

Research Article

Spectral Expansion Method for Cloud Reliability Analysis

K. Kotteswari ¹ and A. Bharathi ²

¹Department of Computer Science and Engineering, Annai Mira College of Engg & Techn., Vellore, India

²Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam, India

Correspondence should be addressed to K. Kotteswari; kottikarthikeyan17@gmail.com

Received 28 May 2019; Accepted 6 August 2019; Published 2 September 2019

Guest Editor: Mazdak Zamani

Copyright © 2019 K. Kotteswari and A. Bharathi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cloud computing is a computing hypothesis, where a huge group of systems is linked together in private, public, or hybrid network, to offer dynamically amendable infrastructure for data storage, file storage, and application. With this emerging technology, application hosting, delivery, content storage, and reduced computation cost are achieved, and it acts as an essential module for the backbone of the Internet of Things (IoT). The efficiency of cloud service providers (CSP) could be improved by considering significant factors such as availability, reliability, usability, security, responsiveness, and elasticity. Assessment of these factors leads to efficiency in designing a scheduler for CSP. These metrics also improved the quality of service (QoS) in the cloud. Many existing models and approaches evaluate these metrics. But these existing approaches do not offer efficient outcome. In this paper, a prominent performance model named the “spectral expansion method (SPM)” evaluates cloud reliability. The spectral expansion method (SPM) is a huge technique useful in reliability and performance modelling of the computing system. This approach solves the Markov model of cloud service providers (CSP) to predict the reliability. The SPM is better compared to matrix-geometric methods.

1. Introduction

Cloud computing is a computing model for enabling expedient, on-demand network service to a pool of computing resources such as servers, storage, networks, services, and applications that can be promptly planned and released with minimum executive effort or service provider interface. Cloud computing has novel characteristics such as service-oriented computing, huge scale resource sharing, and on-demand services [1, 2]. Many cloud applications are developed using multitier architecture [3]. In the traditional model, one-tier, two-tier, 3-tier, to N -tier models are available in the cloud to provide n no. of services based on the tier type. Among these models, Multiple-tier or N -tier cloud is a significant model because it can provide N no. of service together in a single-tier model. As it provides multiple services, the maintenance of performance is more required to this model. To maintain the capacity of the model, the reliability evaluation is necessary for this multitier cloud system. For example, Microsoft Azure Cloud service in [4] and Amazon cloud service (EC2) in [5]

uses the multitier cloud environment for their web application. Compared to traditional architecture, the multitier cloud model guarantees the service level agreement (SLA) with combined service in a single physical machine. To manage the SLA, the reliability of the system must be predicted. A multitier model can simultaneously manage multiple users with a high workload. To reduce the usage of power and storage, the main concept called virtualization is evolved in each tier for better performance of the system. In [6], the author comes out with the outcome that multitier architecture is the best suitable with virtualization technology. This combination can save power, storage, and reliability.

Although various studies have proposed approaches recently on maintaining availability and balancing power performance [7–10], no one considered reliability, another major factor of cloud service. Reliability is the assertion that cloud provisions are free from software faults, hardware faults, and other faults that break down the system’s efficiency. Reliability modelling is necessary for the cloud environment. In [11], the author evaluates the reliability model

for maintaining the performance and power of the cloud virtualization environment. The reliability maintenance of the system indirectly affects performance and energy. In [12], a quality model called CLOUDQUAL is proposed to manage QoS among cloud service. The quality factors considered in this model are availability, reliability, usability, security, and elasticity. From this, reliability is a significant factor to achieve QoS in a cloud environment. In all existing studies, the reliability is calculated in a different way and considered as an important parameter for QoS of the cloud.

In this work, the high-level performance modelling called the spectral expansion method is proposed to estimate the reliability of the cloud computing environment. The spectral expansion method is a mathematical performance technique for analysis of two-dimensional Markov process of finite state space. This Markov model occurs commonly in reliability and performance problem of computing systems. The Markov model of the computing system is first represented in matrix form to compute the eigenvalues λ and eigenvectors Ψ . The eigenvalues and vectors form a linear equation v_i and solve it. This solution of the linear equation v_i determines the reliability of the system in graphical form. The remaining section of the paper is described as follows: Section 2 discusses existing studies; Section 3 explains the background of multitier cloud environment; Section 4 describes the spectral expansion model and algorithm; Section 5 illustrates numerical analysis and comparison between other models; and, finally, Section 6 concludes with a conclusion and future work.

