An Optimal Virtual Machine Placement Method in Cloud Computing Environment

Original Scientific Paper

Ashalatha Ramegowda
Faculty, Department of Computer Science
Gulbarga University, Kalaburagi
Karnataka, India
ashalatha.dsce@gmail.com

Abstract *– Cloud computing is formally known as an Internet-centered computing technique used for computing purposes in the cloud network. It must compute on a system where an application may simultaneously run on many connected computers. Cloud computing uses computing resources to achieve the efficiency of data centres using the virtualization concept in the cloud. The load balancers consistently allocate the workloads to all the virtual machines in the cloud to avoid an overload situation. The virtualization process implements the instances from the physical state machines to fully utilize servers. Then the dynamic data centres encompass a stochastic modelling approach for resource optimization for high performance in a cloud computing environment. This paper defines the virtualization process for obtaining energy productivity in cloud data centres. The algorithm proposed involves a stochastic modelling approach in cloud data centres for resource optimization. The load balancing method is applied in the cloud data centres to obtain the appropriate efficiency.*

Keywords*: Cloud computing, Virtualization, Stochastic modeling, Energy efficiency, Cloud service provider, Resource optimization*

1. INTRODUCTION

Cloud is a large server for storing different services and data of the users. Cloud is a concept of using services not stored on your computer. The virtualization process consolidates many workloads into smaller physical servers in the data centres of the cloud to meet the Service Level Agreement (SLA) standards using Virtual Machines (VMs) [1]. The Virtualization process allows multiple users to share a physical server. Major companies include VMware, Hyper-V, HP, F5, Nuage, Nicira, etc. The virtualization technology uses a Virtual Machine Monitor (VMM) at the software level for abstraction purposes [2]. The Cloud Service Providers (CSPs) make the infrastructure as a Service (IaaS) for cloud consumers. They use VMs and Virtual Clusters (VCs) for computing cloud resources. Primary cloud providers include Amazon EC2 and IBM [3].

The user, service, and cloud computing infrastructure are significant entities involved in the cloud environment. Primary attacks in the cloud can be service attacks, including browser attacks, phishing attacks, and SSL certificate spoofing attacks. User attacks can be accomplished during spoofing or the cloud infrastructure services. Significant Quality of Service (QoS) matter has service availability considered a cloud environment. Therefore, the virtual data centre is becoming popular because of providing IaaS in the field of the cloud [4]. IaaS is the most crucial delivery model developed in cloud computing. IaaS offers virtual infrastructure like servers and data storage. Cloud providers can use virtualization techniques for virtual data centres in cloud computing. Amazon is an excellent example of providing a cloud platform that offers infrastructure at an affordable price to its customers [5].

A VM is a single computer with a dedicated platform environment with limited resources. The resource virtualization process is the simplified version of traditional resource management. The virtualization system encourages replication procedures in the cloud to perform elasticity processes within a given system. They run on a hardware platform controlled by VMM. Each VM runs under VMM, which can change the virtual status from one data centre to another. A VMM is hypervisor software that divides the computation resources into the number of VMs through the guest operating system [6]. VMM has various operating systems to execute particular hardware simultaneously for resource isolation. The significant benefits of using VMM include high-security performance using multiple services simultaneously. VMM uses a Live VMM scheme to transfer the cloud server from one hardware platform to another using system modification. Mobile devices consume less energy to achieve resource optimization in the cloud network. The parameter metrics to consider are data size and delay constraints for optimal

solutions in the mobile cloud. Different energy optimal models for mobile devices include:

- The mobile execution models.
- The execution model for the cloud.
- The optimal application execution policy model.

This research aims to observe the energy efficiency of cloud data centers using a stochastic modelling approach. The principal enrichment of this paper is as follows.

- Increasing energy productivity in cloud data centers using a resource optimization approach.
- Load balancing scheme using resource virtualization method.
- A stochastic modeling method for energy efficiency in cloud data centers.

The remaining part of this work is as follows. Section 2 presents the overall literature survey part. Section 3 depicts a system model with architecture. Section 4 details the methodology part. Section 5 provides the performance analysis, and section 6 concludes the proposed work.

2. RELATED WORK

An enormous amount of literature has been reviewed for energy efficiency for achieving high performance in cloud computing. The stochastic models accomplish service requests and maintenance of the server. Zhang et al. provide a Markov chain process scheme for optimum policy schedule and scaling problems of cloud servers [7]. CSPs must reduce the energy usage in cloud data centres, and a reliable system can produce with less cost for operation purposes.

Xia et al. have presented energy efficiency in the cloud data centre using the VM migration process's stochastic method for high performance [8]. Han et al. have used the VMM policy in cloud data centres to achieve high performance and robustness. A dedicated stochastic process model has been used for energy efficiency with high production [9].

Ait Salaht et al. have used a technique called the hysteresis queuing model, which is used for cloud data centres. The stochastic bounding models provide performance analysis in the cloud [10]. Anastasopoulos et al. have operated the optical networks and cloud infrastructures considered for service provision in the cloud. A stochastic linear-based programming approach has been used for resource provisioning in the cloud to evaluate and use renewable sources [11].

