https://doi.org/10.53964/jia.2024005

https://www.innovationforever.com

Journal of
Information Analysis

ISSN 2959-1295 (Online)

Open Access

Research Article

Research on Data Risk Control Strategies for Hybrid Cloud

Yiqian Zhao[1], Sheng Zhu[1*]

[1]School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo, Henan Province, China

[*]**Correspondence to: Sheng Zhu, Associate Professor**, School of Mathematics and Information Science, Henan Polytechnic University, Shanyang District, Century Avenue, Jiaozuo, 454003, Henan Province, China; E-mail: shengzhu_ms@sina.com

Received: February 29, 2024 **Revised:** April 25, 2024 **Accepted:** May 17, 2024 **Published:** May 31, 2024

Abstract

Objective: Hybrid cloud provides an efficient and relatively secure service, widely used in commercial and public affairs. There are already several specific risk control strategies for hybrid clouds in the existing cloud computing literature, but there is still a need to design data risk control strategies that cater to users with different risk preferences.

Methods: Firstly, by using a classification method, we propose four data risk control strategies that are suitable for different cloud environments. Then, we use the queueing theory to model the hybrid cloud system and derive some relevant indicators of system performance.

Results: We propose four new risk control strategies for the personalized needs of enterprises in different scenarios, namely, L-control strategy, LOS$_1$-control strategy, LOS$_2$-control strategy, and (L,q)-control strategy.

Conclusion: It is more practical to use a hybrid cloud system for the protection of data security. This work proposes four risk control strategies for different situations to help companies with risk control, namely, L-control strategy, LOS$_1$-control strategy, LOS$_2$-control strategy, and (L,q)-control strategy. These risk control strategies provide theoretical assistance for exploring the allocation of cloud resources.

Keywords: cloud computing, hybrid cloud, risk control strategy, queueing system

Citation: Zhao Y, Zhu S. Research on Data Risk Control Strategies for Hybrid Cloud. *J Inform Anal*, 2024; 2: 5. DOI: 10.53964/jia.2024005

1 INTRODUCTION

Cloud computing is an Internet-based business computing method in which computing resources and services can be provided through virtualization[1] and distributed technology[2]. Cloud computing aims to connect businesses with the Internet. Previously, companies placed their business data in local data centers. However, due to the growing businesses and huge costs of maintaining local data storage space, more companies purchase storage space and other cloud products from the cloud service providers to meet their computing resource needs[3]. According to the type of service, the cloud computing platform provides three levels of service patterns[4]: infrastructure as a service , platform as a service , and software as a service.

As a new information technology (IT) service industry, cloud services manage resources in an elastic way, which improves the effectiveness of storage and enhances the utilization of servers and the quality of service. Faced with the different needs of customers, cloud services provide flexible service resources. Cloud computing has been rapidly developing and is gradually penetrating the traditional IT industry. According to the statistics of Gartner, the global cloud computing market is growing steadily, and global public cloud service end-user spending is expected to grow to 600 billion in 2023, with an increased rate of about 21.7% compared to 2022. According to Internet Data Center, the proportion of non-cloud expenses is expected to drop from 43.0% in 2022 to 31.4% in 2026. Global cloud infrastructure expenses are expected to surpass non-cloud expenses in 2023. This indicates that the real-time global cloud services industry is slowly replacing the traditional IT services industry. Based on the different needs of users, cloud computing is deployed in three ways: private cloud, public cloud, and hybrid cloud[5].

Private cloud. Private cloud is a proprietary service system that exists behind a company's firewall or is deployed in a secure hosting location. It is managed by the company and provides the most secure data processing. However, the private cloud has finite computing resources and is not flexible enough to adapt to changes in company demand.

Public cloud. Due to the expanding demand of companies, cloud services are offered commercially to customers. Users who request services only need to rent cloud resources by paying a fee to service providers. Public clouds can meet most of the basic needs of companies and are more elastic and scalable. Cloud service providers offer high performance, scalability, security, and availability. Users for public clouds only need to concentrate on their own jobs, while the cloud providers[6] will take care of cloud construction and maintenance. However, using public clouds does not guarantee the security of private and confidential data. When companies store data in a public cloud, they lack control over the resources in the cloud. So data and information are not adequately protected.

