Harvest: An International Multidisciplinary and Multilingual Research Journal

E-ISSN: 2582-9866 Volume III Issue I January-March 2023

A Review of Dynamic Pricing Methods for Service Providers Under the Multi-Server Cloud Environment

[1*]**A. Rajesh,** *Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Thuduppathi, Perundurai, Erode-638057.*

[2]**Dr. R. Kalaivani,** *Professor, Department of Electronics and Communication Engineering, Erode Sengunthar Engineering College, Thuduppathi, Perundurai, Erode-638057.*

[3]**C. Senthilkumar,** *Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Thuduppathi, Perundurai, Erode-638057.*

Abstract

In the corporate sphere, which is highly well-known in the modern world, the word used is "cloud computing." Users and other service providers are linked to cloud services without considering their hardware requirements. The users may select multiple servers in this method to obtain the most advantageous plans for the resources they effectively store in the cloud. A multi-server setup can provide a better level of availability and dependability than a single-server environment, which is its principal benefit. In cloud computing, a service provider's pricing strategy is based on two factors: revenue and expense. A service provider's revenue comes from the service it charges customers, while its expenses are made up of the rent plus any utilities it must pay infrastructure vendors. To maximize income for cloud services and improve customer satisfaction, SLAs are necessary to guarantee that the service price and the quality of the services offered (QoS) are continuously consistent and interrelated. Even if one server dies in a multi-server system, the remaining servers can still give users access to the applications and services they need. Furthermore, because the load may be spread across numerous servers, a multi-server architecture can provide a greater performance level than a single-server setup.

Keywords: Multi-server environment, Cloud computing, SLA, Multi-cloud, Service provider, dynamic pricing scheme.

Introduction

A cloud is a type of concurrent and distributed system comprised of several linked and virtualized computers that will continually be offered and shown as one or more pooled computing resources as part of service-level agreements agreed upon between the service provider and the consumer. Figure 1 below shows the fundamental types of cloud computing.

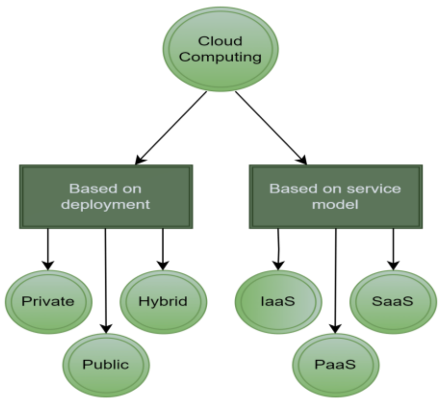

Figure. 1. Basic forms of cloud computing

Three services are primarily offered by the cloud such as infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). Fundamental cloud computing consists of four kinds, namely: Public Cloud, Private Cloud, Hybrid Cloud, and Multi-Cloud. Every server in a multi-server system is in charge of its own duties and responsibilities. Employing these servers would lessen how dependent the resources are on one another. It would not affect the database servers, for instance, if you have a separate database server and your main server does. An offer between a service provider and clients known as SLA assures that a certain level is ensured at all times. A cloud SLA guarantees that cloud providers adhere to a set of enterprise-level standards and offer customers a precisely specified set of deliverables. The failure of the supplier to uphold the guaranteed terms is also described, along with the financial consequences, such as service time credits. Pricing strategies fall into one of two categories: fixed pricing model or dynamic pricing model. In fixed or static pricing models long-term price charging changes are impossible. The cloud provider has previously decided on the costs of various resources. And in a dynamic pricing model, the prices are constantly altered based on the state or condition of the market. Because of a change in price and a shift in both demand and supply [1]. The advantages, benefits, and drawbacks of the dynamic pricing models are shown in Table 1 below.

Table 1. Benefits and drawbacks of the dynamic pricing method

Pricing model	Characteristics	Benefits	Drawbacks
Dynamic pricing Strategy	It also goes by the name of Real-time pricing.	It enables boosting earnings or income.	It makes users feel alienated.
	Due to pricing variations throughout time, there is a lot of flexibility in determining the cost of the services.	It modifies the pricing in consideration of cost and time.	More sophisticated technology is required for altering pricing and profit calculations.