Ghosh et al. have used cloud host services to reduce costs in the data centres. The VMs provide IaaS cloud service, which shares the instances of Physical Machines (PMs) instances within cloud data centres. An optimal PM has been chosen to minimize operational costs with excellent infrastructure in the cloud. The stochastic model has been used to analyze value and optimize the framework within the IaaS cloud [12].

Zhou et al. use cloud computing with a sophisticated infrastructure and comprehensive data-sharing service. The stochastic process with high-quality evaluation and modelling has been proposed in work. The assessment of the IaaS cloud performs quality metrics based on the criteria such as completion time of user requests, system overhead rate and rejection time probability [13].

Maguire et al. provide the stochastic analysis model, which uses the load balancing process and the VM and is a scheduled concept of cloud. In Cloud Computing Cluster (CCC) theory, every job uses VMs under a stochastic process. A dedicated algorithm for load balancing and VM scheduling is analyzed for high capacity within the system [15]. Chen et al. say a scalable and flexible approach is designed for high-scale cloud computing. The multi-data centre model provides substantial data processing for large applications and top computing resources.

High performance attained by using workload scheduler under VM in cloud data centres. The QoS-based approach model is designed for energy efficiency and resource optimization [16]. Vasileios et al. have presented an intelligent city framework using the Internet of Things [17]. Modern computing techniques are adopted to increase usability and are also helpful in preserving confidential details [18]. Tahar et al. use a VM placement scheme for cloud data centres. The integer linear model saves energy and hence increases QoS [19].

Every VM is decided through a scheduling strategy to achieve budgetary deadlines. Resource provisioning is provided for accessing the tasks with execution time. The stochastic-based scheme uses multi-objective scheduling criteria for making energy-efficient in a cloud [20]. A stochastic approach is used for energy and cost minimization purposes. LP and LDPP schemes are used for cost-savings sake. CCDF method has been proposed for high performance. A stochastic optimization model is given for green data centres [21]. Stochastic Petri nets are used for QoS and any time system availability. The VM placement strategy achieves good accuracy in the cloud. The SRN model performs VM migration and placement strategy using proposed algorithms [22].

3. SYSTEM MODEL

Cloud system network varies from traditional network distributed systems. They are characterized by many resources that can span different administrative domains. Various clouds appropriate to one particular or different organizations can dynamically join each other to achieve a common objective, usually represented using the optimization of cloud resources. This method is known as cloud federation [23]. The system architecture comprises data users, owners, trusted authority, cloud proxy, and Third-Party Auditor (TPA). The cloud users are connected to share multimedia files

through the wireless access point to the cloud proxy server. The cloud data owner uses data file ciphertext to store the contents of the cloud data centres. TPA is used in cloud data centres to achieve additional security purposes.

The trusted authority uses privilege data management requests and privilege update requests by using public and attributes keys in cloud data centres. TPA is used in cloud data centres to achieve additional security purposes. The trusted authority uses privilege data management requests and privilege update requests by using public and attributes keys in cloud data centres. The system security model for public auditing scheme cloud servers, data users, and TPA [24]. The proposed system model represents computer systems composed of many resources, making it possible to describe physical and virtual resources. Each system's general stochastic data model comprises M = m X N VMs running parallel. Here, m refers to the number of virtual machines in the cloud.

M is the maximum number of VMs running in a particular method of a cloud data centre for resource optimization. The stochastic process maintains three primary servers in the cloud. They include images, video and document servers [25].

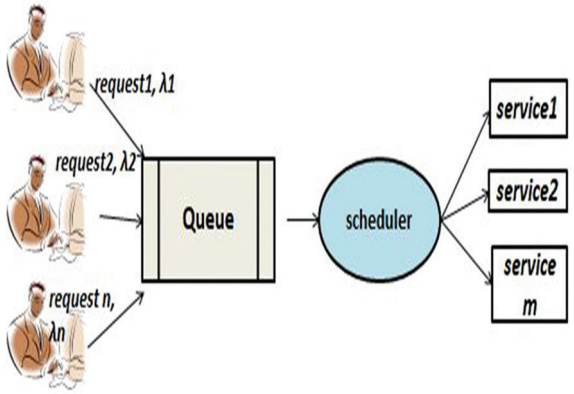

Fig. 1. Basic system model

Figure 1 depicts the overall cloud system security model for various services and user requests. The significant entities included in the system architecture are cloud users, input queue, resource scheduler, and many services to be computed in the cloud server [26]. It depicts the queuing performance model in the cloud for the service requests to be performed. Figure 2 represents the proposed system storage model in cloud systems. Primary operations involved in the model include cloud users, data owners, trusted authority, cloud proxy servers and TPA.

A data owner is responsible for generating the encrypted files and uploading the files to the cloud server [27]. Trusted authority checks for the incoming requests and sends the required key to the data owner. The proxy server requests the needed key, gets the essential key, selects the file to encrypt, decrypts the confidential data, downloads the received files and shares the files with the cloud users [28]. TPA can view all the files on the server and perform the auditing process for cloud security. The overall system architecture has the following stages.