Hybrid cloud. To balance security and efficiency, companies usually use two types of cloud services. Private clouds serve some jobs that require higher security protection, and public clouds serve other jobs with high resource elasticity requirements. This type of service is called hybrid cloud service. It helps users manage their IT infrastructure to achieve a balance among operational risk, cost, and expense. As shown in Figure 1, the private cloud and the public cloud form a collective cloud service. In contrast to a single private cloud, the hybrid cloud is more suitable for complex IT environments.

Low-cost. Private clouds have better security, but their computing resources are finite. When a company's business grows, the choice to expand the computing resources of a private cloud will incur huge costs. In a hybrid cloud service system, a single cloud breakdown does not lead to the loss of all jobs, which lowers the cost of disaster recovery. In addition, if the company chooses to implement a multi-cloud architecture, it can help it better control costs. Multi-cloud services identify cloud-suitable service providers for cloud computing by finding a balance between work requirements and cost and selecting the lowest cost to meet the workload requirements.

Security. The implementation of a hybrid cloud has led to significant changes in IT infrastructure. The hybrid cloud strategy provides companies with a more secure and robust service environment. IT professionals maintain sensitive data and core jobs served by the private cloud, improving the overall security of the service[7].

Disaster recovery. The loss of core data will have a devastating impact on the company, so the continuity of job services is becoming more and more essential. One of the advantages of hybrid cloud services is that they have disaster recovery and backup capabilities. Under the hybrid cloud architecture, it can quickly switch the service to other clouds when a single cloud breaks down. The hybrid cloud system backups all of the company's historical data at the most efficient cost. Not only can the historical data be restored to the original site, but the private cloud can also be directly restored to the cloud server. The jobs can continue to operate on the cloud, and the disaster recovery of the jobs can be realized at extremely low costs.

Flexible scalability. The hybrid cloud helps companies manage their data and applications in the most flexible way. It stores the most valuable data in private clouds, enabling absolute control over the data. In the face of short-term job emergencies, the hybrid cloud provides flexible and efficient resources to address high-growth data computing. In addition, diversified cloud services are often more efficient and flexible than a single one. The company can personalize its cloud service facilities according to different job needs and corporate goals[8].

The hybrid cloud combines the advantages (Table 1) of private and public clouds and has attracted the interest of

https://doi.org/10.53964/jia.2024005

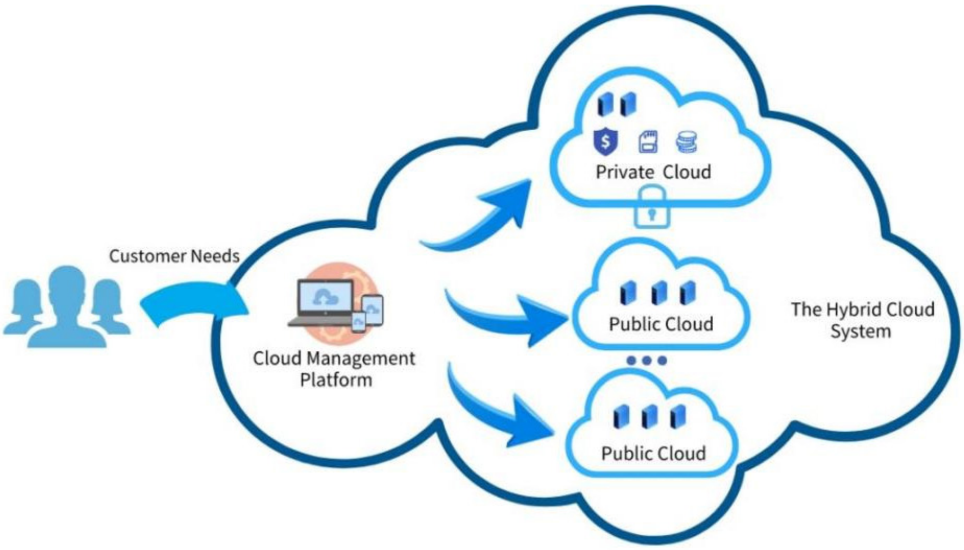

Figure 1. Illustration of hybrid cloud system.