Various digital research article sources were explored to get relevant publications. More than 20 journal articles were assessed using various inclusion and removal criteria in order to determine the items that were significant to the literature study. We commonly utilized phrases such as "multi-server environment," "dynamic pricing schemes," "cloud environment pricing schemes," and so on. This research examines 27 peer-reviewed papers. This study will be a helpful review for future researchers.

Literature Review

As defined by California State University, Berkeley, "cloud computing pertains to both the hardware and system software in the data centers that serve such services as well as the apps supplied as cloud-based services." [2]. A three-tier organization with users, service providers, and infrastructure vendors is one appealing cloud services environment [3]. A vendor for infrastructure keeps up the fundamental hardware and software infrastructure [4]. In order to provide clients with a variety of services, cloud services rent services from infrastructure providers. According to specific SLAs, a client sends a service call to a service provider, obtains the required result from the provider, and settles for the service depending on its cost and quality.

Wu et al. [5]: In cloud computing, shared resources, software, and data are made available for other devices and computers as a metered service across a network rather than being sold as a product (Generally the Internet). A dynamic pricing strategy combining two data mining techniques is suggested to increase the income of cloud service providers. The k-means algorithm divides chronological data into groups and is the primary data mining technique. The second is a Bayes decision, which has the ability to anticipate the trend of user-favored cloud service packages. The BPneutral network is used in the suggested pricing model to estimate the price that would optimize revenue. The suggested approach can better match consumer demand and beat models without the K-means algorithm and static pricing models in terms of maximizing revenue.

Al-Roomi et al. [6] explain the procedure of figuring out what a service provider will get from an end user in compensation for its service is known as pricing. According to Weinhardt et al. [7], the only way to make cloud computing profitable in the IT industry is to establish effective pricing strategies. The pricing procedure may be either fixed, in which case the client is always billed the same amount, dynamic, where in case the price required to pay fluctuates, or market-dependent, so in that case, the consumer is invoiced in accordance with the current market circumstances [8]. Pay-per-use methods are one type of fixed pricing scheme where clients are charged based on how much of a product or how long they utilize a particular service. Another form of fixed pricing is a subscription, that the consumer makes a one-time payment in order to utilize the service for longer lengths of time at any suitable time or sum. Another type of set pricing is list pricing, which refers to a fixed price that may be found in a compendium or list. Differential or dynamic pricing, then suggests that the price varies constantly in response to the service features, client attributes, number of purchases, or consumer preferences.

An SLA fundamentally denotes a contract between a consumer and a provider to obtain a specific service [9]. The QoS (Quality of service) characteristics, like reaction time, bandwidth, storage, dependability, throughput, delay, deadline, and cost, are included in SLAs and must be upheld by a provider [10]. For various application domains, a service provider can construct several multiserver systems, allowing service requests of various types to be routed to various types of multi-server systems [11]. Each of its systems has numerous servers that can be allocated to handle a certain kind of service request or application. Two fundamental characteristics, namely the count of servers in a multi-server setup (its selection of cloud services) and the performance of a multiserver system, were used to define its setup (the execution speed of the servers).

Studies that have already been done on cloud computing prices make the assumption that cloud servers are uniform [11], [12], and [13]. Servers for homogeneous cloud computing have the same amount of storage, processing power, energy supply, and bandwidth. But, homogenous servers did not accurately represent the real-world use case for cloud server systems. Systems for cloud servers are made up of a variety of servers that offer varying levels of service [14]. The requirement for extremely responsive, high-performance systems which also connect using external sources was obtained from the requirement for growing heterogeneity in computer systems [15]. As a result, the need to create pricing models for multiserver structures that are mixed arose results that pricing models based on homogeneous multiserver systems cannot adequately reflect the current deployment scenarios of cloud servers, which deploy servers with various service rates and capabilities.

