



Private Cloud in 6G Networks: A Study from the Total Cost of Ownership Perspective

Yuanfang Chi¹, Wei Dai², Yuan Fan¹, Jun Ruan¹, Kai Hwang^{2,3},
and Wei Cai^{2,3}(✉)

¹ Alibaba Group, Beijing, China

{cyf170506, fanxi.fy, jun.ruanj}@alibaba-inc.com

² The Chinese University of Hong Kong, Shenzhen, Shenzhen, China

³ Shenzhen Institute of Artificial Intelligence and Robotics for Society,
Shenzhen, China

weidai@link.cuhk.edu.cn, {hwangkai, caiwei}@cuhk.edu.cn

Abstract. Security and privacy concerns are increasingly important when massive data processing and transferring becomes a reality in the era of the Sixth Generation (6G) Networks. Under this circumstance, it becomes a trend that the enterprises tend to host their data and services on private clouds dedicated to their own use, rather than the public cloud services. However, in contrary to the well-investigated total cost of ownership (TCO) for public clouds, the analytic research on the cost of purchase and operation for private clouds is still a blank. In this work, we first review the state-of-the-art TCO literature to summarize the models, tools, and cost optimization techniques for public clouds. Based on our survey, we envision the TCO modeling and optimization for private clouds by comparing the differences of features between public and private clouds.

Keywords: Cloud computing · Total cost of ownership · Case study

1 Introduction

It took ten years for the network infrastructure to evolve from 4th generation (4G) to 5th generation (5G). According to Moore's Law, the expected waiting time for the 6th generation (6G) network will be even shorter. According to predictions, the bandwidth will be increased to terabytes, with which the users are able to exchange a large volume of data in a short time. Under this circumstance, security and privacy are increasingly important, since the risks of data leakage and the potential damages to service hijack are dramatically grown at the same time.

This work was supported by Project 61902333 supported by National Natural Science Foundation of China, by the Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS).

Meanwhile, cloud computing has been widely adopted by startups and well-established enterprises, thanks to its prominent elastic and on-demand features. The flexibility of on-demand usage reduces development, service deployment, and maintenance costs. Huge data computation, backup, and recovery tasks have become easier with the cloud. The cloud computing market is expanding. Many large, medium, and small enterprises will choose to invest in cloud services [24] to enjoy the benefits of cloud computing, but risks often accompany the investment. For a company that provides cloud services or uses cloud services, how to stabilize development and revenue is also a key success factor. On the one hand, operating an enterprise with cloud-based IT infrastructure can improve the stability of the enterprise and bring intuitive benefits. A number of articles in the financial and business domain have reported the cloud computing business models, such as the VE model [46] and Cloud Business Model Framework (CBMF) [69]. More discussions and studies about business models can be found at [5, 7, 13, 16, 38, 40, 48]. On the other hand, analyzing Return On Investment (ROI) can also increase the value of investment. ROI analysis and modeling methods in cloud computing have been discussed in [4, 37, 64]. Cost analysis and optimization modeling can effectively help decision-making and increase profit under the premise of guaranteeing service quality [39].

Nevertheless, the privacy and security problems of cloud services are never fully solved. The tenants have no choice but surrender their data and programs to the public cloud providers, who may not be trustworthy if the data is extremely sensitive. To address this concern, cloud providers propose private cloud service, which is typically deployed inside the organization and is typically behind the firewall [74]. Private cloud enhances the enterprise customers' confidence to embrace the cloud since it is an isolated solution that can prevent data from being transmitted to the public network. It maximizes the customers' control over their data, provides improved security and service quality [52]. In addition, the local network within a private cloud also reduces network latency and improves transmission quality. The isolation, low latency, and security of private cloud make many applications possible to deploy on the cloud, such as the Industrial Internet of Things (IIoT) [9], Federal Learning [29], and Cluster Learning [35].

From the users' perspective, the cost efficiency is always a key metric when considering cloudization. On the other hand, cloud providers also need to evaluate their profits in the cloud business. Hence, the total cost of ownership (TCO), a standard approach in analyzing the cost of purchase and operation of an asset, is a critical topic for both parties. By definition, the TCO in cloud computing covers the capital and operational costs of building, using, and maintaining a cloud data center. It involves all direct and indirect costs, such as the cost of cloud service providers purchasing, deploying, operating, and maintaining assets, and the cost of cloud users renting cloud resources. Due to the importance of TCO, the analysis and optimization methods for public cloud and hybrid cloud have been well investigated. However, few work has been devoted to the cost analysis associated with private clouds. In this work, we survey the literature on

public cloud TCO analysis methods and the optimization solutions that minimize the TCO of a cloud service. We compare and contrast the similarities and differences between public and private clouds from a TCO perspective and provide an in-depth discussion on what else can be further optimized to accomplish a lower TCO in a private cloud.

The remainder of the paper is organized as follows. We present an overall review of existing works on TCO in Sect. 2. Afterwards, we compare the public and private cloud from TCO optimization perspective in Sect. 3. Section 4 concludes the work and envision the TCO research for private clouds.