most companies nowadays. In addition, companies can also choose to build multi-cloud technologies on hybrid cloud systems, which means they can use two or more public clouds at the same time to serve jobs. For example, users can host their databases on Amazon AWS and run applications on Microsoft Azure. Multi-cloud enables users to choose the most appropriate and customized cloud service platform for their business. The hybrid multi-cloud strategy has become a trend in cloud computing development in recent years. In May 2022, Cisco published the 2022 Global Hybrid Cloud Trends Report, indicating that 82% of companies choose to use a hybrid multi-cloud strategy to support applications. Statistical data also shows that in terms of infrastructure, 58% of companies choose to use 2-3 public clouds to serve their jobs, 31% choose to use 4-10 public clouds, and 3% use more than 10 public clouds. The report also shows that several smaller companies prefer to use multiple public cloud services to achieve optimal cost-effectiveness.

The cloud management platform[9] was first proposed by Gartner in the form of a product that is integrated and manages each cloud service in a unified manner. Based on the statistical data in Gartner's report, with the rapid growth of the cloud services industry, the market size of cloud management platforms is also gradually developing. In terms of investment, the services investment size of global cloud management has grown from 23.17 billion dollars in 2017 to 42.73 billion dollars by 2020. The advantages of the cloud management platform in computing, networking, and storage make it a vital tool for the future of cloud services.

As digital development increases, the issue of cloud resource management is attracting more and more attention from companies. Effective resource management has also become a concern for companies. Meanwhile, how to effectively manage cloud computing resources, reasonably arrange expense costs, and improve security have also become challenges for cloud platforms. In this paper, we provide various security risk control strategies for the cloud platform.

2 LITERATURE REVIEW

A lot of researchers from many fields have been paying close attention to the issue of cloud computing, and the related literature includes[10-18], and among others. Several critical research issues in cloud computing have also been proposed, including cloud resource allocation[10-14], performance analysis[15], and cloud security[16-18].

Various resource allocation strategies have been proposed for cloud computing. Kayalvili et al.[10] presented a model of cloud resource allocation based on the genetic algorithm. This model utilized the genetic algorithm's high-speed convergence and the ability for global optimization to solve the optimal allocation of cloud resources. Ben et al.[11] developed a new method that integrated the dynamic queues and the meta-heuristic algorithm, achieving the optimization of waiting time in the cloud, maximizing resource utilization, and providing good load balancing. Beegom et al.[12] proposed a new integer particle swarm algorithm and analyzed the performance of the proposed algorithm on more number of tasks and VM pairs. This method considered multi-objective optimization scenarios for task scheduling in cloud computing systems to achieve globally optimal resource scheduling. Oddi et al.[13] studied cloud resource allocation for heterogeneous virtualized cloud management agents. They proposed a resource allocation algorithm based on the Markov decision process to maximize the expected benefits of cloud management agents. Teng[14] provided a new resource pricing and allocation policy. Experimental results proved the resource price would gradually converge to an equilibrium state through a dynamic game.

At present, research on cloud computing resource allocation based on queueing theory becomes a hot issue.

Table 1. Comparison among Three Deployment Patterns

Deployment Pattern	Advantages	Disadvantages
private cloud	high security	high cost
public cloud	inexpensive, scalable on-demand	low security
hybrid cloud	reasonable cost, high disaster recovery	relatively low security

Many works have been published about the queueing theory, such as Ref.[19-22], and among others. Goswami et al.[19] constructed a queueing model with finite multi-servers, which controlled the number of virtual machines in real-time by cloud architecture. Nan et al.[20] considered resource optimization problems in the single-service case, the multiple-service case, and the priority-service case, respectively. They constructed a queueing model to characterize the service process in multimedia clouds and presented the relation between service time and resource allocation at different stages. Vilaplana et al.[21] obtained the relevant variables that the response distribution, customer arrival rate, number of servers, and service rate of a cloud system modeled on the M/M/m and M/M/1 queues. Recently, the strategic behavior of customers has evoked the interest of many researchers. Zhu et al.[22] proposed a two-tier cloud service system with a queue-length-based admission control mechanism and analyzed the equilibrium joining probability of customers' behavioral strategies. Interested readers can refer to Ref.[23-27]. Compared with Ref.[19-22], our contributions are listed as follows.