2. Related Work

Reliability and availability are a critical requirement for cloud services and must be represented with appropriate planning and modelling [13, 14]. In [15], the reliability model is focused on a virtual machine (VM) of the cloud environment to accomplish high performance. Vishwanath and Nagappan [16] considered the reliability parameter to maintain the consistency of the cloud-based hardware system. The quality of service (QoS) is improved using the reliability metric in performance analysis. In [17], the author states that reliable service and availability services are strongly associated with the performance analysis of cloud-based service.

A reliability optimization algorithm [18] is introduced to maintain the performance of the top banking sector developed using a cloud system. Xia et al. [19] determined the reliability analysis of the computing system using a non-Markov stochastic Petri net (NMSPN). This method takes service invocation and message as model input. Reliability modelling is done for predicting the fault tolerance of computers. A small change in design or insight may cause high variation in performance modelling. To prevent the variation, reliability modelling is estimated in this work [20]. Peng and Huang [21] propose the reliability framework that collaborates failure dependencies as well as considers individual services and failure source. Reliability of the computing system [22] prevents failure and dependency

from occurring. Heimann et al. [23] compute the dependability analysis of the computer system which combines concepts like maintainability, availability, and reliability. System reliability measures the instance of offensive events in the computing system. From this instance, the prevention measure can be applied. In [24], a survey has been presented on reliability and energy efficiency in the cloud computing system. They predict that key challenge in the cloud is to manage the reliability of the system. Reliability is defined in the context of resource failure, in the context of VM failure, in the context of service failure, or in the context of security. Nachiappan et al. [25] have used reliability for cloud storage scheduling in big data. From the existing studies, very few reliability modelling is proposed for a cloud-based system. These studies consider only a few parameters and do not produce efficient outcomes.

In this paper, the spectral expansion method (SPM) is evolved for estimation of reliability modelling. This method has been used for performance modelling of the network-based system. The main task of SPM is to solve any class of the Markov model. This method provides a better solution compared to the matrix-geometric method [26]. Chakka [27] used the spectral expansion technique for finding the solution for the Markov model in some finite queue. This model exacts solution for analysis of the Markov process. The results include the performance analysis of the computing system. The spectral expansion method is applied for the performance and dependability analysis of the computing system. In this work, the SPM is applied to predict the reliability of cloud-based service.

3. Background

In this section, virtualization technology and multitier architecture designed are described.

3.1. Virtualization Technology. In recent years, the virtualization technology has been well developed in business and academics. Virtualization technology desires to provide background where VMs can be operated effectively and share the resource to another host efficiently. This leads to improvement in QoS with minimized cost [28]. Each VM handles whole software development including operating system, middleware, and application. Virtualization also appends the abstraction layer to computer architecture for multiplexing resources among other VMs. Several VMs can be virtualized into a physical system based on its capacity. The VMs can be arranged in series or parallel or distributed according to user's needs. The VMs provide many services such as application service, database service, security service, and software service. Based on the architectural design, each VM acts as a particular cloud service provider. By virtualization of VM into a physical system [29], we can provide a lot of service within a system. VM behaviour [30] is monitored using a virtual monitor machine. The VM can also facilitate physical resources to the operation and isolate the environment for execution of virtualization. VM can also reduce cost consumption, power consumption [31], and

workloads [32, 33]. Therefore, VM hosts as a logical machine that is similar to the real computer system.