2 Literature Review

TCO was first being studied in the business domain by Ellram since 1993 [15]. She claimed that TCO modeling is constructive for business parchment decision making, ongoing supplier management, and understanding of indirect costs. However, she also pointed out that the quantification and measurement of TCO is complexity [14]. In 1999, Milligan suggested that accurate total cost measurement is elusive due to the lack of analysis methods [43]. Thus, one of the leading research directions in TCO is to find an accurate modeling and measurement method. Later, Degraeve and Roodhooft illustrated the application of TCO in the supply chain domain. They split purchasing activities into three levels corresponding to three different expenses. Then, they built a mathematical model to optimally select suppliers such that the total cost of ownership is minimized [72]. As the IT industry rising in 2004, few scholars began to study TCO in the information technology domain. They have worked on topics such as revenue analysis in IT by taking advantage of TCO [4], how to utilize TCO to decide whether to adopt open-source software in IT companies [31,45], and optimization models to decrease the cost of IT infrastructure [1]. After cloud computing entered the market, some researchers analyzed the business value of cloud computing and mentioned that cloud providers should pay attention to the cost [27], while few studies took the modeling of TCO in cloud computing in to consideration. Although Patel and Shah presented a cost model to calculate the cost of building a data center, they did not take indirect costs such as maintenance cost, operation cost, and labor cost into consideration [50]. TCO in cloud computing was finally studied by Li et al. in 2009. A cloud cost amortization model is built to make it possible to calculate the direct and indirect cost of cloud computing infrastructures. Their model can be divided into eight parts: server cost, software cost, network cost, support and maintenance cost, power cost, cooling cost, facilities cost, and real-estate cost. They also implemented an interactive tool for cloud providers to calculate cloud TCO [33]. Scholars begin to pay much more attention to models and tools to analyze TCO in cloud computing and the optimization to reduce TCO.

In order to give the readers better comprehension on the landscape of the cloud computing TCO research, we classified the surveyed papers into groups as shown in Fig. 1. The selected papers are mainly divided into two groups: models

and tools for cloud TCO, and cost optimization methods. For the papers related to models and tools, we will further divide into two subclasses from the perspective of cloud computing providers and cloud computing users respectively. For the papers related to cost optimization methods, we will further classify them into three topics: task scheduling, resource scheduling, and heterogeneous resource optimization. More details on the classification system are in the following.

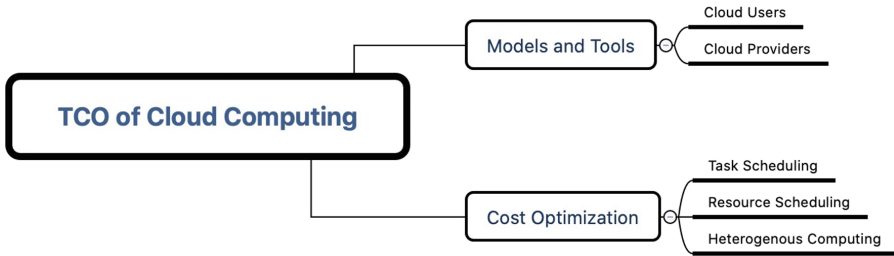


Fig. 1. Our proposed classification system of cloud computing TCO papers.

- Models and Tools for Cloud Computing TCO (Sect. 2.1): We consider papers that provide modeling methods or tools to measure or calculate TCO for cloud computing providers or users. These studies can be further categorized into two classes based on their viewpoints: models and tools for cloud providers and models and tools for cloud users.
- Cost optimization (Sect. 2.2): We consider papers that optimize cloud computing TCO by optimizing one aspect of resources in cloud computing. Each work focus on different optimization directions. Such studies can be further categorized into three groups: task scheduling, resource scheduling, and heterogeneous resource optimization.

2.1 Models and Tools

In cloud computing, a model and tool that measures the total cost of ownership can help analyze investment risk, manage cloud resources, and improve the effectiveness of decision making. The modeling and tools related papers are roughly categorized into two classes: (i) Modeling and tools for Cloud Providers and (ii) Modeling and tools for Cloud Users. They are surveyed in the following.

Models and Tools for Cloud Providers. The cost has been known as one of the most important factors for cloud providers. Moreover, there is significant capital consumption, such as cloud data center construction, service software development, maintenance. So, cost evaluation will be helpful for revenue estimation, investment risk analysis, and decision making. From the perspective of

cloud providers, the research of cloud TCO modeling and tools can be further classified into two subclasses, (i) TCO breakdown and (ii) energy cost models and tools.

TCO Break Down. The primary purpose of cloud total cost of ownership breakdown research is to provide cloud providers with a measurement of cost. It focuses on the modeling approaches of server cost, software cost, network cost, support and maintenance cost, power cost, cooling cost, facilities cost, and real-estate cost.

In 2009, Li et al., as the pioneers, first solved the problem that the lack of a method to measure TCO. They proposed a cloud cost amortization model to calculate the total cost, and further use the model to find VM utilization cost given the number of running VMs [33]. Later, researchers studied cloud TCO breakdown under different business models. In 2015, Filiopoulou et al. applied the system of system (SoS) method to reorganize each part of the total cost. They regraded each type of cost as a subsystem of cloud computing and illustrated a cost modeling framework for each subsystem [17]. In 2016, Simonet et al. demonstrated a TCO breakdown modeling method for distributed cloud computing (DCC). Besides, they categorized DCC actors into five levels and presented related cost models [56]. It provides a detailed reference for cloud providers who intend to invest distributed cloud computing.

Through the above studies, researchers showed an overview of cloud TCO breakdown. However, they either overlooked some detailed cost measurements for some terms in the formulation or overlooked the cost calculation under some specific business scenarios. In real business, the cost can be significantly altered by different business scenarios and other dynamic factors such as application migration, timely changed resource utilization rate, and satisfactory network dependability and availability. In the rest of the subsection, we will review researches that focus on cost modeling for one aspect of TCO breakdown.