L-control strategy. Based on queueing theory, we propose several cloud resource control strategies and provide the most appropriate resource allocation scheme for the different service needs of companies. When the computing resources of a private cloud are sufficient or the level of job security requirements is relatively low, we use L-control strategy, arranging servers based on the available computing resources in the private cloud.

LOS_1- and LOS_2-control strategies. We design LOS_1-control strategy to send all high-security jobs to the private cloud. The level of security greatly meets the companies' compliance requirements, but the service is less efficient. In LOS_2-control strategy, most jobs with high-security requirements are sent to a private cloud, so companies' risk levels can meet the regulations. However, companies need to invest in infrastructure costs to build orbit space.

(L,q)-control strategy. We propose the (L,q)-control strategy, where jobs with low-security requirements can be sent to the public cloud without waiting. Although more costly, the service is more efficient.

3 SEVERAL RISK CONTROL STRATEGIES FOR HYBRID CLOUD SYSTEMS

Traditionally, a private cloud is an in-house server. It provides the most secure guarantees for the quality of service. Public cloud services are provided by cloud service providers, which have lower costs but usually do not meet security compliance requirements. A hybrid cloud is a combination of public cloud and private cloud. To balance efficiency and security, most companies have started to adopt a hybrid cloud to request computing resources. In this paper, we propose several different hybrid cloud risk control strategies.

We consider a two-tier hybrid system, which consists of a private cloud and a public cloud. The private cloud offers finite computing resources, and we assume that the queue-length limitation of the private cloud is L. When the queue length of the private cloud does not reach the threshold L, we claim that the computing resources of the private cloud are available. When the queue length of the private cloud reaches the threshold L, companies rent a public cloud from a cloud service provider. Jobs arrive at the cloud service system according to a random process with intensity λ. Companies decide how to triage service requests to different servers, for example, public or private cloud servers, based on real-time information feedback from monitors. Intuitively, when the real-time queue length in the private cloud reaches L, the service request will be rejected by the private cloud. Then these jobs will be sent to different servers based on different security compliance requirements.

3.1 L-control Strategy

There is no need to classify the types of jobs when the security level is not too high or when there are sufficient computing resources in the private cloud. In these cases, we consider using L-control strategy to control the cloud service system. The hybrid cloud system with L-control strategy is described as follows (Figure 2). When a company receives a service request/job, the controller first checks the real-time queue-length feedback in private cloud from the monitor unit. If the real-time queue length in the private cloud is less than the threshold L, the cloud controller will send the job into the private cloud; otherwise, the arriving job will be sent to the public cloud.

As the private cloud is an in-house service system, companies have invested a lot of money upfront. Even if the utilization of the private cloud is less than 100%, the system also needs to be maintained. So it is reasonable to prioritize the use of private clouds. When the computing resources of a private cloud are large, it is able to satisfy most of the service requests. The fact that a very small proportion of jobs are sent to the public cloud does not affect the security compliance. When the computing resource load of the private cloud reaches the threshold level, jobs are sent to the public cloud for obtaining service resources. Short-term rental services in the public cloud adapt to short-term changes. There is no need to expand the computing resources of the private cloud to adapt to the