3.2. Multitier Architecture. Partitioning the system architecture into layers is a fundamental approach to addressing the complexity, and the order of magnitude increases the rate of change. The system is constructed as a continuous sequence of layers, each of which is independent of adjacent ones and communicates with the previous tier via interfaces. The multitier application is a scalable architecture [6]. The architecture design of the multitier application cloud is good compared [34] to 3-tier application although the implementation complexity is more compared to 3-tier. In 3-tier architecture, it is limited and only three fundamental components or tiers such as presentation layer, business layer, and data layer are used to provide services, but in multitier, it is not limited. It includes these three fundamental tiers and can add many components for providing service. In 3-tier, security component is part of the presentation layer, but in multitier, security tier is created as an independent layer to provide secure applications. Like this, N -component and their services can be integrated into multitier application to provide more application and services to the client with less cost and high performance [35]. In 3-tier application, the user interfaces are less interactive compared to the multitier application. It provides more scalability and availability of services compared to the 3-tier application. Due to more interaction, it provides more on-demand services. This on-demand service is well suited in cloud computing design. Thus, the multitier application is well suited for scheduling and provisioning cloud application in the cloud.

The layers deployed in multitier architecture are as follows: The presentation layer provides an interface between user and cloud server and maintains the user interface between the client and the application server. In the multitier pattern, the presentation layer provides more interaction to the end-user. The security layer provides secure processing of the application to the user. It is a framework that provides authentication and data security. It offers data integrity and data confidentiality services. The applications are validated and authorized to the requested client. The business or domain layer implements client application logic. It applies the application logic. This layer provides all services related to application-specific functionality. The domain layer is executed in one or more application servers which communicate with the security, presentation, and data layers. In each application, a separate virtual machine is deployed. The data layer maintains consistency storage and simplifies access to data stored in persistence storage. The details of each application are stored in the data layer. The applications are developed with the help of information required. The data layer is implemented in one or several database servers. The data layer uses many techniques to store and retrieve data efficiently. Several approaches like

caching, master-slave replication, SQL, and NoSQL database are used.

4. Spectral Expansion Method (SPM)

The initial process of evaluating the reliability of the system is to design the Markov model for multitier cloud environment. Consider the multitier cloud architecture as shown in Figure 1, with M identical virtual machines (VMs) and host machines (HMs) processing different and unbounded queue of job. The VM and HM can fail from time to time. The failure may be single and independent or multiple and dependent. In this system design, both single, independent repair and multiple, dependent repair are allowed. In execution of tasks in queue, the VM and HM process a single job at a time and each job requires only one VM processor at a time. The policy applied for the failure of the processor is to resume, repair, and re-sample. The service rate, failure rate, and repair rate follow an exponential distribution.

The system is modelled by SPM. $X(t)$ is the functioning state of the system, denoting the number of VM processors running at time t which varies from 0 to M . $Y(t)$ is the number of processes waiting in the queue system and served processes at time t . The irreducible Markov model representing the multitier cloud system is defined as $I = \{[X(t), Y(t)]; t > 0\}$, with state space $\{0, 1, \dots, M\} \times \{0, 1, \dots\}$. Tasks are assumed to arrive based on the Poisson process with rate λ when the operative state $X(t) = i$. Two kinds of failures are possible. One is an individual processor breakdown independently at rate δ and is repaired independently at rate θ . The second is an "Overall" breakdown of all current VM and HM processors, at rate δ_0 , and the "Overall" repair rate of all current non-processing processors is at θ_M . The processing state of the system never changes during the arrival and departure of the tasks unless if there is an independent attack towards the changes. Hence, the alteration in VM and HM state of the system is done only in matrices P and P_y .

The single-move upward transition is generated by a single task arrival. Hence, Q and Q_y , the single-move upward transition matrices rate, are defined as

$$Q = Q_y = \text{diag} [\lambda_0, \lambda_1, \dots, \lambda_M]. \quad (1)$$

This is applicable for all values of y ($y=0, 1, \dots$).

The single-move downward transition occurs by decamping of the single task that is arrived. R and R_y are the single-move downward matrices rate. Let μ be the service rate of VM and HM processor. The decamping rate of tasks at time t depends on $X(t) = x$ and $Y(t) = y$, and that is defined as $R_y(x, x)$. If $x > y$, then every task is allocated to the processor for service provision and all VM and HM processors are not occupied; thus, the decamping rate $R_y(x, x) = y\mu$. Otherwise if $x \leq y$, then all the VM and HM processors are occupied with tasks; therefore, the decamping rate $R_y(x, x) = x\mu$. Thus, we conclude that $R_y(x, x)$ does not rely on y if $y \geq x$ and R_y does not depend on y if $y \geq M$. Therefore, we have