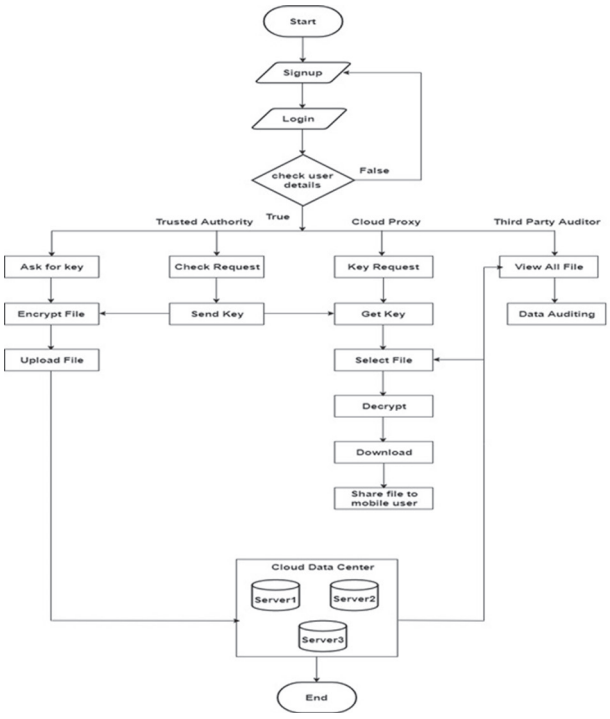

Fig. 2. System storage model

Figure 2 represents the system storage model that includes the following steps.

i. Data owner who uploads the data files to the cloud data centre in an encrypted manner.
ii. Cloud proxy who shares files from cloud server for cloud users.
iii. A third-party auditor who is used for maintaining the cloud data files
iv. A trusted authority that provides the attribute key for cloud proxy for file sharing.
v. Data users could be mobile phones, laptops, personal computers or Tablet PCs.

4. METHODOLOGY

The dynamic load balancing (DLB) approach uses only the present machine state for balancing the current workloads to achieve high-performance satisfaction and complete deployment of cloud resources. The load balancing scheme for data centres improves the energy efficiency of the cloud resources in the cloud. DLB is a method that distributes scalable workloads evenly among all the system nodes in the cloud, used to create new instances [29]. The massive amount of energy from the industry and companies leads to high-cost cloud data centres. Thus, the cloud data centres must change according to the energy used to gain energy efficiency using the virtualization process. A

stochastic model that uses the queuing theory concept is used to achieve high performance and energy consumption. Dynamic right-sizing of the data centres can be gained using stochastic modelling in cloud data centres [30]. The cloud maintains the central storage of data in its remote data centres. Data management has been made easier through cloud storage. The queuing model concept applies to managing and storing applications in the cloud [31].

4.1 SYSTEM ARCHITECTURE

The following section describes the overall system architecture for the stochastic system process. Figure 3 represents the stochastic system model. It includes a load balancing process, resource optimization and a data centre management module. VM scheduler works on VM section strategy using placement process. The cloud data centres use hypervisors for VM management [32].

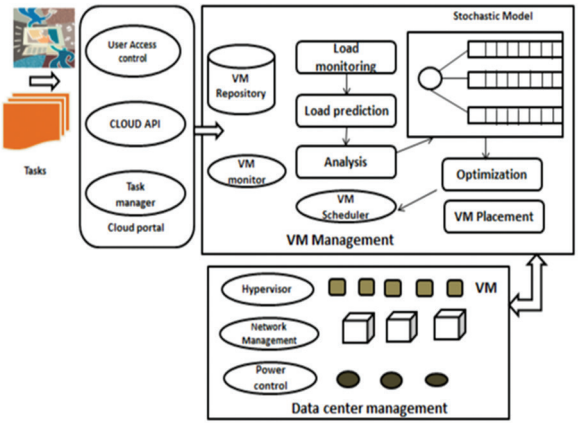

Fig 3. System architecture

4.2 MODEL ANALYSIS AND SOLUTION

The steady-state includes the probability distribution analysis method, which accomplishes essential information about the given data centre model.

Let J (t) is given as number of VMs for the system at time t.

Let L (t) is given as the number of working servers for the time t, then

L(t) and J(t) provides the input with a process with the given state space.

$$\Omega = \{(i,j): 0 \le j \le M-1, 0 \le i \le j\} \cup \{(i,j): j \ge M, 0 \le i \le N\} \quad (1)$$

The stochastic process solution includes the following entities.

Let X(*t*) be the stochastic process where t ≥ 0

S: finite state space

R: Reward function

π_i: Steady-state probability of state S_i

The stochastic modelling is a process of evolution, where $r_{X(t)}$ gives the stochastic process using system reward rate at time t.