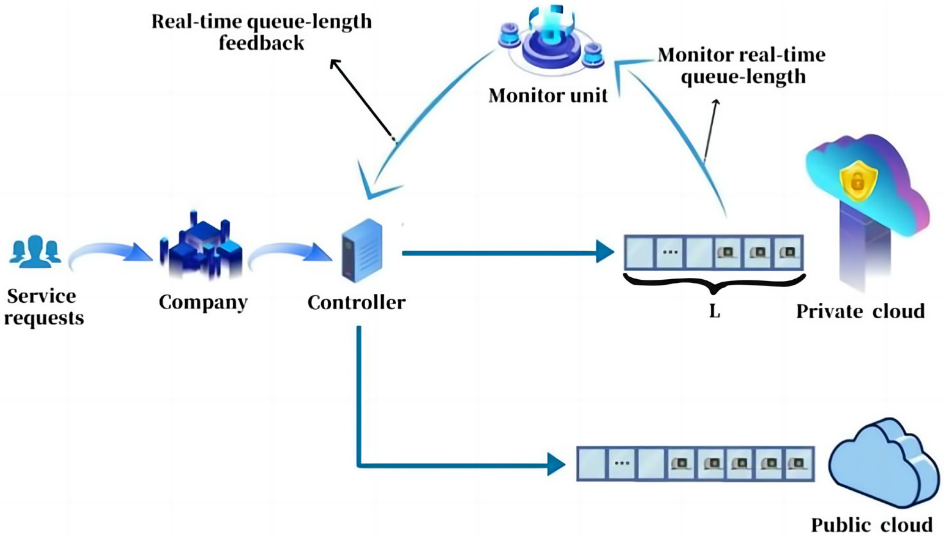

Figure 2. Hybrid cloud model under the L-control strategy.

unstable demand, which will not waste resources on free time. For example, at the end of the year, companies need to generate year-end summaries for a large amount of data. Users can go directly to short-rent data statistical resources from public cloud service providers without the need for private cloud resource expansion.

At times of peak business, companies only need to rent services from cloud service providers for a short period of time. The companies can reduce their rentals to public clouds after peak demand periods have passed. The cost of architecting cloud systems can be reduced. However, due to the uncertainty of business needs, companies may not be able to rent a public cloud in time when business demand increases. So the efficiency of service may be affected. In addition, the controller will reject jobs with high-security requirements after the queue length in the private cloud reaches the threshold. As the level of security requirements increases, the security level may not meet the compliance requirements of the companies.

3.2 LOS$_1$-control Strategy

With the increased security compliance of a company, we categorize jobs according to their security requirements. A controller sends jobs with different security requirements to different servers to ensure that the company's job security requirements are met. To state simplicity, a probabilistic method is used to describe the level of security requirements for the jobs. The percentage of jobs with high-security requirements among arrival jobs is q. The percentage of jobs with low-security requirements is 1-q.

We consider a hybrid service system with a private cloud and a public cloud with an orbit space (Figure 3). Before the controller sends jobs to different servers, it will classify those jobs according to their security level and then arranges suitable servers to serve them. When the private cloud has the available resources, this job is sent directly to the private cloud. Otherwise, the controller arranges servers according to the security type of the job. If the job has a low-security level, it will be sent to the public cloud; if it has high-security requirements, the controller will arrange for this job to stay in an orbit space, and then retry to join the queue in the private cloud with a rate of θ (called the retrial rate).

The controller schedules all jobs with high-security requirements to enter the private cloud. When computing resources in the private cloud are available, all jobs are sent to the private cloud; when computing resources are not available, jobs with high-security requirements are sent to the orbit space and then retry to join the queue in the private cloud with rate θ. These jobs are not sent to public cloud services, thereby reducing the cost that companies spend on public clouds. The orbit space provides opportunities for future access to services for jobs that may be lost due to a lack of computing resources. In addition, when the service resources of the private cloud are available, some jobs with lower security requirements can be sent to the private cloud, making full use of the private cloud's computing resources. However, since we allow ordinary jobs to have access to the private cloud when the computing resources of the private cloud are available, the queue length in the private cloud will quickly reach the threshold L. As the number of jobs with high-security requirements increases, the time they spend in the orbit space also increases. Although the service security of the company's business can be guaranteed, the efficiency of service and customer satisfaction decrease.

3.3 LOS$_2$-control Strategy

In this subsection, we consider the LOS$_2$-control strategy. In real-life situations, private clouds only accept service requests for jobs with high-security requirements when the number of jobs with high-security requirements is relatively low. Different from the LOS$_1$-control strategy, the orbit space is set on the private cloud in this subsection (Figure 4).