In 2011, Mencer et al. demonstrated a cost model for software development in cloud computing. They believed that the computational efficiency and development cost of the software would be affected by different abstraction level programming languages. By analyzing software development cost and TCO, they concluded that TCO is not necessarily minimized by minimizing programming effort because programming efforts might be related to conditions of servers [41]. In 2013, Sun and Li illustrated a labor cost model to calculate labor efforts in a service migration scenario. They quantified the skill level of employees and built a probabilistic model to estimate the number of persons needed per day to finish migration [61]. Omana et al. proposed an analyzing method for cloud providers to determine whether to replace aged assets. By measuring the cost of power, cooling, physical space, asset attachments, and IT support, they utilized statistical methods to examine the relationship between asset cost and resource capacities, such as the number of CPU cores and memory. They finally mined some useful results to help decision making and device management [24].

Thanakornworakij et al. did a cloud TCO research on the premise of ensuring system availability and satisfying service-level agreement (SLA), and quality

of service (QoS). They modeled the cost of servers, network, software, power, cooling, facilities, maintenance, and availability. Then, they obtained the relationship between the number of devices needed and revenue under 99% system availability, which is beneficial for cloud providers to define the right size of the data center [62]. Sousa's research team presented a modeling strategy under a similar scenario. They introduced Mean Time to Failure (MTTF) and Mean Time to Repair (MTTR) to build a detailed maintenance cost model. They further derived a Stochastic Model Generator for Cloud Infrastructure Planning (SMG4CIP) which concerned both dependability and cost requirements [58, 59].

Furthermore, some studies presented modeling methods for dynamic factors in cloud TCO such as resource utilization cost. Although [33] tried to use the ratio of running VMs to model utilization cost at a given time point, they can only get a rough result because the cost of running VMs depends on other frequently changed variables like CPU rate. As a follow-up study of [33], in 2013, Vrček and Brumec pointed out that many cloud TCO related studies ignored some hidden variables to build cost models, such as CPU utilization rate, data transmission rate, system load. They utilized CPU rate per hour to build a cost model and analyzed how the CPU rate affects the total cost in cloud computing [66]. Their study offered some vital insight into cloud computing TCO. Since time dimension can be added into the cost model, cloud providers can generate flexible strategy by analyzing per hour cost. Molka and Byrne had a similar idea that an accurate prediction model of resource utilization rate can derive an accurate utilization cost calculation. They demonstrated a prediction model using an online nonlinear autoregressive method and then formulated a model to calculate the real-time cost using CPU utilization data [44]. If there are some more accurate prediction model that can be applied and plugged in [44], cloud providers can obtain more accurate information to make decisions. Some research about resource prediction can be found in [54]. In 2017, Singh et al. also took account of the resource utilization rate to calculate utilization cost. They captured the effect of the relinquishment of cost and revenue [57].

Energy Cost Modeling and Tools. With the development of cloud computing and its convenience, the energy and electricity power consumed by cloud data center per year increase dramatically [51], which leads to a higher expense to operate a data center. An investigation indicated that energy cost account for 42% of all expenditure per month in data center until 2011 [22]. Thus, energy cost, as part of the total cost of ownership, gradually becomes a focus of cloud providers' concern. We survey some representative energy cost modeling methods and tools as follows.

In 2010, Yu and Bhatti thought [33] ignored approaches to gain specific values in the energy cost formulas. They pointed out there is a lack of an information model that collects energy usage and resource usage data for devices for cloud providers to manage assets. They proposed a Scalable Energy Monitor (SEM) architecture to measure energy consumption and further to calculate energy cost [71]. Later in 2012, Uchechukwu et al. studied a detailed energy cost modeling method. They divided energy into static and non-static parts where energy con-

sumed by storage, computation, and communication is modeled. By analyzing cost using their measurement, they found it is possible to save energy costs [65]. In the same year, Chen et al. had a similar idea. They built a model with a static part and a dynamic part, and further developed an analytic tool to measure and summarize energy consumed by different tasks [8]. Recently, some researchers focus on establishing more advanced models. In 2018, Jawad et al. illustrated a smart Power Management Model (PMM) to schedule power by adapting a Nonlinear Autoregressive Network with Exogenous Inputs and Neural Network based forecasting algorithm (NARX-NN) [26].

Models and Tools for Cloud Users. In general, cloud users are small companies or startups who rent cloud services from cloud providers to operate their businesses. There are three types of service in cloud computing: Infrastructure as a Service (IaaS) that provides computation power and data storage, Platform as a Service (PaaS) that provides developers platforms and Software as a Service (SaaS) that allows users access software service by light-weight clients [40]. Besides, there are different cloud providers in the market with different pricing models [55], such as Amazon EC2, Microsoft Azure, Google Compute Platform. It is not trivial for cloud users to choose the optimal leasing solution from the intricate market. Several representative papers are selected to illustrate methods and tools to analyze TCO from the perspective of cloud users.

In 2009, Kondo et al. first found monthly cost can be decreased drastically by deploying tasks on cloud servers instead of on Volunteer Computing (VC) platform [28]. Later in 2011, Han introduced the concept of cloud TCO from the perspective of cloud users. He presented a detailed analysis of TCO by comparing cloud service with local storage and servers [23]. Inspired by [23, 28], in 2012, Martens et al. first contributed the TCO measurement for cloud users. They summarized cost variables to be considered by cloud users when renting cloud services, and proposed cost models for IaaS, PaaS and SaaS [39]. In 2013, Martens and his team continued their research on TCO models. They demonstrated a deployment planning strategy for cloud users and presented more detailed cost models for IaaS, PaaS and SaaS [67].

