad \text{(Eq.1)}$$

Further, assuming that, each and every instance is also collection of virtual machines, VM. Hence, with m number of virtual machines on each instance, this can be formulated as,

$$I_x = \sum_{j=1}^{m} VM_j \qquad \text{(Eq.2)}$$

It is to be understood, that the number of virtual machines in each instance is not identical with other instances.

Also, the data centre will host the applications, A_P, which must have multiple sub-components as, A_k. This relation can be formulated as,

$$A_p = \sum_{k=1}^{p} A_k \qquad \text{(Eq.3)}$$

Assuming that, the number of applications hosted on the data centre are p.

Naturally, the applications will consume and generate data during the life cycle process. Assuming the entire data on the data centre, which has affinity to the applications, can be denoted as, D_p and the count of the data instances can be denoted as logarithmic value range of the number of applications, p. Hence, this relation can be formulated as,

$$D_p = \sum_{k=1}^{p^{\log p!}} D_k \qquad \text{(Eq.4)}$$

During the complete life cycle of the data and the applications must go through a number of auditing processes at various time instances. Any auditing process shall validate the application component, data component and the behaviour of the complete set at any given point, which can be formulated as,

$$\oint (A_k, t, D_k) \rightarrow \lambda \qquad \text{(Eq.5)}$$

Here, the characteristics or the behaviour of the this set, λ, *must be taken under consideration for validation or auditing.*

The characteristics, λ, *is again a collection of* the time stamp, t, the nature of the operation on the data, A, and finally the frequency of the operation on the data, F. This relation can be formulated as,

$$\lambda \rightarrow \oint (t_1, A_1, F_1) \quad (Eq.6)$$

Similarly, assuming for another instance of the application and data, the formulation,

$$\tau \rightarrow \oint (t_2, A_2, F_2) \quad (Eq.7)$$

Here, during the auditing process, the characteristics, λ and τ must be verified and validated. The formulation of the verification process can be represented as,

$$If \; |t_1 \neq t_2 \; and \; A_1 = A_2$$
$$Then, F_1 = F_2 = \frac{F_1 + F_2}{\Phi(F_1, F_2)} \quad (Eq.8)$$

Hence, the auditing of the data and applications on the cloud data centre can realized.

II. Recent Research Outcomes

Based on the analysis in the prior section of this work on fundamentals of data auditing on cloud, the knowledge for analysing the recent research outcomes have gathered. Henceforth, in this section of the work, the parallel research approaches and their outcomes are analysed critically.

The first issues with the data auditing process was encountered by the medical record storage and processing systems as reported by J. Sun et al. [9]. The medical data is highly mission critical and must be highly accurate for the further processing. Hence the auditing must be performed in very frequent intervals. Also, in the other hand, the as the data is highly confidential, thus trusting any third-party auditors for this process is also risk provoking. Hence, the process is highly criticised.

In the other hand, there are some application scenarios, where the data continuously under change. Hence, detecting or performing the auditing process on such data is highly complex. The work presented by G. Ateniese et al. [10] demonstrates a standard auditing process for highly updatable data. The process is fairly static and cannot comply with the dynamic changes on the data. Hence, the auditing process must consider the change characteristics of the data and the applications deployed on the same data. This problem was resolved in the work by C. Erway et al. [11] to some extends.

Further, the challenge with dynamic data on the cloud is that the data can be accessed and updated by various applications and these applications again can be consumed by various users. Hence, the auditing cannot be performed by a single entity. Thus, the content of public auditing is bought into the research and practice by Q. Wang et al. [12].

Also, significant amount of contributions was made by the works of J. Yu et al. [13], J. Yu and K. Ren et al. [14], J. Yu and H. Wang [15] and J. Yu and R. Hao et al. [16] have demonstrated sufficient reasons for further studies in this research directions.

Finally, the work by B. Wang et al. [17] and G. Yang et al. [18] have made the foundation of the further work in this research direction by specifying the demand of the machine learning methods and analytical research on the behavioural pattern of the data and auditing users for public auditing processes.

Henceforth, with the detailed understanding of the bottlenecks of the parallel recent research outcomes and realizing the future research directions with trends, this work formulates the actual research problem in the next section of this work.

III. Problem Formulation

After the deep analysis of the fundamental principle of cloud data auditing and understanding of the parallel research outcomes in the prior sections of this work, in this section the problem is formulated mathematically using mathematical lemma.

Lemma: To increase the security during the auditing process, the selective auditor access permissions must be incorporated using the auditor characteristics.