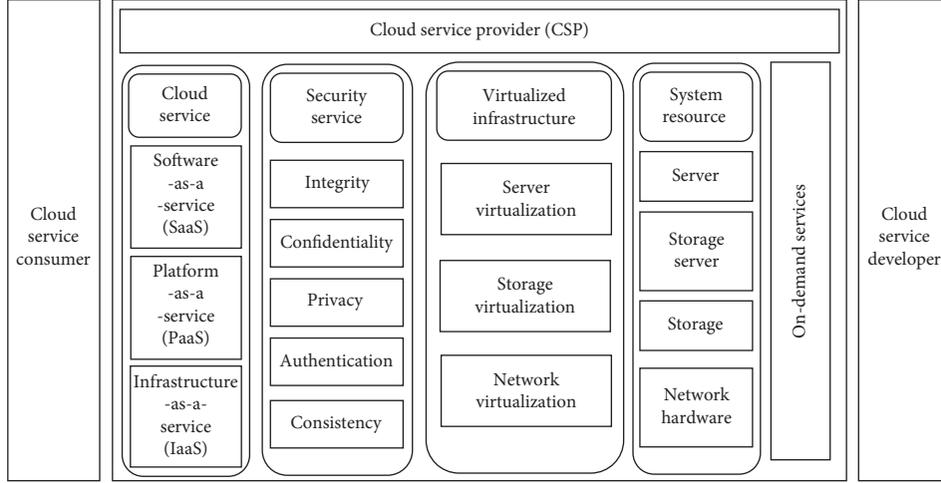


FIGURE 1: Architecture of the multitier cloud system.

$$R_y = \text{diag}[0, \min(y, 1)\mu, \min(y, 2)\mu, \dots, \min(y, M)\mu],$$

$$0 < y < M,$$

$$R = \text{diag}[0, \mu, 2\mu, \dots, M\mu], \quad y \geq M.$$
(2)

R_0 is equal to zero.

To evaluate the reliability of the cloud-based system, three different states are considered based on breakdown and repair rate in the form of matrices P and P_y . The three states are as follows.

Case 1. In this state, VM and HM processors lead to failure independently at a rate of δ per processor. Each non-functional processor has a repair rate of θ . There is no synchronization between failure or repair of multiple VM and HM processors.

The matrix P_y and P are defined as

$$P = P_y (y = 0, 1, \dots) = \begin{bmatrix} 0 & M\theta & & & \\ \delta & 0 & (M-1)\theta & & \\ & 2\delta & 0 & \ddots & \\ & & \ddots & \ddots & \theta \\ & & & & M\delta & 0 \end{bmatrix}.$$
(3)

Case 2. (independent failure and repairs as in state 1, and “overall” failure occurring at rate δ_0). All currently functioning processors fail at this rate except the independent failure:

$$P = P_y (y = 0, 1, \dots) = \begin{bmatrix} 0 & M\theta & & & \\ \delta_0 + \delta & 0 & (M-1)\theta & & \\ \delta_0 & 2\delta & 0 & \ddots & \\ \vdots & & \ddots & \ddots & \theta \\ \delta_0 & & & & M\delta & 0 \end{bmatrix}.$$
(4)

Case 3. It is similar to state 2, but in addition to this, the currently nonfunctioning processor gets repairs simultaneously. This “Overall” repair occurs at θ_M :

$$P = P_y (y = 0, 1, \dots) = \begin{bmatrix} 0 & M\theta & & & \theta_M \\ \delta_0 + \delta & 0 & (M-1)\theta & & \theta_M \\ \delta_0 & 2\delta & 0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \theta \\ \delta_0 & & & & M\delta & 0 \end{bmatrix}.$$
(5)