Additionally, some researchers focused on developing tools and methods that compare different cloud providers and estimate resources acquired by users' services or applications to further provide an optimal leasing plan. Liew and Su proposed a tool called CloudGuide which can predict the cloud computing resources required, and helps users select the most suitable leasing scheme from multiple cloud service products with different pricing modes according to user policy, such as maximizing performance or minimizing cost [34]. Aniceto et al. gave a more detailed deployment strategy for small IT companies as cloud users. They abstracted the resource into the number of instances, and built a statistical model using historical data to predict the number of instances demanded at the moment. The authors further categorized cloud products into reserved instances and on-demand instances according to the pricing model. They finally reduced 32% cost compared with only adapting on-demand deployment by formulating

a mixed deployment strategy [47]. In 2018, Ghule and Gopal further proposed a comparison framework for cloud users to analyze the difference between IaaS cloud providers by defining suggestive comparison parameters, such as reliability, performance, serviceability [18].

2.2 Cost Optimization

In above sections, we reviewed modeling methods and tools to measure TCO. It may be not enough for providers or users to make an optimal decision. In order to sustain a thriving business in enterprise, it is an effective way to maximize profit by optimizing and reducing the total cost of ownership. In cloud computing, many papers studied cost optimization from different aspect. We survey and summarize several representative papers from the aspect of task scheduling, resource scheduling and heterogeneous computing.

Task Scheduling. As cloud providers, all users' operation requests will be executed in provider back end. Effectively scheduling tasks requested by users can improve service quality, reduce system latency, and reduce costs. We select several typical studies. They adopted different optimization methods in different application scenarios and finally achieved certain results.

Pandey et al. optimized the general task scheduling problem in cloud computing, and solved the Task-Resource Scheduling Problem by using the Particle Swarm Optimization (PSO)-based Heuristic optimization method. They minimized the cost of task execution to reduce the operating costs of the TCO. The proposed PSO method can achieve lower cost than BRS (best resource selection) algorithm, and the PSO can converge faster than the GA (Genetic Algorithm) [49]. Zhang et al. further considered the scheduling strategy when tasks can be executed parallelly. They proposed the DeCloud architecture which is responsible for scheduling users' data requests from a centralized data storage center. They presented Generic Searching Algorithm (GSA) and Heuris Searching Algorithm (HSA) to schedule parallel and non-parallel tasks respectively. Finally, their method is better than greedy search and random search [73].

Resource Scheduling. As a typical resource scheduling problem in cloud computing, elastic computing has always been a concern of cloud providers. Over-provisioning will lead to lower resource utilization, thus increasing cloud computing operating costs. Under-provisioning will result in unmet user requirements and reduction of service quality. We survey two cost-related resource scheduling studies as below.

Each research team has different focuses and directions on this issue. Wu et al. demonstrated an optimization model to schedule resource in SaaS. They minimized the cost by minimizing resources allocated to VMs while meeting SLA and QoS requirements. Their proposed algorithm can reduce the cost by 50% compared to the base algorithm, ProfminVio [70]. Mao et al. viewed VM (Virtual Machine) as the unit of recourse, and considered several factors including the

type of VM, the startup latency of the VM, and the deadline of the task to construct an auto scaling schedule strategy. They performed different types of tasks by scheduling different kinds of VMs, and finally got a lower cost compared with using fixed VM type under the condition of satisfactory performance and deadline [36].

Heterogeneous Computing. Companies may encounter business situations that require a mix of local computing resources and exogenous public cloud computing resources, which is common in hybrid clouds. Some studies have found that it is possible to optimize the cost of reducing the amount of investment required by the enterprise by optimally allocating the workload or task quota for local and heterogeneous resources.

In 2010, Trummer et al. utilized the COP (Constraint optimization problem) method to minimize the rental cost of external cloud services such that the cost of the enterprise is minimized [63]. However, their research did not analyze the cost of local resources. Thus, it cannot give a global optimal cost solution. Bittencourt and Madeira proposed HCOC (Hybrid Cloud Optimized Cost) scheduling algorithm. It enables computing tasks in a hybrid cloud to be dynamically assigned to local resources (private clouds) or external sources (public clouds). Their method can effectively improve the efficiency of task completion, and reduce the cost compared to greedy schedule algorithm [3]. In 2015, Laatikainen et al. focused on the optimization of the storage costs of hybrid clouds. They first formulated the measurement of hybrid cloud storage costs and then analyzed the impact of refining the reassessment interval on the cost savings attainable by using hybrid cloud storage. Finally they obtained that shortening reassessment interval and the acquisition of public cloud storage capacity allows the volume variability to be reduced, yielding a reduction of the overall costs [30].

3 Overall Comparison Between Public and Private Cloud

Many cloud service providers start to deliver the entire data center as a whole cloud computing solution to customers, such as Alibaba Cloud Apsara Stack [21], Dell EMC [25], HPE Helion Open-Stack [11], Microsoft Azure Stack [42], which deploys public cloud software in a smaller private cloud, providing customers with customized, secure, low-latency private cloud services.

In the reviewed works above, most of them are related to cost modeling and optimization methods in public or hybrid clouds while private cloud is seldomly mentioned. In this chapter, we hope that by comparing and contrasting the public and private cloud environment, readers can draw a better understanding of the differences and similarities between public and private cloud, thus better understand the total cost of owning a private cloud and possible ways that can be attempted to optimize the cost. Table 1 gives a brief comparison of characteristic of public and private cloud.

As shown in the table above, both public and private clouds can provide IaaS, PaaS, and SaaS services, but they have significant differences in the other

Table 1. Comparison of characteristics between public cloud and private cloud.