Proof: Firstly, assuming that, the group of auditors, set U, assigned for a data centre is a collection of the auditing users, where each user can be represented as, U_x. Hence for a total number, k, of auditing users on the data centre can be presented as,

$$U[] = \sum_{x=1}^{k} U_x \quad (Eq.9)$$

Further, each auditing user can be characterised as, the time of the action performed on the virtual machine data, t, the type of action performed on the VM data, A, the frequency of the VM data read actions, FR, and finally the frequency of the update operation on the virtual machine data, FU. Hence, this relationship can be formulated as,

$$U_x = \oint (t_x, A_x, FR_x, FU_x) \quad (Eq.10)$$

Henceforth, for any two random users, $U_{x=1}$ and $U_{x=2}$, the Eq. 10 can be reformed as,

$$U_{x=1} = \oint (t_1, A_1, FR_1, FU_1) \quad (Eq.11)$$

And,

$$U_{x=2} = \oint (t_2, A_2, FR_2, FU_2) \quad (Eq.12)$$

Again, the virtual machine data characteristics can be reformed for any two random virtual machines, $VM_{x=1}$ and $VM_{x=2}$, from Eq. 6.

The Eq. 6 can be re-written as,

$$VM_{x=1} \rightarrow \oint (t_1, A_1, F_1) \quad \text{(Eq.13)}$$

And,

$$VM_{x=2} \rightarrow \oint (t_2, A_2, F_2) \quad \text{(Eq.14)}$$

Assigning the correct auditing users to the correct virtual machine is the key to the successful selective secure auditing process.

Henceforth, the following formulation can be realized,

$$U_x (U_{x=1} \neq U_{x=2}) : \exists VM[] \rightarrow \left| \sum_{k=1}^{n} A_k \subseteq A_x \right| \cup \frac{\left| \sum_{k=1}^{n} FR_k \subseteq FR_x \right|}{\left| \sum_{k=1}^{n} FR_k \subseteq FR_x \right| + \left| \sum_{k=1}^{n} FU_k \subseteq FU_x \right|} \quad \text{(Eq.15)}$$

Thus, it is natural to realize that using the characteristics method, for a specific auditing user, U_x, the set of correct virtual machines, $VM[]$, can be assigned using the action and frequency of that action characteristics to make the auditing process highly secure and selective.

This mathematical formulation of the problem will help in building the secure selective algorithm for cloud-based data auditing process, which is elaborated in the next section of the work.

IV. Novel Secure Selective Auditing Algorithms

The mathematical formulation of the problem statement in the prior section of this work helps in building the robust algorithms for the secure selective auditing process. The total security component is distributed in few of the following algorithm sets, which are furnished here.

Firstly, the virtual machine data characteristics analysis algorithm is furnished here.

Algorithm - 1: **Extraction of the VM Characteristics using Deep Learning** Algorithm (EVMC-DL)

Step - 1. Accept the list of instances in the data centre as I[]
Step - 2. For each I[]
- a. Accept the list of virtual machine data as D[]
- b. For each D[] access the data centre log
  - i. Accept the total operation on the VM set as A[]
  - ii. Accept the total duration of each operation as T(A[])
  - iii. Accept the total data transfer for each operation as S(A[])
  - iv. For each A[]
    1. If A[i].Lock == "Shared Lock" && A[i].Lock != "Exclusive Lock" && A[i].Lock != "No Lock"
    2. Then,
    a. Update the A[i].L[i] = "Read"
    3. Else,
    a. Update the A[i].L[i] = "Update"
    4. If A[i].L[i] == "Read"
    5. Then,
    a. Calculate the time frequency, TF(A[i].L[i] = "Read") = (Sum(T(A[A[i].L[i]])))/(Count of T(A[]))
    b. Calculate the size frequency, SF(A[i].L[i] = "Read") = (Sum(S(A[A[i].L[i]])))/(Count of S(A[]))
    6. Else,
    a. Calculate the TF(A[i].L[i] = "Update") = (Sum(T(A[A[i].L[i]])))/(Count of T(A[]))
    b. Calculate the SF(A[i].L[i] = "Update") = (Sum(S(A[A[i].L[i]])))/(Count of S(A[]))
  - v. Report the TF & SF for all applications
- c. Report the TF[] & SF[] for all virtual machines

a data distribution centre to clear up the structure, substance, connections, and inference principles of the data. Profiling just gets peculiarities and survey data quality, yet in addition to find, register, and evaluate undertaking metadata. The aftereffect of the examination is utilized to decide the reasonableness of the hopeful source frameworks, typically giving the reason for an early go/no-go choice, and furthermore to recognize issues for later arrangement plan.

Secondly, the auditing user characteristics analysis algorithm is furnished here.