$r(r:S \rightarrow R)$ is given as reward function

Probability π_i allows $i \in S$ as steady state and *r(i)* denotes reward and is written as r_i

The model solution has $t(\pi_i(t))$ and $t_i(t)$ presents the expected reward rate given in equation 2.

$$E[r_x] = \sum_{i \in S} r_i . \pi_i \quad (2)$$

4.3 MARKOVIAN CHAIN PROCESS

The following equation gives the markovian continuous model.

$$w, P1, P2, \ldots, Pm, V1, V2, \ldots, Vm, F1, F2, \ldots, Fm \quad (3)$$

Where, w is the number of VM requests in the waiting queue

P_i is the position of the virtual machine level

V_i is the number of virtual machines

F_i=0 means the virtual machine is alive in the process else failure.

The queuing theory scheme can be used for modelling the data and applications in the cloud. The model solutions are built using an analytic method using the probability vectors. The model analysis provides a key by using an M/M/N queuing model using a Markovian chain process in the cloud environment [33]. This model helps develop the algorithm for a specific VM during the execution time of the cloud resources.

4.4 STEADY STATE VECTOR

Here, πij is the probability vector stating that there are j VMs on the system and i working servers. The steady state vector is given as π and can be considered as:

$$\begin{matrix} \{\Pi Q = 0 \\ \sum_{j=0}^{\infty} \pi j e = 1\} \end{matrix} \quad (4)$$

The steady state vector is given by:

$$\begin{matrix} \Pi = [\Pi_0 \Pi_1 \Pi_2 - - - - \Pi_N] \\ \Pi_j = [\Pi_0 \Pi_1 \Pi_2 - - - - \Pi_{jj}] \text{ for j < N+1} \\ \Pi_j = [\Pi_{0j} \Pi_{1j} \Pi_{2j} - - - - \Pi_{Nj}] \text{for j > N} \end{matrix} \quad (5)$$

4.5 THE STOCHASTIC MODEL

Each system's general stochastic data model comprises M = m X N VMs running parallel. Here, m refers to the number of virtual machines in the cloud. M is the maximum number of VMs running in a particular method of a cloud data centre for resource optimization [34]. The stochastic process maintains three central servers in the cloud. They include images, video and document servers [35]. Here, each server can consider one of the following states.

i. serving the VMs. The service rate always gives the energy spent in this specific state. Suppose

there is m' VMs present where m' ≤ m is the number of cloud servers running in the system, the CPU utilization can be given as μm'. The energy consumption can be provided by:

$$\mu m' P(on) \quad (6)$$

ii. *OFF state*: Here, the server is not serving any VMs. The image, video and document servers are made off as there are insufficient VMs. Hence, the energy consumption for any given server is 0.

ii. *Setup state*: The state, which goes from off state to on form, is a setup state. Here, all the servers are made for file accessing purposes. Therefore, the energy consumption for the given server is given as P_{on}.

ii. *Failed state*: The server has failed due to some erroneous failure in the server or the data centres. The system can crash due to some catastrophic losses or damage due to disasters, and hence, the server goes to a failed state. Here, energy consumption can be given as P_{fail}.

The performance efficiency can be achieved using scalable analytic models for cloud resources. The optimization algorithms are used in cloud data centres to achieve optimal values such as service rate μ, optimized servers N and the performance function F (μ, N). The dynamic workflow uses a critical path VM selection strategy for optimization sequence. The resource optimization algorithm is defined in algorithm 4.1.

Algorithm 4.1: Resource optimization algorithm

Step 1: Procedure measure1

Step 2: Input: performance metric function F (μ, N), N, λ, θ, with an initial point μ_0 and a positive tolerance Δ.

Step 3: Output: Approximate solution μ*

Step 4: Calculate unit matrix $H=I_n$ (7)

Step 5: Compute the gradient
$g_0=\nabla F(\mu_0,N)$ at point $x_0=\mu_0$

Step 6: Set k to 0

Step 7: While $|\nabla F(x_k+1)\leq\Delta|$ do (8)

Step 8:Generate the search direction
d_k=-H_k g_k (9)

Step 9: Search along d_k from point X_k, find the step-length $\propto_k$ by satisfying
$F(X_k+\propto_k d_{-k})=\min\{F(X_k+\propto_k d_k)\}$ (10)

Step 10: let $g_{k+1}=\nabla F(X_k+1)$, $p_k=X_{k+1}-X_k$, $q_k=g_{k+1}-g_k$ (11)

Step 11: $H_{k+1}=H_k+1+p_k^T q_k/q_k^T H_k q_k)$ (12)

Step 12: $X_{k+1}=X_k+\alpha_k d_k$ (13)

Step 13: $k=k+1$

Step 14: end while

Step 15: return X_k

Step 16: end Procedure

Step 17: Stop

The optimization procedures are used in algorithms using various measures. The optimal solution can be found using an optimization procedure to see the complexity of the optimization algorithms. The main factors used are variables, validity and effectiveness in our optimization algorithm. The resource optimization algorithm for stage 2 is given in the algorithm.