Feng et al. [13] adaptively distributed the cloud resources across various service instances relying on the frequently gathered data in an effort to optimize revenue. Each service circumstance, a virtual computer linked to a user, is characterized in their study as an M/M/1/FIFO queue system that is FIFO (first in first out). Another statistic in cloud computing that might be utilized to calculate prices was Mean response time (MRT) and immediate reaction time (IRT), two customer-oriented pricing mechanisms, which were presented by the authors. Under these mechanisms, consumers are paid in accordance with the realized quality of service measured in terms of mean response time. It was possible to determine the ideal number of servers needed to improve profit.

A dynamic pricing strategy for federated clouds, wherein the services were shared among several cloud service providers, was presented by Mihailescu and Teo [16]. The effectiveness of this strategy was assessed through simulations by contrasting it with a fixed pricing mechanism. When buyer welfare and the amount of fulfilled requests increased, they discovered that dynamic pricing produced superior overall results. But, the fixed price was more scalable when there was a significant market requirement.

Two distributed net profit optimization algorithms—net profit optimization for indivisible tasks (NPOI) and net profit optimization for divisible jobs (NPOD)—were proposed by Wang et al. [17] (NPOI). The difference between a divisible and an indivisible work is that the latter can be interrupted. By contrasting the two algorithms to the largest job first (LJF) method, the authors were able to demonstrate through simulations that they can boost profits and lower power costs. The authors, therefore, solely took into account static job arrivals and exits. Additionally, they used the assumption that all data centers' servers were identical, which does not accurately reflect the actual deployment situations for cloud servers. A homogeneous multiserver system's drawback was that it had worse performance overall and required longer execution times for a multitude of tasks.

The author's employed mean response time as the performance indicator and considered that the multi-server system was homogenous, i.e., that each server was identical. The heterogeneity displayed by cloud servers, nonetheless, cannot be captured by a homogenous multi-server system [18]. Additionally, profit maximization depends on mean response only indicating the benefit from larger-sized requests, not from all requests. This study developed a revenue model that takes the heterogeneity of the servers into account and bases prices on the obtained service quality in terms of mean response time or means slow down.

According to mean response time and mean slowness, Nansamba et al. [19] study generated models for profit for the heterogeneous multiserver system. When contrasted with the homogenous revenue model, the functioning of the offered pricing models is evaluated. A heterogeneous multiserver system processes more profitability than a homogeneous multiserver system, according to the numerical findings from the derived models. Additionally, it has been shown that mean slowness generates more money when employed as a pricing statistic than mean response time. One class of consumers in a heterogeneous multiserver system was the sole system examined in the study. It could be insightful to expand the paradigm in the coming years to include different kinds of customers, some of whom are given priority. It uncovers price strategy depending on the mean slowdown as well as potential methods of estimating profit for heterogeneous multiserver cloud computing systems. This study is anticipated to aid academics in identifying novel methods of cloud computing user billing as well as potential approaches to modeling revenues for heterogeneous multiserver cloud computing environments.

A pricing method for cloud computing services was presented by Li et al., [20]. Because the cloud bank agent approach would offer accurate evaluation and support for all members, this model is utilized as a resource firm. To establish the price, the authors employed a continuous price modification technique. It examined the historical ratio of resource use, current pricing has continuously repeated many times, evaluated the resource availability for the following round, and calculated the ultimate price. A service request dealer, cloud banking, a cloud service agent (CSA), and a cloud resource agent were all included in the system. Due to its inability to adjust to the quick changes that the market frequently experiences, the suggested pricing scheme was comparably rigid. Nonetheless, it may lower provider expenses and increase their profits, freeing up resources for more efficient usage.

A genetic method for pricing in the cloud computing industry was established by Macias et al., [21]. There are three basic processes in choosing a suitable pricing model using genetic algorithms: specify the chromosome, assess it, and, the best chromosomal pairs for reproduction are chosen, then the poorest chromosome pairs are discarded. The simulation's results showed that, in most circumstances, genetic pricing generates higher income. Utilizing genetic pricing, service companies may increase revenues by up to 100% compared to other dynamic pricing techniques and by up to 1000% compared to fixed pricing strategies. The genetic model that was provided has a flexible genome and has been shown to be more stable over noise and profitable than the one with a rigid genome. The suggested genetic model is simple to use, adaptable, and versatile enough to take into account a range of different factors that affect the price scenario. By establishing relationships between the factors that affect pricing, the genetic pricing method might be further investigated.