Algorithm - 2: **Extraction of the Auditing Users Characteristics using Deep Learning** Algorithm (EAUC-DL)

Step - 1. Accept the total list of auditing users as U[]
Step - 2. For each U[]
Step - 3. Accept the total operation as A'[]
Step - 4. Accept the total duration of each operation as T'(A[])
Step - 5. Accept the total data transfer for each operation as S'(A[])
Step - 6. For each A'[]
- i. If A'[i].Lock == "Shared Lock" && A'[i].Lock != "Exclusive Lock" && A'[i].Lock != "No Lock"
- ii. Then,
  a. Update the A'[i].L[i] = "Read"
- iii. Else,
  a. Update the A'[i].L[i] = "Update"
- iv. If A'[i].L[i] == "Read"
- v. Then,
  a. Calculate the time frequency, TF'(A'[i].L[i] = "Read") = (Sum(T'(A'[A'[i].L[i]])))/(Count of T'(A'[]))
  b. Calculate the size frequency, SF'(A'[i].L[i] = "Read") = (Sum(S'(A'[A'[i].L[i]])))/(Count of S'(A'[]))
- vi. Else,
  a. Calculate the TF'(A'[i].L[i] = "Update") = (Sum(T'(A'[A'[i].L[i]])))/(Count of T'(A'[]))
  b. Calculate the SF'(A'[i].L[i] = "Update") = (Sum(S'(A'[A'[i].L[i]])))/(Count of S'(A'[]))
- vii. Report the TF & SF for all applications

Report the TF'[] & SF'[] for all users

In auditing user-cantered plan, personas are made to speak to the sorts of users. It is now and again indicated for every persona which sorts of user interfaces it is fine with, and what specialized skill and level of information it has in explicit fields or teaches. At the point when a couple of requirements are forced on the end-user class, particularly when planning programs for use by the overall population, usually practice anticipating negligible specialized skill or past preparing in end users.

Finally, the secure selective auditing algorithm is furnished here.

Algorithm - 3: **Secure Selective Auditing Assignment of VMs using Deep Learning** Algorithm (SSAAVM-DL)
Step - 1. Accept all the TF[] & SF[] and TF'[] & SF'[] Step - 2. Assign Error_Rate = 0 Step - 3. For each U[] as U[i] a. If TF'[] similar to Mean of TF[] ± Error_Rate and SF'[] similar to Mean of SF[] ± Error_Rate b. Then, i. Calculate the VM[] as 1. If VM[i].TF[] similar to Mean of TF[] and VM[i].SF[] similar to Mean of SF[] 2. Then, a. Add VM[i] to VM[] set 3. Else a. Ignore VM[i] and add to VM'[] ii. Return the total VM[] iii. If total of VM[].TF[] similar to Mean of TF[] and total of VM[].SF[] similar to Mean of SF[] iv. Then, 1. Map VM[] to U[i] v. Else 1. Calculate the Error_Rate as VM[] - VM'[] vi. Repeat From Step - 3 until Error_Rate = 0 Step - 4. Report the final U[] with VM[]

Cloud security engineering is powerful just if the right protective usage is set up. A productive cloud security engineering ought to perceive the issues that will emerge with security the board. The security the board tends to these issues with security controls. These controls are set up to protect any shortcomings in the framework and decrease the impact of an assault. While there are numerous kinds of controls behind a cloud security design, they can more often than to be found in one of the numerous classifications.

Henceforth, with the furnishing of the algorithms in this section of the work, in the next section, the obtained results from the secure selective auditing process is elaborated and discussed.

V. Results and Discussion

The results obtained from the proposed novel algorithms are highly satisfactory and the results are furnished here in this section of the work with discussions.

A. *VM Data Read Characteristics Analysis*

Firstly, the read characteristics analysis on the virtual machine data is furnished and discussed [Table – 1]:

TABLE I
Read Access Analysis

VM Data	Number of Reads	Freq. of Read
Dataset1.dat	41	0.000408961
Dataset2.dat	38	0.000379037
Dataset3.dat	40	0.000398987
Dataset4.dat	42	0.000418936
Dataset5.dat	36	0.000359088
Dataset6.dat	39	0.000389012
Dataset7.dat	34	0.000339139
Dataset8.dat	36	0.000359088
Dataset9.dat	28	0.000279291
Dataset10.dat	41	0.000408961

The essential preferred standpoint of utilizing data access objects is the moderately straightforward and thorough partition between two significant pieces of an application that can however ought not know anything of one another, and which can be relied upon to develop much of the time and autonomously. Changing business rationale can depend on the equivalent DAO interface, while changes to steadiness rationale don't influence DAO customers as long as the interface remains effectively executed.

The result is visualized graphically here [Fig – 3].