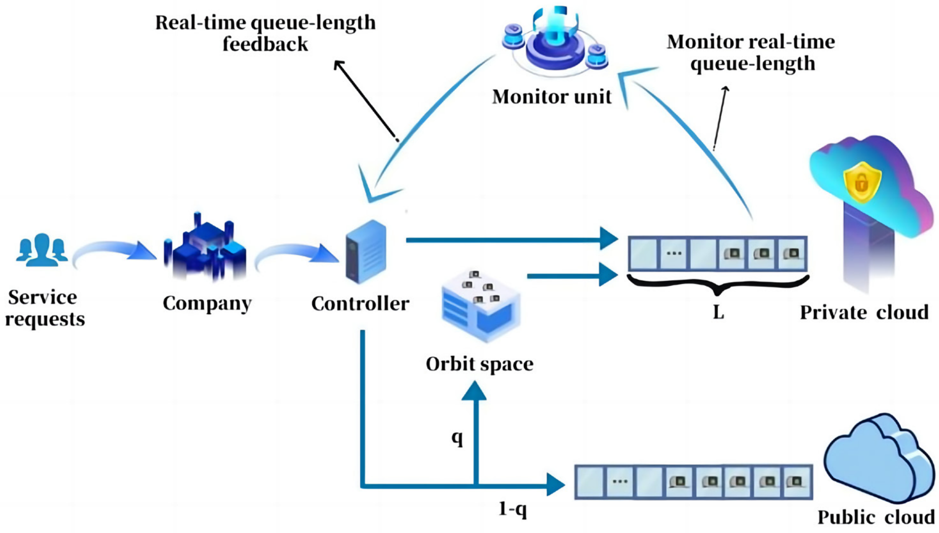

Figure 3. Hybrid cloud model under the LOS_1-control strategy.

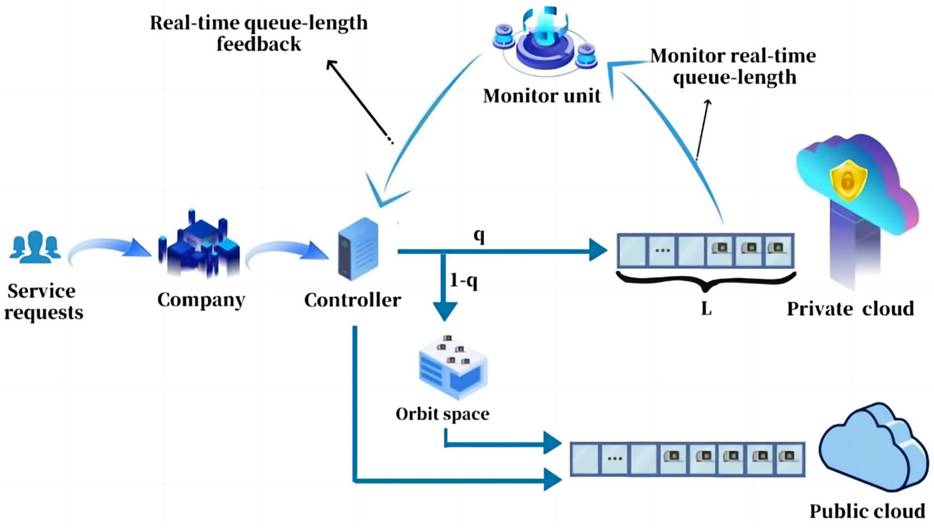

Figure 4. Hybrid cloud model under the LOS_2-control strategy.

When the queue length of the private cloud does not reach the threshold L, jobs with high-security requirements will be sent to the private cloud with probability *q*, and low-security requests will be sent to the public cloud with probability 1-*q*. However, the computing resources of the public cloud are not available until the queue length of service requirements in the private cloud reaches the threshold, thus service requests that are sent to the public cloud can only stay in an orbit space. These jobs that stay in orbit space retry to join the queue in the public cloud at a rate of θ. It may immediately enter the public cloud when public cloud computing resources are available. In addition, when the real-time queue length in the private cloud is greater than the threshold L, all job requests can only be sent to the public cloud.

All computing resources in the private cloud are occupied by jobs with high-security requirements. When the number of jobs with high-security requirements is small, the computing resources of the private cloud can meet these job service requests. So the system's security level can reach the company's compliance. Instead of allowing private clouds to accept all service requests, the controller diverts these requests to be sent to the orbit space with a certain probability. So the available computing resources in the private cloud will reach the threshold later than in the public cloud with orbit space. However, since the computing resources in the private cloud are limited, jobs with high-security requirements are also sent to the public cloud for service when the computing resources in the private cloud are not available. As the number of jobs with high-security requirements increases, the security level of the system may not be able to meet the company's security compliance. In addition, jobs with lower security requirements need to remain in the orbit space when private cloud computing resources are available. As a result, the service efficiency of these jobs may not be satisfactory, and jobs that remain in the orbit space require additional storage space from the system, which increases the cost of building the infrastructure for the company.