Characteristics	Private cloud	Public cloud
Service Type	IaaS, PaaS, SaaS	IaaS, PaaS, SaaS
Service Users	Large Organizations	General Public or Small Startups
Device Ownership	Users	Cloud Service Providers
Deployment Location	User's Data Center	Cloud Provider's Data Center
Quality of Service	Stable	Unstable
Data Security	High	Low
Use Cost	High	Low
Scalability	Scale-out Only	Elastic Computation
Maintenance	Technician/Service Provider	Public Cloud Service
Procurement	High Customization	Low Customization

nine areas. The users of private cloud services are large organizations, communities and enterprises, such as government departments while public cloud orients to general public users, small companies and startups. In general, private cloud users need to pay more in the initial stage because they need to purchase a complete cloud computing infrastructure, while public cloud users only pay for the rental service in on demand price. For private cloud users, high costs usually are accompanied by stable cloud service quality and high data security. Since the private cloud is deployed in the user's data center, the user can have an isolated network built within the domain of an intranet to stably run the service. Meanwhile, data security is guaranteed because users can manage data autonomously. In contrast, since the public cloud is deployed in the data center of the service provider, the public cloud user needs to access the cloud service through the public network (Internet), and the quality of service often depends on the network quality between the user side and the public cloud data center. Since public cloud users don't know where their data is physically stored, there is a hidden risk of data leakage. In terms of procurement, since private cloud users may need to carry some personalized services, their infrastructure will also need to be customized. Servers in the private cloud can be highly customized where users can select the model specification according to the business requirement, while the public cloud server has a low degree of customization. Generally public providers purchase servers suitable for the multi-tenant technology.

3.1 The Difference from Cloud Provider's Perspective

In this section, we will analyze the similarities and differences between public cloud providers and private cloud providers in analyzing TCO from two aspects.

Procurement, Deployment and Maintenance. When creating a data center strategy, it is hard for public cloud providers to predict the exact user demands. So they need to adopt high-performance computing servers, implement scalable storage and networks to meet the increasing demand for

compute-intensive, storage-intensive, or IO bandwidth-intensive workloads. Private clouds, on the other hand, can get their customer requirements directly from their customers so the demands are more predictable than public clouds. However, to better meet each customer's business objects, each private cloud data center has to be customized in a certain level. This customization leads to huge personnel costs for private cloud providers. Furthermore, public cloud is deployed in carefully chosen sites that can be managed and continuously improved by providers, whereas the private cloud is deployed on customers' sites. In order to maintain the quality of service (QoS) and meet the Service Level Agreement (SLA), both public and private cloud service providers need to provide sufficient support on hardware maintenance, disaster recovery or the ability of service scale-in/scale-out, etc. On-site supports are offered by private cloud providers, which also increases the cost of owning a cloud data center for cloud provider's aspect. Therefore, a more flexible planning tools and automated deployment and maintenance tools are crucial for private cloud providers to lower their TCO, compared to public cloud providers.

Resource Scheduling. Public cloud service providers have realized resource pooling and on-demand cloud instances by leveraging virtualization, multi-tenant and elastic computing technology. By optimizing the resource scheduling algorithms, resource utilization is evidently increased. Equipment idle time and equipment operating costs can be reduced. Virtualization technology is also the fundamental in private clouds. Recently, many companies use light-weighted virtual machine, dockers, as the container to carry their applications [2]. Private cloud providers still need to develop resource scheduling strategies to optimally allocate resources to each docker carrying user business applications. The effectiveness of the resource scheduling strategy becomes even more crucial for private cloud since any waste of resource will reflect directly on the customer bill.

3.2 The Similarities and Differences from Cloud User's Perspective

As summarized in Table 1, there is a significant difference between public cloud users and private cloud users in nature. Generally, public cloud users are general public, small companies, and startups, and they usually cannot build a data center of its own due to the lack of funds. The public cloud's pay-as-you-go mode can meet the user's demand for resource on-demand, so that such users can quickly deploy their own business. In contrast, most of private cloud users are governments, large and medium-sized enterprises and communities. They have higher requirements for network service quality and data security, and initially have sufficient funds to purchase data centers. Compared with public cloud users, the cost spent by private cloud users in a short period of time would be much greater than that of public cloud users by nature. According to some studies, in the case of Amazon, if users need to operate their business for more than 18 months, the invested cost of public cloud users may be equal to or more than that of private cloud users [10].

Use and Maintenance. Both public cloud users and private cloud users need to develop their own commercial applications on the top of infrastructure using IaaS. The cost of application development depends on the technology stack and technology architecture used by the users. Some PaaS cloud services enable users to quickly complete application development or allows users to adapt the outsourcing API interface to complete product release. Thus, the development costs can be further reduced. For example, Google App Engine [20] can help users quickly establish a web application development; Cloud Healthcare API [19] can provide an intelligent healthcare interface that enables users to complete the development related to smart healthcare. On the SaaS side, public cloud users can use the software on the cloud to complete business tasks directly, such as Dropbox [12] to complete the task of cloud storage. PaaS and SaaS private cloud users may need to work harder if they choose to develop their own PaaS and SaaS applications for security reasons.

From the perspective of operation and maintenance, public cloud users need to have certain operation and maintenance knowledge of servers and networks. For private cloud users, there are usually professionals from private cloud providers side to sustain the basic operation and maintenance, which greatly reduces the cost of learning and training professionals.

Security and Network Reliability. Security and confidentiality are long-term issues in the public cloud [60]. Since public cloud users' data is stored in remote data centers, users may have concerns about confidentiality of data. If the data center crashes, users will also face data loss issues. Although public cloud multi-tenant technology increases the utilization of hardware resources through virtualization technology, multi-tenant environments may lead to potential data leakage caused by side-channel attacks [53]. The public cloud's latency may also affect the user's profit, if the user's business is latency sensitive [68]. Compared to the public cloud, private cloud users can manage and back up data autonomously because the infrastructure of the private cloud is completely owned by users. Since the network is relatively independent and the network traffic does not go through the public network, the delay will be much lower. Therefore, the risk of private cloud users in terms of security, network reliability and latency will be much lower than that of public cloud users.