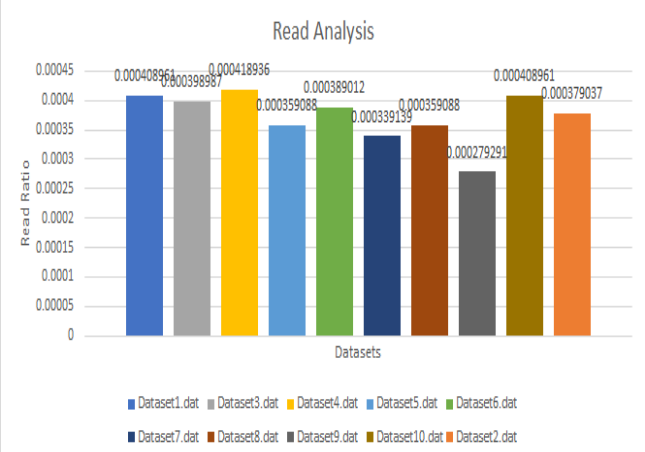

Fig. 3 Read Ratio (Frequency) Analysis on VM Data

B. *VM Data Read Size Analysis*

Secondly, the read size analysis on the virtual machine data is furnished and discussed [Table – 2]:

TABLE II
Read Size Analysis

VM Data	Number of Reads	Mean Size (MB)
Dataset1.dat	41	1585.275391
Dataset2.dat	38	1582.652344
Dataset3.dat	40	1583.521484
Dataset4.dat	42	1580.914063
Dataset5.dat	36	1581.783203
Dataset6.dat	39	1581.160156
Dataset7.dat	34	1578.055664
Dataset8.dat	36	1579.424805
Dataset9.dat	28	1576.301758
Dataset10.dat	41	34.00390625

The result is visualized graphically here [Fig – 4].

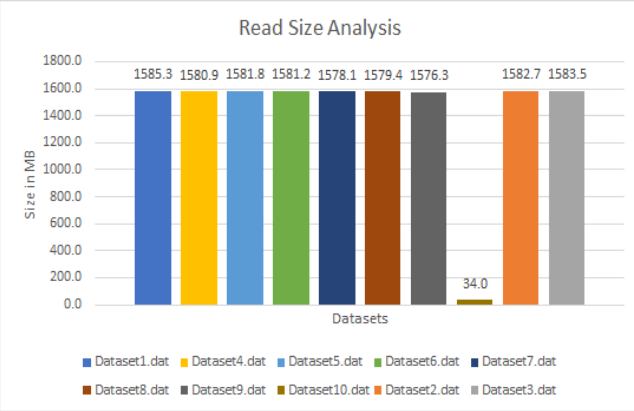

Fig. 4 Read Size Analysis on VM Data

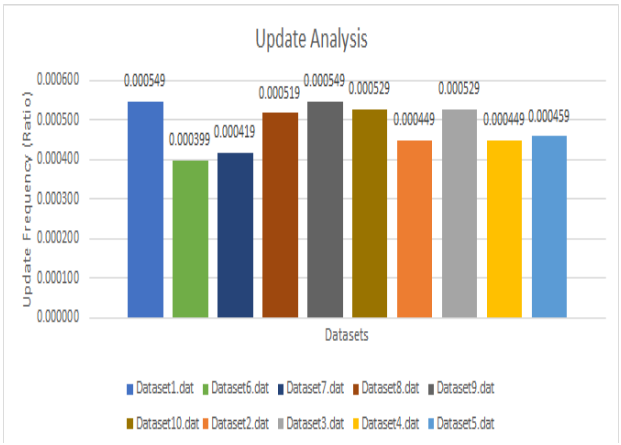

Fig. 5 Update Access Analysis on VM Data

C. VM Data Update Characteristics Analysis

Third, the update analysis on the virtual machine data is furnished and discussed [Table – 3]:

TABLE III
UPDATE ACCESS ANALYSIS

VM Data	Number of Updates	Freq. of Update
Dataset1.dat	55	0.000548607
Dataset2.dat	45	0.00044886
Dataset3.dat	53	0.000528657
Dataset4.dat	45	0.00044886
Dataset5.dat	46	0.000458835
Dataset6.dat	40	0.000398987
Dataset7.dat	42	0.000418936
Dataset8.dat	52	0.000518683
Dataset9.dat	55	0.000548607
Dataset10.dat	53	0.000528657

Frameworks perform input/yield utilizing a unique program for an I/O channel, a processor devoted to controlling fringe stockpiling gadget access and data exchange to and from primary memory. Channel programs are made out of channel direction words (CCWs). Programming those is a mind-boggling task requiring nitty gritty information of the equipment qualities. Channel programs are started by a START IO guidance issued by the working framework. This is typically front finished by the Execute Channel Program (EXCP) full scale for application software engineer accommodation. EXCP issues an SVC (chief call guidance) that guides the working framework to issue the START IO for the application's sake.

The result is visualized graphically here [Fig – 5].

D. VM Data Update Size Analysis

Fourthly, the update size analysis on the virtual machine data is furnished and discussed [Table – 4]:

TABLE IV
ACCESS SIZE ANALYSIS

VM Data	Number of Reads	Mean Size (MB)
Dataset1.dat	41	279.2785347
Dataset2.dat	38	245.6970857
Dataset3.dat	40	277.0116943
Dataset4.dat	42	224.3418666
Dataset5.dat	36	266.8870985
Dataset6.dat	39	215.8573718
Dataset7.dat	34	259.1162109
Dataset8.dat	36	305.2644314
Dataset9.dat	28	413.7660435
Dataset10.dat	41	243.2276343

The result is visualized graphically here [Fig – 6].