Algorithm 4.2: Optimization algorithm

Step 1: Procedure measure2

Step 2: Input: F (μ, N), μ*, λ, θ and M, where M is a sufficient large number

Step 3: Output: Approximate solution (μ*, N*)

Step 4: F*= ∞

Step 5: for (N=1;N<M; N++) do

Step 6: Use algorithm 1 to calculate F^* (mu^*, N)

Step 7: if ($F^*(mu^*, N) < F^*$ then

Step 8: $F^*=F^*(mu^*, N)$

Step 9: S= (mu^*, N)

Step 10: end if

Step 11: end for

Step 12: return S

Step 13: end procedure

Step 14: Stop

The resource optimization algorithm is used to gain optimal values of service rate μ, and optimal server N. The process uses a minimum performance metric function to achieve efficiency.

4.6 SYSTEM MODULES

The system design comprises the following system modules in cloud data centres.

i. *Load monitoring module*: The load monitoring module is used for calculating the statistics of λ, θ and σ for the resource optimization in the cloud system.

ii. *Load prediction module*: The load prediction module predicts the actual load by using λ (i + 1) for the next optimization period Δ.

iii. *Analysis and optimization*: These modules are used to implement the mathematical model and solution for resource optimization. The mathematical system model for this system design has been presented in Table 1.

Table 1: Mathematical model

μ	Service rate
m'	Virtual machines
N	Optimal web servers
E (W)	Execution time
E (P)	Energy consumption
F(μ, N)	Metric function
μ*	Optimal value

5. PERFORMANCE ANALYSIS

The efficiency of the proposed work is shown through performance metrics which are defined as follows.

i. *Accessing data files*

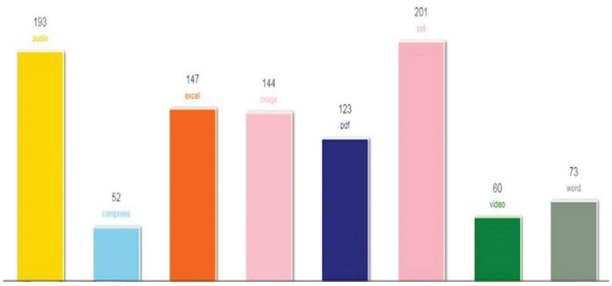

Fig. 4. Proxy accessing cloud

Figure 4 depicts the number of users accessing a proxy server for obtaining various types of files in the cloud. Cloud proxy can view the uploaded files and request for the file to download. Once the 'key' is requested, the request is sent to the trusted authority. Data files can be accessed more efficiently using a proxy server than the actual cloud. A cloud proxy server can view the uploaded files and request for the file to download. Once the key is requested, the request is sent to the trusted authority. Various multimedia servers involve video, word, pdf, ppt, image, excel, compress, audio servers etc.

ii. *Downloading multimedia files*

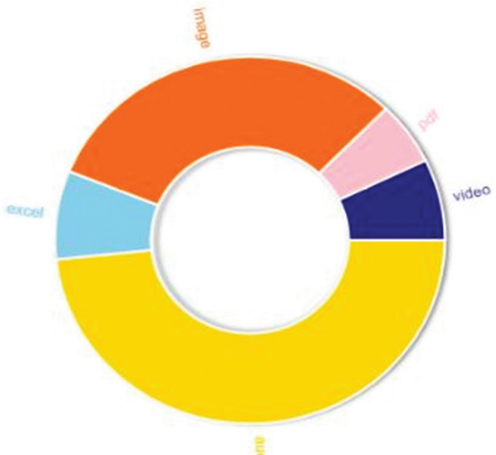

Fig. 5. Users accessing proxy server

Figure 5 depicts the number of users using a proxy server for downloading the data. The data owner can upload the data in a given format such as video, audio, image or document format. The process of downloading multimedia files can be made through a cloud proxy. The pie chart represents various multimedia files used for downloading from the cloud proxy server, including documents, video, audio, image servers, etc.

iii. *Resource access time with several cloud users*

In the following figure, the resource access audit time from the cloud proxy server is compared with access time from the cloud server. Bandwidth has been used broader, and user access has now improved through the proposed method. The audit access time of the cloud data can be reduced relatively by using the proxy server than the cloud server, as shown in Figure 6.

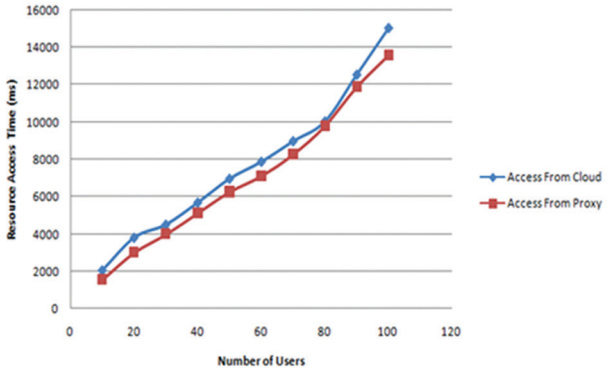

Fig. 6 Resource access auditing time

In the comparison graph of Figure 6, the resource access time from the cloud proxy server is compared to that of access time from the cloud server in milliseconds. In this approach, bandwidth has been used higher, and user access is improved through the proposed method. The trusted authority approves the file request from the proxy cloud. The cloud proxy can receive the mail with the token and attribute key to decrypt in registered mail id. Once the cloud proxy gets the mail, it can solve and view the file. And cloud proxy has additional work, which means they must upload the file to access the end mobile user. The Android mobile app is created to view and access this detailed data for accessing the file.