https://doi.org/10.53964/jia.2024005

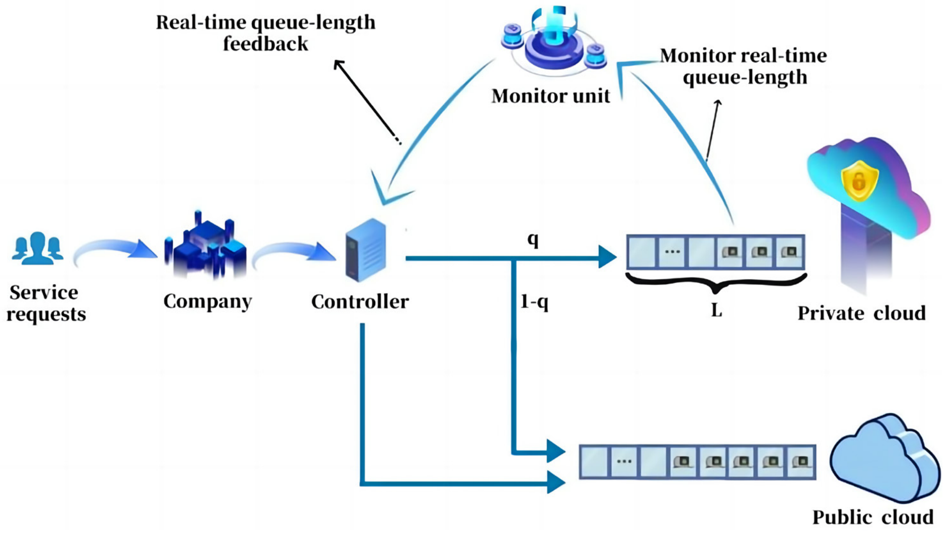

Figure 5. Hybrid cloud model under the (L,q)-control strategy.

3.4 (L,q)-control Strategy

In this subsystem, we consider a so-called (L,q)-control strategy (Figure 5). To improve the service efficiency for low-security level jobs, the service resources of the public cloud may also be available when the queue length in the private cloud does not reach the threshold. The hybrid cloud system will classify these jobs by security level and then arrange them on different servers, which can be summarized as follows. When private cloud computing resources are available, the controller sends jobs with low-security requirements to the private cloud and jobs with high-security requirements to the public cloud. Otherwise, after the queue length in the private cloud reaches the threshold, all jobs are sent to the public cloud.

We need to measure the service efficiency and security requirements of the companies. If the wait time for low-security level jobs in the private cloud is too long to meet service efficiency requirements, the company should choose to rent a public cloud. When the queue length in the private cloud does not reach the threshold, the controller arranges jobs with low-security requirements into the public cloud to improve the service efficiency. Different from the L-control strategy, the (L,q)-control strategy allocates servers by security level rather than simply by the available resources of the private cloud. The computing resources of the private cloud are fully utilized by jobs with high-security requirements, which increases the effective utilization of the private cloud. For jobs with high-security requirements, they are prioritized to be sent to the private cloud. Only if the resources of the private cloud are not available will these jobs be sent to the public cloud service. If high-security requirement jobs are few, the private cloud can serve all the high-security requirement jobs. The level of system riskiness meets the compliance requirements of companies. Compared to LOS$_2$-control strategy, jobs with low-security requirements do not stay in the orbit space when the private cloud queue length has not reached L. However, companies need to purchase public cloud space in the beginning, which is relatively costly in terms of cost. The system needs to find a balance among service efficiency, cost, and security to help companies adopt the optimal control strategy for hybrid cloud system.

4 IMPLEMENTATION AND APPLICATION SCENARIOS OF RISK CONTROL

In this section, we consider the implementation of risk control and its application scenarios, and pointed out which system parameters can be adjusted to achieve risk control levels of companies.

4.1 Implementation of Risk Control

The implementations of risk controls for all strategies are described in this section. The risk control of the hybrid cloud model is achieved by adjusting different parameters in our proposed models.