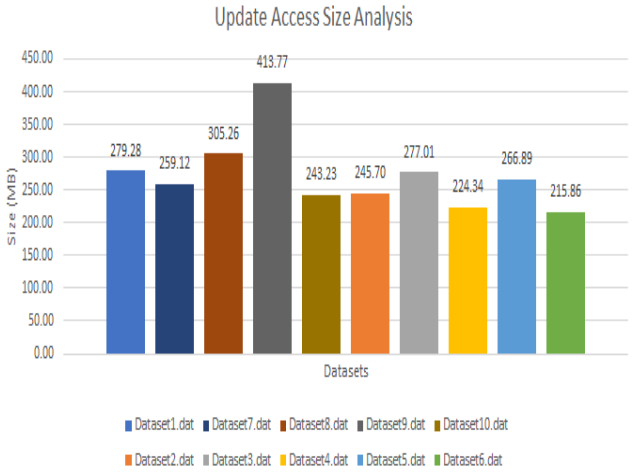

Fig. 6 Update Size Analysis on VM Data

E. Auditing User Analysis

Fifth, the auditing user analysis on the virtual machine data is furnished and discussed [Table – 5]:

TABLE V
AUDITING USER ANALYSIS

Users	Number of Reads	Freq. of Read	Mean Size (MB)	Number of Updates	Freq. of Update	Mean Size (MB)
1	3	0.0002825	906.3774	10	0.0009415	4562.4564
2	3	0.0002825	968.7005	8	0.0007532	2879.2112
3	3	0.0002825	1164.0531	6	0.0005649	3142.4922
4	4	0.0003766	1413.8480	5	0.0004708	1434.1389
5	6	0.0005649	794.3825	7	0.0006591	1171.5460
6	7	0.0006591	844.1648	5	0.0004708	1053.2839
7	2	0.0001883	1339.9080	2	0.0001883	840.3167
8	3	0.0002825	1130.2103	9	0.0008474	3648.1150
9	3	0.0002825	1511.3605	8	0.0007532	2254.7693
10	3	0.0002825	21.0266	3	0.0002825	28.9203

Observing all movement at the information layer is a key segment of an information driven security system. It gives perceivability into the sorts of activities that clients and apparatuses have mentioned and been approved to on explicit information components. Constant observing at the information layer joined with exact access control can contribute altogether to the ongoing location of information ruptures, constrains the harms perpetrated by a break and can even stop the interruption if legitimate controls are set up.

F. Auditing User Analysis – Read Access on VM Data

Further, the auditing user analysis on the virtual machine data is furnished and discussed [Table – 6] with few samples from the result:

TABLE VI
USER ANALYSIS – READ ACCESS

User_Seq	User Read Frequency	VM Dataset	VM Read Frequency
1	0.000282459	Dataset1.dat	0.000408961
1	0.000282459	Dataset10.dat	0.000408961
1	0.000282459	Dataset2.dat	0.000379037
1	0.000282459	Dataset3.dat	0.000398987
1	0.000282459	Dataset4.dat	0.000418936
1	0.000282459	Dataset5.dat	0.000359088
1	0.000282459	Dataset6.dat	0.000389012
1	0.000282459	Dataset7.dat	0.000339139
1	0.000282459	Dataset8.dat	0.000359088
1	0.000282459	Dataset9.dat	0.000279291
2	0.000282459	Dataset1.dat	0.000408961
2	0.000282459	Dataset10.dat	0.000408961
2	0.000282459	Dataset2.dat	0.000379037
2	0.000282459	Dataset3.dat	0.000398987
2	0.000282459	Dataset4.dat	0.000418936
2	0.000282459	Dataset5.dat	0.000359088
2	0.000282459	Dataset6.dat	0.000389012
2	0.000282459	Dataset7.dat	0.000339139
2	0.000282459	Dataset8.dat	0.000359088
2	0.000282459	Dataset9.dat	0.000279291
3	0.000282459	Dataset1.dat	0.000408961
3	0.000282459	Dataset10.dat	0.000408961
3	0.000282459	Dataset2.dat	0.000379037
3	0.000282459	Dataset3.dat	0.000398987
3	0.000282459	Dataset4.dat	0.000418936
3	0.000282459	Dataset5.dat	0.000359088
3	0.000282459	Dataset6.dat	0.000389012
3	0.000282459	Dataset7.dat	0.000339139
3	0.000282459	Dataset8.dat	0.000359088
3	0.000282459	Dataset9.dat	0.000279291
4	0.000376612	Dataset1.dat	0.000408961
4	0.000376612	Dataset10.dat	0.000408961
4	0.000376612	Dataset2.dat	0.000379037
4	0.000376612	Dataset3.dat	0.000398987
4	0.000376612	Dataset4.dat	0.000418936
4	0.000376612	Dataset5.dat	0.000359088
4	0.000376612	Dataset6.dat	0.000389012
4	0.000376612	Dataset7.dat	0.000339139
4	0.000376612	Dataset8.dat	0.000359088
4	0.000376612	Dataset9.dat	0.000279291
5	0.000564919	Dataset1.dat	0.000408961
5	0.000564919	Dataset10.dat	0.000408961
5	0.000564919	Dataset2.dat	0.000379037
5	0.000564919	Dataset3.dat	0.000398987
5	0.000564919	Dataset4.dat	0.000418936
5	0.000564919	Dataset5.dat	0.000359088
5	0.000564919	Dataset6.dat	0.000389012
5	0.000564919	Dataset7.dat	0.000339139
5	0.000564919	Dataset8.dat	0.000359088
5	0.000564919	Dataset9.dat	0.000279291