iv. *Virtualization graph*

The VMM policy can be used to transfer the VMs taken from one DC to another in the distributed systems. To evaluate the system behaviour in multiple data centres, the analysis of VMM time and balancing of the incoming load is required. The virtualization graph specifies resource type and the number of files in the cloud. It analyses which resource type occupies more on cloud and proxy. If the resources are stored in the proxy cloud, the speed will automatically increase. Here, the proxy cloud acts as a virtual server. The consolidation ratio denotes the virtual servers running independently on the host machine.

The number of multimedia resources, their count, and specific resource type are mentioned here. Figure 7 depicts the virtualization process within the cloud data centres. The graph shows the number of users accessing proxy servers for accessing various types of files in the cloud. Cloud proxy can view the uploaded files and request for the file to download. The demand for the specific file is sent to the trusted authority by the cloud proxy server.

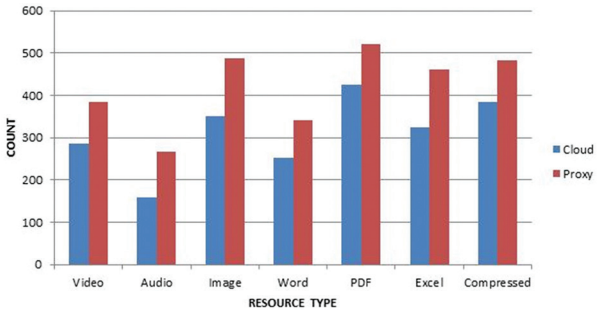

Fig 7. Virtualization graph

v. *Energy consumption*

The energy usage in various DCs can be reduced using the dynamic optimization scheme. The energy consumption graph is given between the resource name and the energy taken to access the resource. The chart emulates the cloud's energy consumption and the cloud proxy server to access the resource.

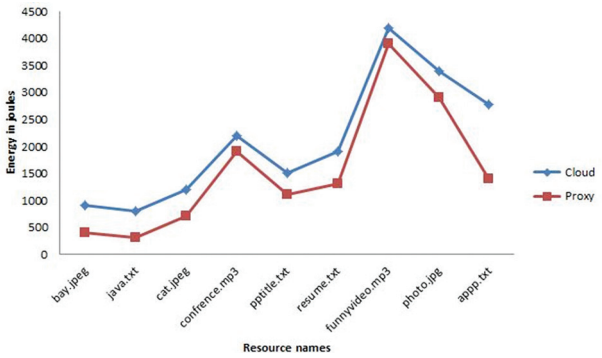

Fig. 8. Energy consumption

The stochastic type programming models reduce the energy consumption in large DCs in terms of joules. Figure 8 represents the energy consumption in the cloud centres for various files and applications. The numbers of resource names and the energy used by the cloud server are depicted in the figure.

vi. *Internet load monitoring*

Load Monitoring analyses resource size and time taken to access the particular resource in that size. It gives a more precise picture of how long it will take to obtain the specific volume.

Figure 9 gives the load monitoring process with the file size in kilobytes. The time taken in milliseconds from each file in the cloud is given. A large number of resource types with size, along with the time taken to load the files onto a cloud server, are shown in the graph.

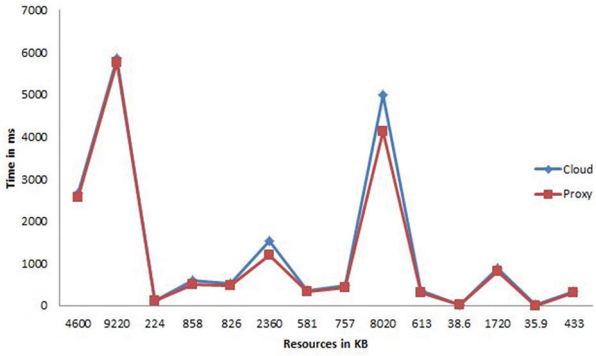

Fig. 9. Load monitoring

viii. *Load balancing*

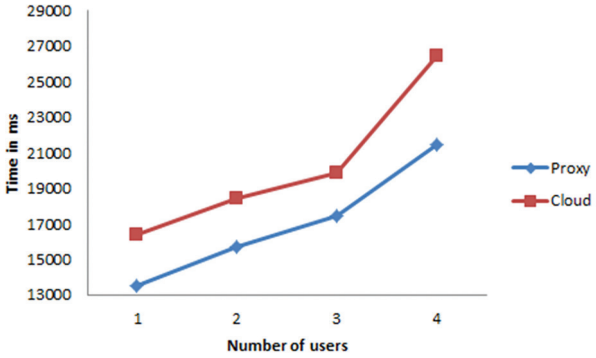

Fig. 10. Load balancing

Figure 10 depicts the load balancing feature involving several VMs in the cloud and proxy server. Time has been calculated in milliseconds for different VMs. The result shows that the proxy server takes less time than the cloud server.

ix. *Resource utilization comparison*

Figure 11 compares the resource utilization for both proxy and cloud servers using a virtualization system. The result shows that a proxy server utilizes less time for considering the cloud resources than a cloud in milliseconds.