Cloud Evaluation. For public cloud users and private cloud users, due to different charging standards of different providers, the choice of cloud service providers directly affects the cost of their investment. For public cloud users, there are some tools to help them choose the service provider and rental solution that best suits their requirements, such as [18, 32, 34]. However, there are no tools to provide users with a comparison of private cloud providers. This is not a trivial problem because private cloud users need to compare private cloud service offerings in terms of investment costs, business value benefits, security and stability, and IT cost savings. On the other hand, cloud users need to estimate the amount of computing and storage resources consumed by their own business

before adopting cloud computing solutions. Public cloud and private cloud users can control their costs by evaluating and predicting the resource requirements. Rodrigo N. Calheiros et al. proposed EMUSIM that can help users evaluate and predict the resources that will be consumed by their own business after migrate to cloud [6].

4 Conclusion

Nowadays, cloud computing is widely adopted for personal or organization use. How to help users and providers better understand their costs of owning a cloud service, so they can further optimize their usage or ways to build the service to lower the overall TCO has become the next question. Many have provided breakdowns of TCO calculation for public cloud, but few has discussed how TCO is different for the private cloud. In this work, by discussing the similarities and differences between public and private clouds in terms of server procurement and deployment, software, operations and maintenance and resource allocation algorithms, we hope to shade a light on the unique points that people can look at for TCO optimization of private clouds.

References

1. Ardagna, D., Francalanci, C., Trubian, M.: A cost-oriented approach for infrastructural design. In: SAC (2004)
2. Bhimani, J., Yang, Z., Leeser, M., Mi, N.: Accelerating big data applications using lightweight virtualization framework on enterprise cloud, 09 2017. <https://doi.org/10.1109/HPEC.2017.8091086>
3. Bittencourt, L.F., Madeira, E.R.M.: HCOC: a cost optimization algorithm for workflow scheduling in hybrid clouds. *J. Internet Serv. Appl.* **2**, 207–227 (2011)
4. Brocke, J.V., Lindner, M.: Service portfolio measurement - a framework for evaluating the financial consequences of out-tasking decisions, pp. 203–211, 01 2004. <https://doi.org/10.1145/1035167.1035197>
5. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging it platforms: vision, hype, and reality for delivering computing as the 5th utility. *Future Gener. Comput. Syst.* **25**, 599–616 (2009)
6. Calheiros, R., Netto, M., De Rose, C., Buyya, R.: EMUSIM: an integrated emulation and simulation environment for modeling, evaluation, and validation of performance of cloud computing applications. *Softw. Pract. Exp.* **43**, 595–612 (2013). <https://doi.org/10.1002/spe.2124>
7. Chang, V.I., Wills, G.B., Roure, D.D.: A review of cloud business models and sustainability. In: 2010 IEEE 3rd International Conference on Cloud Computing, pp. 43–50 (2010)
8. Chen, F., Schneider, J., Yang, Y., Grundy, J., He, Q.: An energy consumption model and analysis tool for cloud computing environments. In: 2012 First International Workshop on Green and Sustainable Software (GREENS), pp. 45–50, June 2012. <https://doi.org/10.1109/GREENS.2012.6224255>
9. Choo, K.R., Gritzalis, S., Park, J.H.: Cryptographic solutions for industrial Internet-of-Things: research challenges and opportunities. *IEEE Trans. Industr. Inf.* **14**(8), 3567–3569 (2018). <https://doi.org/10.1109/TII.2018.2841049>

10. Dantas, J., Matos, R.R.M., Araujo, J., Maciel, P.R.M.: Eucalyptus-based private clouds: availability modeling and comparison to the cost of a public cloud. *Computing* **97**, 1121–1140 (2015)
11. Development, H.P.E.: HPE Helion OpenStack (2019). <https://www.hpe.com/us/en/product-catalog/detail/pip.hpe-helion-openstack-cloud-software.1010838414.html>. Accessed 15 Oct 2019
12. Dropbox: Dropbox (2019). <https://www.dropbox.com/>. Accessed 15 Oct 2019
13. Centre for Economics and Business Research Ltd.: The cloud dividend: part one. The economic benefits of cloud computing to business and the wider EMEA economy-France, Germany, Italy, Spain and the UK. *Business & Information Systems Engineering* (2010)
14. Ellram, L.M.: Total cost of ownership: a key concept in strategic cost management decisions (1993)
15. Ellram, L.M.: Total cost of ownership: elements and implementation (1993)
16. Etro, F.: The economic consequences of the diffusion of cloud computing (2010)
17. Filiopoulou, E., Mitropoulou, P., Tsadimas, A., Michalakis, C., Nikolaidou, M., Anagnostopoulos, D.: Integrating cost analysis in the cloud: a SoS approach. 2015 11th International Conference on Innovations in Information Technology (IIT), pp. 278–283 (2015)
18. Ghule, D., Gopal, A.: Comparison parameters and evaluation technique to help selection of right IaaS cloud. In: 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), pp. 1–6, November 2018. <https://doi.org/10.1109/UPCON.2018.8597059>
19. Google: Cloud Healthcare API (2019). <https://cloud.google.com/healthcare/>. Accessed 15 Oct 2019
20. Google: Google App Engine (2019). <https://cloud.google.com/appengine/>. Accessed 15 Oct 2019
21. Group, A.: Apsara Stack (2009). <https://www.alibabacloud.com/product/apsara-stack>. Accessed 15 Oct 2019
22. Hamilton, J.: Cooperative expendable micro-slice servers (CEMS): low cost, low power servers for internet-scale services, 01 2009
23. Han, Y.: Cloud computing: case studies and total cost of ownership (2011)
24. Iglesias, J.O., Perry, P., Stokes, N., Thorburn, J., Murphy, L.: A cost-capacity analysis for assessing the efficiency of heterogeneous computing assets in an enterprise cloud. In: IEEE/ACM 6th International Conference on Utility and Cloud Computing, UCC 2013, Dresden, Germany, 9–12 December 2013, pp. 107–114. IEEE Computer Society (2013). <https://doi.org/10.1109/UCC.2013.32>
25. Inc., D.: Dell EMC (2019). <https://www.dellemc.com/en-us/solutions/cloud/vmware-cloud-on-dellemc.htm#scroll=off>. Accessed 15 Oct 2019
26. Jawad, M., et al.: A robust optimization technique for energy cost minimization of cloud data centers. *IEEE Trans. Cloud Comput.* **1** (2018). <https://doi.org/10.1109/TCC.2018.2879948>
27. Klems, M., Nimis, J., Tai, S.: Do clouds compute? a framework for estimating the value of cloud computing. In: WEB (2008)
28. Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D.P.: Cost-benefit analysis of cloud computing versus desktop grids. In: 2009 IEEE International Symposium on Parallel Distributed Processing, pp. 1–12, May 2009. <https://doi.org/10.1109/IPDPS.2009.5160911>
29. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: strategies for improving communication efficiency. *CoRR abs/1610.05492* (2016). <http://arxiv.org/abs/1610.05492>