The result is visualized graphically here [Fig – 7].

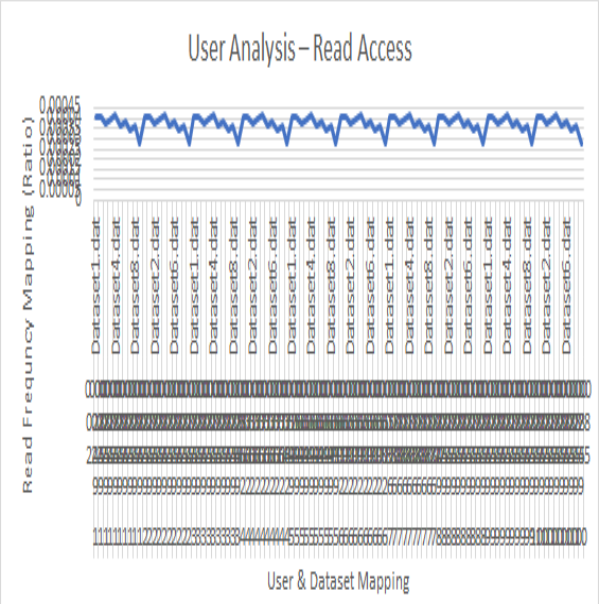

Fig. 7 User Analysis – Read Access

G. Auditing User Analysis – Update Access on VM Data

Further, the auditing user analysis on the virtual machine data is furnished and discussed [Table – 7] with few samples from the result:

TABLE VII
USER ANALYSIS – UPDATE ACCESS

User_Seq	User Update Frequency	VM Dataset	VM Update Frequency
1	0.000941531	Dataset1.dat	0.000548607
1	0.000941531	Dataset10.dat	0.000528657
1	0.000941531	Dataset2.dat	0.00044886
1	0.000941531	Dataset3.dat	0.000528657
1	0.000941531	Dataset4.dat	0.00044886
1	0.000941531	Dataset5.dat	0.000458835
1	0.000941531	Dataset6.dat	0.000398987
1	0.000941531	Dataset7.dat	0.000418936
1	0.000941531	Dataset8.dat	0.000518683
1	0.000941531	Dataset9.dat	0.000548607
2	0.000753225	Dataset1.dat	0.000548607
2	0.000753225	Dataset10.dat	0.000528657
2	0.000753225	Dataset2.dat	0.00044886
2	0.000753225	Dataset3.dat	0.000528657

2	0.000753225	Dataset4.dat	0.00044886
2	0.000753225	Dataset5.dat	0.000458835
2	0.000753225	Dataset6.dat	0.000398987
2	0.000753225	Dataset7.dat	0.000418936
2	0.000753225	Dataset8.dat	0.000518683
2	0.000753225	Dataset9.dat	0.000548607
3	0.000564919	Dataset1.dat	0.000548607
3	0.000564919	Dataset10.dat	0.000528657
3	0.000564919	Dataset2.dat	0.00044886
3	0.000564919	Dataset3.dat	0.000528657
3	0.000564919	Dataset4.dat	0.00044886
3	0.000564919	Dataset5.dat	0.000458835
3	0.000564919	Dataset6.dat	0.000398987
3	0.000564919	Dataset7.dat	0.000418936
3	0.000564919	Dataset8.dat	0.000518683
3	0.000564919	Dataset9.dat	0.000548607
4	0.000470765	Dataset1.dat	0.000548607
4	0.000470765	Dataset10.dat	0.000528657
4	0.000470765	Dataset2.dat	0.00044886
4	0.000470765	Dataset3.dat	0.000528657
4	0.000470765	Dataset4.dat	0.00044886
4	0.000470765	Dataset5.dat	0.000458835
4	0.000470765	Dataset6.dat	0.000398987
4	0.000470765	Dataset7.dat	0.000418936
4	0.000470765	Dataset8.dat	0.000518683
4	0.000470765	Dataset9.dat	0.000548607
5	0.000659072	Dataset1.dat	0.000548607
5	0.000659072	Dataset10.dat	0.000528657
5	0.000659072	Dataset2.dat	0.00044886
5	0.000659072	Dataset3.dat	0.000528657
5	0.000659072	Dataset4.dat	0.00044886
5	0.000659072	Dataset5.dat	0.000458835
5	0.000659072	Dataset6.dat	0.000398987
5	0.000659072	Dataset7.dat	0.000418936
5	0.000659072	Dataset8.dat	0.000518683
5	0.000659072	Dataset9.dat	0.000548607

The result is visualized graphically here [Fig – 8].