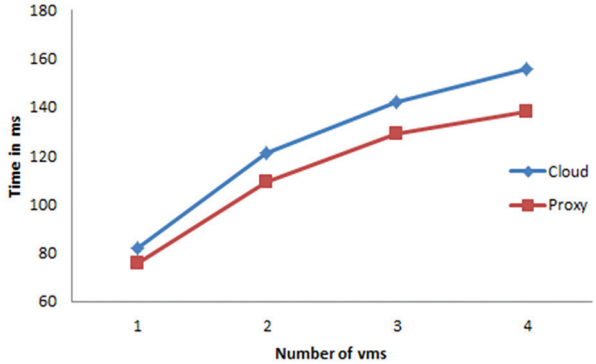

Fig. 11. Resource utilization

6. CONCLUSION

Cloud security has become a significant concern these days in the computing world. Computing and communication security has been taken into consideration by substantial researchers. The availability of cloud data services may fail anytime due to power failures or the occurrence of any catastrophic failures. A third-party error can happen anytime in the cloud due to limited transparency and user control. Hence, it remains paramount to outline the cloud and virtualization process before examining energy efficiency for cloud data centres. The data centre resource management is dealt with and considered a significant critical cloud computing problem. The load balancing approach uses stochastic data modelling for resource optimization in cloud data centres. The VM placement strategy reduces the total energy consumed and undetermined requirements by many cloud servers.

7. REFERENCES:

[1] R. Ashalatha, J. Agarkhed, S. Patil, "Network virtualization system for security in cloud computing", Proceedings of the 11th International Conference on Intelligent Systems and Control, Coimbatore, India, 5-6 January 2017, pp. 346-350.

[2] M. Roohitavaf, R. Entezari-Maleki, A. Movaghar, "Availability Modeling and Evaluation of Cloud Virtual Data Centers", Proceedings of the International Conference on Parallel and Distributed Systems, Seoul, Korea, 15-18 December 2013, pp. 675-680.

[3] R. Ghosh, K. S. Trivedi, V. K. Naik, D. S. Kim, D "End-to-end performability analysis for infrastructure-as-a-service cloud: An interacting stochastic models approach", Proceedings of the IEEE 16th Pacific Rim International Symposium on Dependable Computing, 2010 pp. 125-132.

[4] X. Chang, R. Xia, J. K. Muppala, K. S. Trivedi, J. Liu, "Effective modeling approach for iaaS data center performance analysis under heterogeneous workload", IEEE Transactions on Cloud Computing, Vol. 6, No. 4, 2016, pp. 991-1003.

[5] B. Wei, C. Lin, X. Kong, "Dependability modeling and analysis for the virtual data center of cloud computing", Proceedings of the IEEE 13th International Conference on High Performance Computing and Communications, 2011, pp. 784-789.

[6] W. Zhang et al. "Energy-optimal mobile cloud computing under stochastic wireless channel", IEEE Transactions on Wireless Communications, Vol. 12, No. 9, 2013, pp. 4569-4581.

[7] P. Zhang, C. Lin, K. Meng, Y. Chen, "A Comprehensive Optimization for Performance, Energy Efficiency, and Maintenance in Cloud Datacenters", Proceedings of the Trustcom/BigDataSE/I SPA, Tianjin, China, 23-26 August 2016, pp. 1264-1271.

[8] Y. Xia, M. Zhou, X. Luo, S. Pang, Q. Zhu, "A stochastic approach to analysis of energy-aware DVS-enabled cloud datacenters", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 45, No. 1, 2015, pp. 73-83.

[9] Y. Xia, Y. Han, M. Zhou, J. Li, "A stochastic model for performance and energy consumption analysis of rejuvenation and migration-enabled cloud", Proceedings of the International Conference on Advanced Mechatronic Systems, 2014, pp. 139-144.

[10] F. Aït-Salaht, H. Castel-Taleb, "Stochastic bounding models for performance analysis of clouds", Proceedings of the IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015 pp. 603-610.

[11] M. Anastasopoulos, A. Tzanakaki, D. Simeonidou, "Stochastic energy efficient cloud service provisioning deploying renewable energy sources", IEEE Journal on Selected Areas in Communications, Vol. 34, No. 12, 2016, pp. 3927-3940.

[12] R. Ghosh et al. "Stochastic model driven capacity planning for an infrastructure-as-a-service cloud", IEEE Transactions on Services Computing, Vol. 7, No. 4, 2014, pp. 667-680.

[13] Y. Xia et al. "Stochastic modeling and quality evaluation of infrastructure-as-a-service clouds", IEEE Transactions on Automation Science and Engineering, Vol. 12, No. 1, 2015, pp. 162-170.