The threshold limit is given as H , and it should be $H = M$. The spectral expansion model works until $y = H - 1 = M - 1$. The changes occurred in a functioning state is observed in matrix P ; it is likely to find a simple form for the total average service rate of multiprocessor and the steady-state distribution of the number of functioning processors. Let u be the marginal distribution of the number of functioning processors. Then,

$$u = (q_0, q_1, \dots, q_M) = \sum_{y=0}^{\infty} u_y. \quad (6)$$

This is the probability vector of matrix $P - D^P$ and can be acquired by solving the following equation:

$$u(P - D^P) = 0, \tag{7}$$

$$u\varepsilon = 1.$$

Then, the total average service rate, named as the capacity of multiprocessor service, is $uR\varepsilon$, and average task arrived is $uQ\varepsilon$. It can be emphasized that the cloud-based system is stable when the average task arrival should be less than the capacity of service δ :

$$uQ\varepsilon < uR\varepsilon. \tag{8}$$

5. Numerical Analysis

The reliability of the cloud-based system is implemented using SHARPE tool. The SHARPE tool is an analysis tool in which user-designed architecture is developed and analysis based on spectral Expansion mechanism is done. In this work, the multitier cloud as shown in Figure 2 is developed for the three types of conditions. Using the above mathematical mechanism, the reliability of the multitier cloud for three types of conditions is evaluated. To compare and evaluate the performance of the state, the following parameters are considered such as the capacity of service and the processors with the same service rate. The overall failure rate, δ_0 , of state 2 is reimbursed by increasing the value of θ . In type 3, δ_0 is reimbursed by θ_M . The task arrival rate, the capacity of the processing system, and ergodicity condition are the same for all the three states. The three states are compared to show the performance evaluation among them, and the arrival rate is independent for VM and HM processors. Here, the system with 10 processors is considered for the implementation process.

In Figure 3, the effect of the number of tasks executed $E(Y)$ and the arrival rate of the task are described. In Cases 2 and 3, the repair rate and failure rate of VM and HM processors are more compared to Case 1. This fact causes the differentiation of processing time between them. The processing time required for Case 1 is more than Cases 2 and 3. The failure rate of Cases 2 and 3 and the repair rate of Case 3 tend to increase the variance of processing time compared to Case 1. As a result, the reliability of the cloud system in Case 3 is better compared to Case 2 and Case 1. Due to the simultaneous failure and repair rate maintained in Case 3, the average number of jobs execution is increasing compared to other cases.

In Figure 4, the number of processors $M(Y)$ and their service rate are illustrated. For a small range of $N(Y)$, Cases 2 and 3 show improving result better than Case 1, whereas for large $M(Y)$, Case 1 shows better results. The possible explanation for this is that (i) the processor availability reduces when both failure and repair rates increase, (ii) optimal value of M is larger for Case 1 than for Cases 2 and 3. The proposed reliability modelling using SPM is compared with other standard reliability models such as reliability graph model [36] and hierarchical correlation model (HCM) [11]. These two models evaluate reliability in the cloud-based system. In

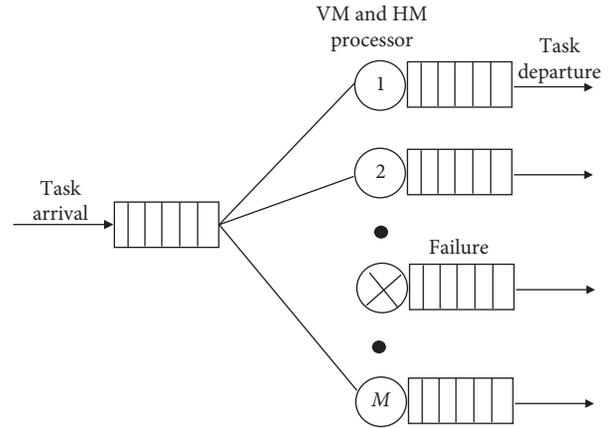


FIGURE 2: Multiprocessor working mechanism.

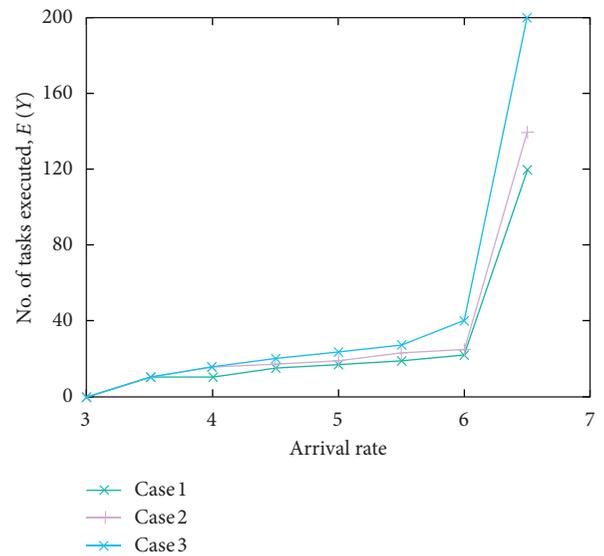


FIGURE 3: Reliability analysis.