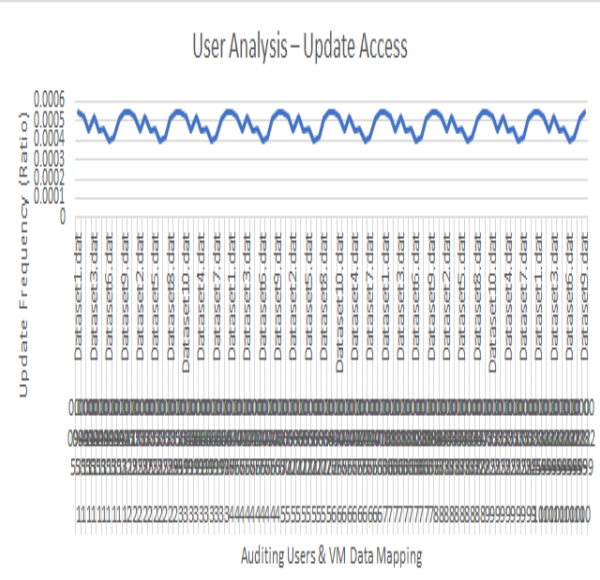

Fig. 8 User Analysis – Update Access

H. Auditing User Analysis – VM Consolidation

Finally, the auditing users & VM consolidation is furnished and discussed [Table – 8]:

TABLE VIII
USER ANALYSIS – VM CONSOLIDATION

User_Seq	VM Data for Read	VM Data for Update
	Access Audits	**Access Audits**
1	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
2	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
3	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
4	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
5	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
6	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
7	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
8	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
9	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat
10	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat	DS1.dat, DS10.dat, DS2.dat, DS3.dat, DS4.dat, DS5.dat, DS6.dat, DS7.dat, DS8.dat, DS9.dat

Henceforth, with the final results of VM to auditing user's consolidation, in the next section of this work, the accuracy of the proposed algorithm is compared with the parallel research outcomes.

VI. COMPARATIVE ANALYSIS

With the detailed analysis of the achieved results from the proposed algorithm components, in this section of the work, the comparative analysis with the parallel research outcomes are discussed [Table – 9].

TABLE IX
COMPARATIVE ANALYSIS

Author Name and Work	VM Data Characteristics Analysis Capabilities and Accuracy (%)	Auditing User Characteristics Analysis Capabilities and Accuracy (%)	VM Consolidation Capabilities and Accuracy (%)
Q. Wang et al. [12]	No	95.33	93.68
J. Yu et al. [13]	No	94	91.23
K. Ren et al. [14]	No	94.43	91.73
H. Wang et al. [15]	No	94.91	95
Proposed Algorithm	99.13	98	100

The result is visualized graphically here [Fig – 9].

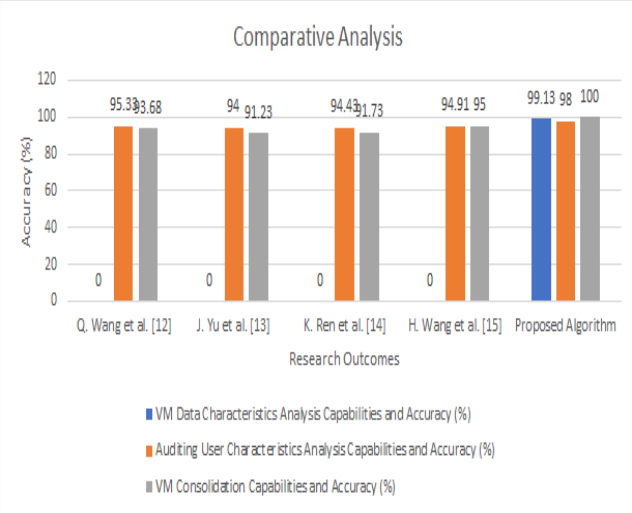

Fig. 9 Comparative Analysis on Accuracy

Henceforth, with the sufficient evidence of the improvements over other parallel research outcomes, this work presents the final conclusion of the research in the next section of this work.