[14] S. El Kafhali, K. Salah, "Stochastic modelling and analysis of cloud computing data center", Proceedings of the 20th Conference on Innovations in Clouds, Internet and Networks, 2017, pp. 122-126.

[15] S. T. Maguluri, R. Srikant, L. Ying, "Stochastic models of load balancing and scheduling in cloud

computing clusters", Proceedings of the IEEE INFOCOM Conference, 2012, pp. 702-710.

[16] Y. Chen et al. "Stochastic workload scheduling for uncoordinated datacenter clouds with multiple QoS constraints", IEEE Transactions on Cloud Computing, Vol. 8, No. 4, 2016, pp. 1284-1295.

[17] V. A. Memos, K. E. Psannis, Y. Ishibashi, B. Kim, B. B. Gupta, "An efficient algorithm for media-based surveillance system (EAMSuS) in IoT smart city framework", Future Generation Computer Systems, Vol. 83, 2017, pp. 619-628.

[18] B. Gupta, D. P. Agrawal, S. Yamaguchi, "Handbook of research on modern cryptographic solutions for computer and cyber security", IGI Global, 2016.

[19] A. Ouammou, M. Hanini, S. El Kafhali, A. B. Tahar, "Energy Consumption and Cost Analysis for Data Centers with Workload Control", Proceedings of the International Conference on Innovations in Bio-Inspired Computing and Applications, 2017, pp. 92-101.

[20] Y. Gao, L. C. Canon, F. Vivien, Y. Robert, "Scheduling stochastic tasks on heterogeneous cloud platforms under budget and deadline constraints", Proceedings of the IEEE International Conference on Cluster Computing, Albuquerque, NM, USA, 23-26 September 2019.

[21] A. Ghassemi, P. Goudarzi, M. R. Mirsarraf, T. A. Gulliver, "A Stochastic Approach to Energy Cost Minimization in Smart-Grid-Enabled Data Center Network", Journal of Computer Networks and Communications, 2019.

[22] Y. Liu et al. "Adaptive Evaluation of Virtual Machine Placement and Migration Scheduling Algorithms Using Stochastic Petri Nets", IEEE Access, Vol. 7, 2019, pp. 79810-79824.

[23] D. Bruneo, "A stochastic model to investigate data center performance and QoS in IaaS cloud computing systems", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 3, 2014, pp. 560-569.

[24] R. Ashalatha, J. Agarkhed, "Dynamic load balancing methods for resource optimization in cloud computing environment", Proceedings of the Annual IEEE India Conference, 2015, pp. 1-6.

[25] D. Shen et al. "Stochastic modeling of dynamic right-sizing for energy-efficiency in cloud data centers", Future Generation Computer Systems, Vol. 48, 2015, pp. 82-95.

[26] Y. Yagawa, A. Sutoh, E. Malamura, T. Murata, "Modeling and Performance Evaluation of Cloud on-Ramp by utilizing a Stochastic Petri-net", Proceedings of the 5th IIAI International Congress on Advanced Applied Informatics, 2016, pp. 995-1000.

[27] A. Uchechukwu, K. Li, K. Li, "Scalable Analytic Models for Performance Efficiency in the Cloud", Proceedings of the IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014, pp. 998-1003.

[28] B. Silva, P. Maciel, A. Zimmermann, "Performability models for designing disaster tolerant infrastructure-as-a-service cloud computing systems", Proceedings of the 8th International Conference for Internet Technology and Secured Transactions, 2013, pp. 647-652.

[29] Y. Tian, C. Lin, Z. Chen, J. Wan, X. Peng, "Performance evaluation and dynamic optimization of speed scaling on web servers in cloud computing", Tsinghua Science and Technology, Vol. 18, No. 3, 2013, pp. 298-307.

[30] J. Wang, S. Shen, "Risk and energy consumption tradeoffs in cloud computing service via stochastic optimization models", Proceedings of the IEEE/ACM Fifth International Conference on Utility and Cloud Computing, pp. 239-246.

[31] M. Ranjbari, M., & J. A. Torkestani, J. A. "A learning automata-based algorithm for energy and SLA efficient consolidation of virtual machines in cloud data centers", Journal of Parallel and Distributed Computing, Vol. 113, 2018, pp. 55-62.

[32] I. Narayanan, D. Wang, A. Sivasubramaniam, H. K. Fathy, "A Stochastic Optimal Control Approach for Exploring Tradeoffs between Cost Savings and Battery Aging in Datacenter Demand Response", IEEE Transactions on Control Systems Technology, Vol. 26, No. 1, 2018, pp. 360-367.

[33] T. Li, B. B. Gupta, R. Metere, "Socially-conforming cooperative computation in cloud networks", Journal of Parallel and Distributed Computing, Vol. 117, 2018, pp. 274-280.

[34] Y. Pan et al. "A Novel Approach to Scheduling Workflows Upon Cloud Resources with Fluctuating Performance", Mobile Networks and Applications, 2020, pp. 1-11.

[35] J. Zhou et al. "Stochastic Virtual Machine Placement for Cloud Data Centers Under Resource Requirement Variations", IEEE Access, Vol. 7, 2019, pp. 174412-174424.