30. Laatikainen, G., Tyrväinen, P.: Cost efficiency of hybrid cloud storage cost efficiency of hybrid cloud storage: shortening acquisition cycle to mitigate volume variation, 12 2015
31. Larsen, M.H., Holck, J., Pedersen, M.K.: The challenges of open source software in it adoption: enterprise architecture versus total cost of ownership (2004)
32. Li, A., Yang, X., Kandula, S., Zhang, M.: CloudCmp: comparing public cloud providers. In: Internet Measurement Conference (2010)
33. Li, X., Li, Y., Liu, T., Qiu, J., Wang, F.: The method and tool of cost analysis for cloud computing. In: 2009 IEEE International Conference on Cloud Computing, pp. 93–100, September 2009. <https://doi.org/10.1109/CLOUD.2009.84>
34. Liew, S., Su, Y.Y.: CloudGuide: helping users estimate cloud deployment cost and performance for legacy web applications, pp. 90–98, 12 2012. <https://doi.org/10.1109/CloudCom.2012.6427577>
35. Liu, Q., Cheng, L., Ozcebe, T., Murphy, J., Lukkien, J.: Deep reinforcement learning for IoT network dynamic clustering in edge computing. In: 2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), pp. 600–603, May 2019. <https://doi.org/10.1109/CCGRID.2019.00077>
36. Mao, M., Li, J., Humphrey, M.: Cloud auto-scaling with deadline and budget constraints. In: 2010 11th IEEE/ACM International Conference on Grid Computing, pp. 41–48, October 2010. <https://doi.org/10.1109/GRID.2010.5697966>
37. Skilton, M., Director, C., Cloud Business Artifacts Project, et al.: Building return on investment from cloud computing. White Paper (2009)
38. Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., Ghalsasi, A.: Cloudcomputing - the business perspective. *Decis. Support Syst.* **51**(1), 176–189 (2011). <https://doi.org/10.1016/j.dss.2010.12.006>
39. Martens, B., Walterbusch, M., Teuteberg, F.: Costing of cloud computing services: a total cost of ownership approach. In: 2012 45th Hawaii International Conference on System Sciences, pp. 1563–1572 (2012)
40. Mell, P., Grance, T.: The NIST definition of cloud computing. *Natl. Inst. Stand. Technol.* **53**(6), 50 (2009)
41. Mencer, O., Vynckier, E., Spooner, J., Girdlestone, S., Charlesworth, O.: Finding the right level of abstraction for minimizing operational expenditure. In: WHPCF@SC (2011)
42. Microsoft: Azure Stack (2009). <https://docs.microsoft.com/en-us/azure-stack/operator/azure-stack-overview?view=azs-1908>. Accessed 15 Oct 2019
43. Milligan, B.: Tracking total cost of ownership proves elusive. *Purchasing* **127**, 22–23 (1999)
44. Molka, K., Byrne, J.: Towards predictive cost models for cloud ecosystems: poster paper. In: IEEE 7th International Conference on Research Challenges in Information Science (RCIS), pp. 1–2, May 2013. <https://doi.org/10.1109/RCIS.2013.6577736>
45. Moyle, K.: Total cost of ownership and open source software (2004)
46. Mvelase, P., Sibiyi, G., Dlodlo, N., Oladosu, J., Adigun, M.: A comparative analysis of pricing models for enterprise cloud platforms. In: 2013 Africon, pp. 1–7, September 2013. <https://doi.org/10.1109/AFRCON.2013.6757870>
47. Orbegozo, I.S.A., Moreno-Vozmediano, R., Montero, R.S., Llorente, I.M.: Cloud capacity reservation for optimal service deployment. In: CLOUD 2011 (2011)
48. Pal, R., Hui, P.: Economic models for cloud service markets: pricing and capacity planning. *Theoret. Comput. Sci.* **496**, 113–124 (2013)