VII. CONCLUSION

The increasing concern of the data security on cloud computing-based data centres, the business of the data centre owners is at a high risk and the challenge of the recent research on this direction is also growing. In order to avoid the security concerns from the application owners, the data centre providers are mostly utilizing the services from the third-party auditing users. Nonetheless, the risk of exposing data to the unauthorised users have not reduced. The parallel research outcomes aiming to reduce the security threats have faced multiple challenges in the reality and been criticised by researchers. Henceforth, this work proposes a novel algorithm to extract and perform the characteristics based secure selective data auditing process for the cloud-based data and the applications on the virtual machine. The proposed algorithm also utilizes the deep learning methods for identifying the core behavioural data access patterns from the auditing users on the same data centres and during a new request process, the same pattern is validated during the access validation process. These algorithm components come into the action during the scenarios where the prior approvals are not granted to access and audit the data files. This novel selective and secure auditor to VM mapper algorithm demonstrates nearly 99.13% accuracy on data characteristics analysis, nearly 98% accuracy on user characteristics analysis and 100% accuracy on secure auditor selection process and builds a much better and trustworthy scenario for the application and data owners.

REFERENCES

1. K. Ren, C. Wang, Q. Wang, "Security challenges for the public cloud", IEEE Internet Comput., vol. 16, no. 1, pp. 69-73, Jan. 2012.
2. G. Ateniese et al., "Provable data possession at untrusted stores", Proc. 14th ACM Conf. Comput. Commun. Secur., pp. 598-609, 2007.
3. A. Juels, B. S. Kaliski, "Pors: Proofs of retrievability for large files", Proc. 14th ACM Conf. Comput. Commun. Secur., pp. 584-597, 2007.
4. H. Shacham, B. Waters, "Compact proofs of retrievability", J. Cryptol., vol. 26, no. 3, pp. 442-483, Jul. 2013.
5. C. Wang, S. S. M. Chow, Q. Wang, K. Ren, W. Lou, "Privacy-preserving public auditing for secure cloud storage", IEEE Trans. Comput., vol. 62, no. 2, pp. 362-375, Feb. 2013.
6. S. G. Worku, C. Xu, J. Zhao, X. He, "Secure and efficient privacy-preserving public auditing scheme for cloud storage", Comput. Electr. Eng., vol. 40, no. 5, pp. 1703-1713, 2014.
7. C. Guan, K. Ren, F. Zhang, F. Kerschbaum, J. Yu, "Symmetric-key based proofs of retrievability supporting public verification" in Computer Security—ESORICS, Cham, Switzerland:Springer, pp. 203-223, 2015.
8. W. Shen, J. Yu, H. Xia, H. Zhang, X. Lu, R. Hao, "Light-weight and privacy-preserving secure cloud auditing scheme for group users via the third party medium", J. Netw. Comput. Appl., vol. 82, pp. 56-64, Mar. 2017.
9. J. Sun, Y. Fang, "Cross-domain data sharing in distributed electronic health record systems", IEEE Trans. Parallel Distrib. Syst., vol. 21, no. 6, pp. 754-764, Jun. 2010.
10. G. Ateniese, R. D. Pietro, L. V. Mancini, G. Tsudik, "Scalable and efficient provable data possession", Proc. 4th Int. Conf. Secur. Privacy Commun. Netw., 2008.
11. C. Erway, A. Küpçü, C. Papamanthou, R. Tamassia, "Dynamic provable data possession", Proc. 16th ACM Conf. Comput. Commun. Secur., pp. 213-222, 2009.
12. Q. Wang, C. Wang, K. Ren, W. Lou, J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing", IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847-859, May 2011.
13. J. Yu, K. Ren, C. Wang, V. Varadharajan, "Enabling cloud storage auditing with key-exposure resistance", IEEE Trans. Inf. Forensics Security, vol. 10, no. 6, pp. 1167-1179, Jun. 2015.
14. J. Yu, K. Ren, C. Wang, "Enabling cloud storage auditing with verifiable outsourcing of key updates", IEEE Trans. Inf. Forensics Security, vol. 11, no. 6, pp. 1362-1375, Jun. 2016.
15. J. Yu, H. Wang, "Strong key-exposure resilient auditing for secure cloud storage", IEEE Trans. Inf. Forensics Security, vol. 12, no. 8, pp. 1931-1940, Aug. 2017.
16. J. Yu, R. Hao, H. Xia, H. Zhang, X. Cheng, F. Kong, "Intrusion-resilient identity-based signatures: Concrete scheme in the standard model and generic construction", Inf. Sci., vol. 442, pp. 158-172, May 2018.
17. B. Wang, B. Li, H. Li, "Oruta: Privacy-preserving public auditing for shared data in the cloud", Proc. IEEE 5th Int. Conf. Cloud Comput. (CLOUD), pp. 295-302, Jun. 2012.
18. G. Yang, J. Yu, W. Shen, Q. Su, Z. Fu, R. Hao, "Enabling public auditing for shared data in cloud storage supporting identity privacy and traceability", J. Syst. Softw., vol. 113, pp. 130-139, Mar. 2016.