Various prospective pricing techniques for services have been discussed in numerous studies, including pay-as-you-go, subscription, pay-for-resources, value-based, cost-based, customer-based, competition-based, and dynamic pricing schemes [6, 22]. The pay-as-you-go, two-part tariff, and subscription pricing schemes are really the most common ones. Numerous academics have investigated a few dynamic pricing schemes [6, 7]. Some of the pricing schemes are shown below in Table 2.

Table 2. Strategies for pricing

Pricing model	**Characteristics & Pricing schemes**	**Fairness and Execution**	**Benefits**	**Drawbacks**
Pay-as-you-go Model [6]	The pricing of services is established by the service provider and remains constant.	Users are being treated unfairly.	Users are informed of the exact pricing of the services.	Prices are not changing in response to demand.
	This model employs a static approach.	Users may be asked to pay more for the same services.	Users get access to resources for the duration of the paid fixed time period.	The resource has been reserved for users who have been underused or over-utilized over an extended period of time.
Value-based-pricing [16]	The price is determined by how the customer perceives it.	It is fair to suppliers because the pricing is determined by how customers perceive the service.	Maximum revenue increase.	Difficulties arise during execution.
	The dynamic technique was employed in this model.	It's a model that's been implemented.		
Algorithm for pricing cloud computing resources [20]	Its foundation is real-time pricing.	For service suppliers, this paradigm is equitable.	With this strategy, profits rise while expenses fall.	Demand and supply fluctuations are too rapid to cause a price shift.
	A dynamic model was utilized.	This method of simulation-based approach is theoretical.		
A genetic pricing model for the cloud computing	Real-time pricing is used to set prices.	Treat service suppliers fairly.	Execution is simple.	Poorly perform under both high and low-demand circumstances.

industry [21]	This employed a dynamic strategy.	Because it maximizes revenue.	Maximum financial benefit	

Conclusion

This review article presented several dynamic pricing mechanisms that ensure service provider fairness in a multi-server environment. Due to high levels of efficiency and dependability, many firms are moving to multi-server environments. In general, the usage of several servers in a firm helps ensure greater performance and uptime, maintain security, and provide more effective resource allocation. The distribution of resources among several servers, each running a different operating system, has numerous additional advantages. Utilizing separate operating systems and separating the operations on many servers might result in a variety of benefits. Effective resource management is crucial for companies that are expanding quickly, including startups with expanding business models. The service pricing method has been enhanced to handle circumstances including demand, reservations, peak demand, and peak supply. The service charge functions now provide indicators for data consumption cost and communication range. The system has a dynamic service function selection model. Analysis of service requests and authorizations is done to determine the service provider's profit margin.

Reference

[1] Samimi P, Patel A. Review of pricing models for grid & cloud computing. In IEEE symposium on computers & informatics 2011 (pp. 634-9). IEEE.

[2] Etro, F., 2009. The economic impact of cloud computing on business creation, employment and output in Europe. An application of the endogenous market structures approach to a GPT innovation. Rev. Bus. Econ. Lit., 54: 179-208. 9.

[3] Lee, Y.C., C. Wang, A.Y. Zomaya and B.B. Zhou, 2010. Profitdriven service request scheduling in clouds. Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, May 17-20, 2010, Melbourne, Australia, pp: 15-24. 12.

[4] Popovici, F.I. and J. Wilkes, 2005. Profitable services in an uncertain world. Proceedings of the ACM/IEEE Conference on Supercomputing, November 12-18, 2005, Seattle, WA., USA., pp: 36.

[5] X. Wu, J. Hou, S. Zhuo, and W. Zhang, “Dynamic pricing strategy for cloud computing with data mining method,” in *Communications in Computer and Information Science*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 40–54.