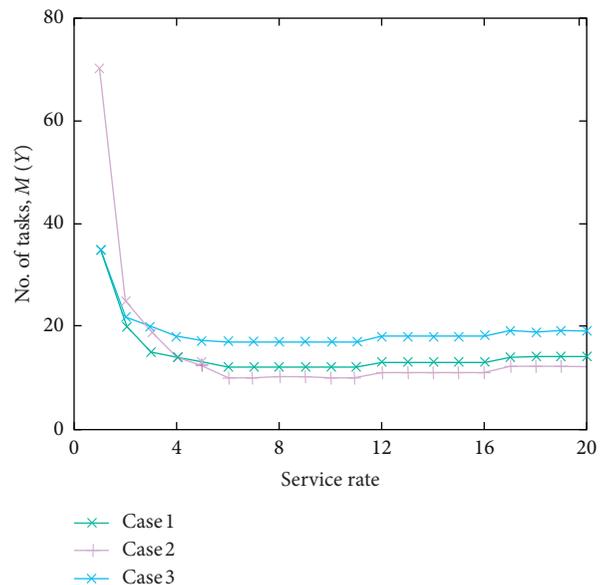


FIGURE 4: Service rate vs $M(Y)$.

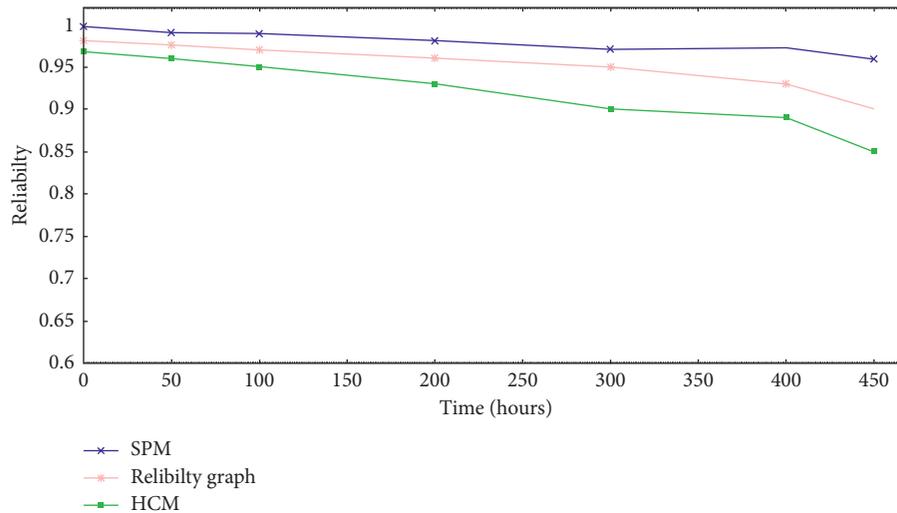


FIGURE 5: Comparison with other reliability models.

this comparison, the three-reliability model is applied to multitier cloud system and measurements are taken individually. These measurements are drawn as graph comparison in Figure 5. The figure illustrates that the proposed reliability model is efficient than the other two models.

The SPM reliability model maintains accuracy in evaluating reliability for a long duration compared with other models. It gives 0.999997% accurate result compared to other models. Analysing the reliability of multitier cloud is a complex task. The SPM reliability model can analyse this complex task with best accuracy. But other models show less accuracy compared with the proposed model.