49. Pandey, S., Wu, L., Guru, S.M., Buyya, R.: A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In: 2010 24th IEEE International Conference on Advanced Information Networking and Applications, pp. 400–407, April 2010. <https://doi.org/10.1109/AINA.2010.31>
50. Patel, C.D., Shah, A.: Cost model for planning, development and operation of a data center (2005)
51. BONE Project: WP 21 tropical project green optical networks: report on year 1 and update plan for activities. No. FP7-ICT-2007- 1216863 BONE project (2009)
52. Qing, L., Boyu, Z., Jinhua, W., Qinqian, L.: Research on key technology of network security situation awareness of private cloud in enterprises. In: 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 462–466, April 2018. <https://doi.org/10.1109/ICCCBDA.2018.8386560>
53. Ren, K., Wang, C., Wang, Q.: Security challenges for the public cloud. IEEE Internet Comput. **16**(1), 69–73 (2012). <https://doi.org/10.1109/MIC.2012.14>
54. da Rosa Righi, R., et al.: A survey on global management view: toward combining system monitoring, resource management, and load prediction. J. Grid Comput. **17**(3), 473–502 (2019). <https://doi.org/10.1007/s10723-018-09471-x>
55. Sharma, U., Shenoy, P., Sahu, S., Shaikh, A.: A cost-aware elasticity provisioning system for the cloud. In: 2011 31st International Conference on Distributed Computing Systems, pp. 559–570, June 2011. <https://doi.org/10.1109/ICDCS.2011.59>
56. Simonet, A., Lebre, A., Orgerie, A.: Deploying distributed cloud infrastructures: who and at what cost? In: 2016 IEEE International Conference on Cloud Engineering Workshop (IC2EW), pp. 178–183, April 2016. <https://doi.org/10.1109/IC2EW.2016.48>
57. Singh, S., Aazam, M., St-Hilaire, M.: RACE: relinquishment-aware cloud economics model. In: 2017 24th International Conference on Telecommunications (ICT), pp. 1–6, May 2017. <https://doi.org/10.1109/ICT.2017.7998279>
58. Sousa, E., Lins, F., Tavares, E., Cunha, P., Maciel, P.: A modeling approach for cloud infrastructure planning considering dependability and cost requirements. IEEE Trans. Syst. Man Cybern. Syst. **45**(4), 549–558 (2015). <https://doi.org/10.1109/TSMC.2014.2358642>
59. Sousa, E., Maciel, P., Medeiros, L., Lins, F., Tavares, E., Medeiros, E.: Stochastic model generation for cloud infrastructure planning. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4098–4103, October 2013. <https://doi.org/10.1109/SMC.2013.699>
60. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. J. Netw. Comput. Appl. **34**, 1–11 (2011)
61. Sun, K., Li, Y.: Effort estimation in cloud migration process. In: 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering, pp. 84–91, March 2013. <https://doi.org/10.1109/SOSE.2013.29>
62. Thanakornworakij, T., Nassar, R., Leangsuksun, C., Paun, M.: An economic model for maximizing profit of a cloud service provider. In: 2012 Seventh International Conference on Availability, Reliability and Security, pp. 274–279 (2012)
63. Trummer, I., Leymann, F., Mietzner, R., Binder, W.: Cost-optimal outsourcing of applications into the clouds. In: 2010 IEEE Second International Conference on Cloud Computing Technology and Science, pp. 135–142, November 2010. <https://doi.org/10.1109/CloudCom.2010.64>
64. Tsalis, N., Theoharidou, M., Gritzalis, D.: Return on security investment for cloud platforms. In: 2013 IEEE 5th International Conference on Cloud Computing Technology and Science, vol. 2, pp. 132–137 (2013)

65. Uchchukwu, A., Li, K., Shen, Y.: Improving cloud computing energy efficiency. In: 2012 IEEE Asia Pacific Cloud Computing Congress (APCloudCC), pp. 53–58 (2012)
66. Vrek, N., Brumec, S.: Role of utilization rate on cloud computing cost effectiveness analysis. In: International Conference on Information Society (i-Society 2013), pp. 177–181, June 2013
67. Walterbusch, M., Martens, B., Teuteberg, F.: Evaluating cloud computing services from a total cost of ownership perspective. *Manage. Res. Rev.* **36**, 613–638 (2013)
68. Wan, Z.: Cloud computing infrastructure for latency sensitive applications. In: 2010 IEEE 12th International Conference on Communication Technology, pp. 1399–1402, November 2010. <https://doi.org/10.1109/ICCT.2010.5689022>
69. Weinhardt, C., et al.: Cloud computing - a classification, business models, and research directions. *Bus. Inf. Syst. Eng.* **1**, 391–399 (2009)
70. Wu, L., Garg, S.K., Buyya, R.: SLA-based resource allocation for software as a service provider (SaaS) in cloud computing environments. In: 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 195–204, May 2011. <https://doi.org/10.1109/CCGrid.2011.51>
71. Yu, Y., Bhatti, S.N.: Energy measurement for the cloud. In: International Symposium on Parallel and Distributed Processing with Applications, pp. 619–624 (2010)
72. Z Degraeve, F.R.: Improving the efficiency of the purchasing process using total cost of ownership information: the case of heating electrodes at Cockerill Sambre SA (1999)
73. Zhang, P., Han, Y., Zhao, Z., Wang, G.: Cost optimization of cloud-based data integration system. In: 2012 Ninth Web Information Systems and Applications Conference, pp. 183–188, November 2012. <https://doi.org/10.1109/WISA.2012.13>
74. Zheng, L., Hu, Y., Yang, C.: Design and research on private cloud computing architecture to support smart grid. In: 2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics, vol. 2, pp. 159–161, August 2011. <https://doi.org/10.1109/IHMSC.2011.109>