[6] M. Al-Roomi, S. Al-Ebrahim, S. Buqrais, and I. Ahmad, “Cloud computing pricing models: A survey,” *Int. j. grid distrib. comput.*, vol. 6, no. 5, pp. 93–106, 2013.

[7] Weinhardt, C., A. Anandasivam, B. Blau, N. Borissov, T. Meinl, W. Michalk and J. Sto$er, 2009. Cloud computing-a classification, business models and research directions. Bus. Inform. Syst. Eng., 1: 391-399.

[8] A. Osterwalder, “The Business Model Ontology – A Proposition in a Design Science Approach”, Doctoral thesis, University of Lausanne, (2004)

[9] Sahal, R., M.H. Khafagy and F.A. Omara, 2016. A survey on SLA management for cloud computing and cloud-hosted big data analytic applications. Int. J. Database Theory Applic., 9: 107-118. 14.

[10] Firdhous, M., S. Hassan and O. Ghazali, 2013. Monitoring, tracking and quantification of quality of service in cloud computing. Int. J. Scient. Eng. Res., 4: 112-117. 15.

[11] Cao, J., K. Hwang, K. Li and A.Y. Zomaya, 2013. Optimal multiserver configuration for profit maximization in cloud computing. IEEE Trans. Parallel Distrib. Syst., 24: 1087-1096. 16.

[12] Zhang, L. and D. Ardagna, 2004. SLA based profit optimization in autonomic computing systems. Proceedings of the 2nd International Conference on Service Oriented Computing, November 15-19, 2004, New York, NY, USA., pp: 173-182. 19.

[13] Feng, G., S. Garg, R. Buyya and W. Li, 2012. Revenue maximization using adaptive resource provisioning in cloud computing environments. Proceedings of the 13th ACM/IEEE International Conference on Grid Computing, Volume 13, September 20-23, 2012, IEEE Computer Society Washington, DC, USA., pp: 192-200. 20.

[14] Narman, H.S., M.S. Hossain and M. Atiquzzaman, 2014. h-DDSS: Heterogeneous dynamic dedicated servers scheduling in cloud computing. Proceedings of the IEEE International Conference on Communications, June 10-14, 2014, Sydney, NSW., Australia, pp: 3475-3480.

[15] Suri, P.K. and M. Sumit, 2012. A comparative study of various computing processing environments: A review. Int. J. Comput. Sci. Inform. Technol., 3: 5215-5218

Harvest: An International Multidisciplinary and Multilingual Research Journal

E-ISSN: 2582-9866 Volume III Issue I January-March 2023

[16] Mihailescu, M. and Y.M. Teo, 2010. Dynamic resource pricing on federated clouds. Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, May 17-20, 2010, IEEE Computer Society, pp: 513-517. 28.

[17] Wang, W., P. Zhang, T. Lan and V. Aggarwal, 2012. Datacenter net profit optimization with individual job deadlines. Proceedings of the 46th Annual Conference on Information Sciences and Systems, March 21-23, 2012, Princeton, NJ., USA. 27.

[18] Crago, S.P. and J.P. Walters, 2015. Heterogeneous cloud computing: The way forward. Computer, 48: 59-61.

[19] B. Nansamba, K. S. Kaawaase, M. Okopa, and B. K. Asingwire, “Pricing scheme for heterogeneous multiserver cloud computing system,” *Australas. J. Comput. Sci.*, vol. 4, no. 1, pp. 32–43, 2016.

[20] Li H, Liu J, Tang G. A pricing algorithm for cloud computing resources. In international conference on network computing and information security 2011 (pp. 69-73). IEEE.

[21] Macías M, Guitart J. A genetic model for pricing in cloud computing markets. In proceedings of the ACM symposium on applied computing 2011 (pp. 113-8). ACM.

[22] Mazrekaj, A.; Shabani, I.; Sejdiu, B. Pricing schemes in cloud computing: An overview. Int. J. Adv. Comput. Sci. Appl. 2016, 7, 80–86. [CrossRef]