6. Conclusion

In this, the reliability of the multitier cloud system is evaluated using the spectral expansion method (SPM). The SPM is developed for three cases of the multitier cloud system. Their reliability modelling is done using SHARPE tool. The comparison of the performance concludes that the system with simultaneous repair and failure rate works with better reliability. Thus, using the SPM, the multitier cloud-based system can be evaluated and analysed. The SPM provides accurate reliability evaluation compared to other standard methods because it considers all parameters for calculating the model. Thus, the graph results show the reliability can be maintained with failure and service rate in processing tasks and also the SPM model can produce accurate result compared to other models. The future work is to implement the SPM reliability model and test on real cloud-based system.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009.
- [2] P. Barham, B. Dragovic, K. Fraser et al., "Xen and the art of virtualization," in *ACM SIGOPS Operating Systems Review*, vol. 37, pp. 164–177, no. 5, ACM, Newyork, NY, USA, 2003.
- [3] W. Lloyd, S. Pallickara, O. David, J. Lyon, M. Arabi, and K. Rojas, "Performance implications of multi-tier application deployments on infrastructure-as-a-service clouds: towards performance modeling," *Future Generation Computer Systems*, vol. 29, no. 5, pp. 1254–1264, 2013.
- [4] S. Turkarlan, *Technical Article for SQL Server and Azure, Application Patterns and Development Strategies for SQL Server in Azure Virtual Machines*, 2014, <http://msdn.microsoft.com/en-us/library/azure/dn574746.aspx#comparison>.
- [5] M. Tavis and P. Fitzsimons, *Web Application Hosting in the AWS Cloud*, Site Point, Melbourne, Australia, 2012.
- [6] Y. Diao, J. L. Hellerstein, S. Parekh, H. Shaikh, and M. Surendra, "Controlling quality of service in multi-tier web applications," in *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06)*, p. 25, IEEE, Lisboa, Portugal, July 2006.
- [7] E. Gelenbe and R. Lent, "Energy-QoS trade-offs in mobile service selection," *Future Internet*, vol. 5, no. 2, pp. 128–139, 2013.
- [8] D. Kliazovich, P. Bouvry, and S. U. Khan, "DENS: data center energy-efficient network-aware scheduling," *Cluster Computing*, vol. 16, no. 1, pp. 65–75, 2013.
- [9] Y. Xia, M. Zhou, X. Luo, S. Pang, and 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, pp. 73–83, 2015.
- [10] G. Nalinipriya, K. G. Maheswari, B. Balusamy, K. Kotteswari, and A. Kumar Sangaiah, "Availability modeling for multi-tier cloud environment," *Intelligent Automation & Soft Computing*, vol. 23, no. 3, pp. 485–492, 2017.
- [11] X. Qiu, Y. Dai, Y. Xiang, and L. Xing, "A hierarchical correlation model for evaluating reliability, performance, and

- power consumption of a cloud service,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 3, pp. 401–412, 2016.
- [12] X. Zheng, P. Martin, K. Brohman, and L. D. Xu, “CLOUDQUAL: a quality model for cloud services,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1527–1536, 2014.
- [13] N. Kryvinska, C. Strauss, and Z. Peter, “A model to provide higher services availability for the mission critical applications,” in *Proceedings of the International Conference on Ambient Systems, Networks and Technologies (ANT-2010), in Conjunction with iiWAS2010*, vol. 8–10, pp. 221–229, Paris, France, November 2010.
- [14] N. Kryvinska and C. Strauss, “Conceptual model of business services availability vs. interoperability on collaborative IoT-enabled eBusiness platforms,” in *Internet of Things and Inter-cooperative Computational Technologies for Collective Intelligence*, pp. 167–187, Springer, Berlin, Germany, 2013.
- [15] Y. S. Dai, B. Yang, J. Dongarra, and G. Zhang, “Cloud service reliability: modeling and analysis,” in *Proceedings of the 15th IEEE Pacific Rim International Symposium on Dependable Computing*, pp. 1–17, Shanghai, China, November 2009.
- [16] K. V. Vishwanath and N. Nagappan, “Characterizing cloud computing hardware reliability,” in *Proceedings of the 1st ACM Symposium on Cloud Computing*, pp. 193–204, ACM, Indianapolis, IN, USA, June 2010.
- [17] E. Bauer and R. Adams, *Reliability and Availability of Cloud Computing*, John Wiley & Sons, Hoboken, NJ, USA, 2012.
- [18] Y. Liu, W. Liu, J. Song, and H. He, “An empirical study on implementing highly reliable stream computing systems with private cloud,” *Ad Hoc Networks*, vol. 35, pp. 37–50, 2015.
- [19] Y. Xia, X. Luo, L. Jia, and Q. Zhu, “A petri-net-based approach to reliability determination of ontology-based service compositions,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 5, pp. 1240–1247, 2013.
- [20] W. G. Bouricius, W. C. Carter, D. C. Jessep, P. R. Schneider, and A. B. Wadia, “Reliability modeling for fault-tolerant computers,” *IEEE Transactions on Computers*, vol. C-20, no. 11, pp. 1306–1311, 1971.
- [21] K.-L. Peng and C.-Y. Huang, “Reliability analysis of on-demand service-based software systems considering failure dependencies,” *IEEE Transactions on Services Computing*, vol. 10, no. 3, pp. 423–435, 2017.
- [22] A. Zhou, S. Wang, B. Cheng et al., “Cloud service reliability enhancement via virtual machine placement optimization,” *IEEE Transactions on Services Computing*, vol. 10, no. 6, pp. 902–913, 2017.
- [23] D. I. Heimann, N. Mittal, and K. S. Trivedi, “Availability and reliability modeling for computer systems,” in *Advances in Computers*, vol. 31, pp. 175–233, Elsevier, Amsterdam, Netherlands, 1990.
- [24] Y. Sharma, B. Javadi, W. Si, and D. Sun, “Reliability and energy efficiency in cloud computing systems: survey and taxonomy,” *Journal of Network and Computer Applications*, vol. 74, pp. 66–85, 2016.
- [25] R. Nachiappan, B. Javadi, R. N. Calheiros, and K. M. Matawie, “Cloud storage reliability for Big Data applications: a state of the art survey,” *Journal of Network and Computer Applications*, vol. 97, pp. 35–47, 2017.
- [26] I. Mitrani and R. Chakka, “Spectral expansion solution for a class of Markov models: application and comparison with the matrix-geometric method,” *Performance Evaluation*, vol. 23, no. 3, pp. 241–260, 1995.
- [27] R. Chakka, “Spectral expansion solution for some finite capacity queues,” *Annals of Operations Research*, vol. 79, pp. 27–44, 1998.
- [28] M. Armbrust, A. Fox, R. Griffith et al., “Above the clouds: a berkeley view of cloud computing,” Technical Report No. UCB/EECS-2009-28, Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA, 2009.
- [29] J. Schad, J. Dittrich, and J.-A. Quiané-Ruiz, “Runtime measurements in the cloud: observing, analyzing, and reducing variance,” *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 460–471, 2010.
- [30] S. Fu, “Failure-aware resource management for high-availability computing clusters with distributed virtual machines,” *Journal of Parallel and Distributed Computing*, vol. 70, no. 4, pp. 384–393, 2010.
- [31] Y. Ding, X. Qin, L. Liu, and T. Wang, “Energy efficient scheduling of virtual machines in cloud with deadline constraint,” *Future Generation Computer Systems*, vol. 50, pp. 62–74, 2015.
- [32] W. Zhang, J. Liu, C. Liu, Q. Zheng, and W. Zhang, “Workload modeling for virtual machine-hosted application,” *Expert Systems with Applications*, vol. 42, no. 4, pp. 1835–1844, 2015.
- [33] X. Xu, H. Jin, S. Wu, and Y. Wang, “Rethink the storage of virtual machine images in clouds,” *Future Generation Computer Systems*, vol. 50, pp. 75–86, 2015.
- [34] M. S. Rehman and M. F. Sakr, “Initial findings for provisioning variation in cloud computing,” in *Proceedings of the 2010 IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 473–479, IEEE, Indianapolis, IN, USA, December 2010.
- [35] X. Bu, R. Jia, and C.-Z. Xu, “A reinforcement learning approach to online web systems auto-configuration,” in *Proceedings of the 2009 29th IEEE International Conference on Distributed Computing Systems (ICDCS’09)*, pp. 2–11, IEEE, Montreal, Canada, June 2009.
- [36] T. A. Nguyen, D. Min, E. Choi, and T. D. Tran, “Reliability and availability evaluation for cloud data center networks using hierarchical models,” *IEEE Access*, vol. 7, pp. 9273–9313, 2019.