L-control strategy. The lower the number of jobs served by the public cloud, the higher the security level of the hybrid cloud. Under the L-control strategy, job requests are sent to the public cloud only after the queue length in the private cloud reaches a threshold of L. Therefore, by adjusting the size of L, the efficient arrival rate into the public cloud can be controlled. In the case of lower security compliance, the larger the threshold of computing resources for a company's in-house server, the less jobs are sent to the public cloud. As a result, the security risk of the cloud service system can be reduced.

LOS$_1$-control strategy. Under the LOS$_1$-control strategy, due to the setting of orbit space, all the jobs with high-security requirements are sent to the private cloud to receive the most secure service. The private cloud is the most secure server. If the computing resource threshold L of the private cloud is set higher, jobs with low-security requirements are also able to be sent to the private cloud even when the queue length in the private cloud has not reached the threshold. In addition, after the queue length in the private cloud reaches the threshold, if the percentage 1-*q* of jobs with low-security requirements is less, the number of jobs sent to the public cloud will also decrease. Therefore, the risk control of the system can be guaranteed by setting an appropriate L and an orbit space.

LOS$_2$-control strategy. Under the LOS$_2$-control strategy, the system risk level is ensured by controlling the number of jobs served by the public cloud. Firstly, until the queue length of the private cloud does not reach the threshold, the jobs being served by the public cloud are the ones staying in orbit space. If fewer jobs are staying in orbit space, fewer jobs will be sent to the public cloud. So the system can control the percentage 1-*q* of jobs with low-security requirements to limit the number of jobs being sent to the public cloud. In addition, after the queue length in the private cloud reaches the threshold, all the jobs are sent to the public cloud. If the computing resource threshold L of the private cloud is larger, the queue length will reach the threshold level later. As a result, the number of jobs being sent to the public cloud will also decrease. The security risk of the system can then be assured.

(L,q)-control strategy. Since the (L,q)-control strategy differs from the LOS$_2$-control strategy only in the existence of the orbit space, the system can adjust the same parameters to control the risk level. In the case of the availability of computing resources in the private cloud, the system controls the number of jobs being sent to the public cloud by adjusting the percentage 1-*q* of service requests with low-security requirements. In addition, the system can increase the queue threshold L for the private cloud, reducing the possibility of sending high-security requirements to the public cloud.

4.2 Mathematical Models and Computational Complexity Analysis

In this section, we present a brief model description for the proposed risk control strategies and analyze mathematical challenges that may exist for each model calculation. We throughout assume that the inter-arrival times of jobs/requests are independent and exponentially distributed with arrival rate λ and the service rate for the private (or public) cloud is μ_1 (or μ_2). In addition, we assume system capacity of the private cloud is L. Performance measures of the hybrid cloud are essential in analyzing the impact of various control strategies on the hybrid cloud. To derive the performance measures under various strategies, the hybrid cloud is modeled as different types of queueing systems. However, in this paper, we have only considered the model construction of the hybrid cloud as well as given only relatively simple system performance measures. We will concentrate on the detailed calculations of the hybrid cloud performance measures in future studies.

L-control strategy. Under the L-control strategy, the hybrid cloud system can be modelled as a two-tier queueing service system where the private cloud (or the public cloud) is characterized as an M/M/1/L queueing system with limited system capacity (or M/M/∞) queueing system with infinite system capacity. Based on basic result of the M/M/1/L queueing (Gross et al.' work[28]), the steady state probability of having i jobs/requests in the private cloud is Equation (1):

$$p_i = \frac{(1-\rho_1)\rho_1{}^i}{1-\rho_1{}^{L+1}}, i = 0, \dots (1)$$

Where $\rho_{1=}\lambda_q/\mu_1$. From Equation (1), the mean number of jobs newly arriving in the private cloud, N$_i$ can be computed from Equation (2):

$$N_i = \sum_{i=0}^{L} ip_i = \frac{\rho_1 - \rho_1{}^{L+1} - (1-\rho_1)L\rho_1{}^{L+1}}{(1-\rho_1)(1-\rho_1{}^{L+1})}, i = 0, \dots (2)$$

According to Little's law, we get the mean waiting time of a job newly arriving in the private cloud can be expressed as follows:

$$W_i = \frac{N_i}{\lambda}, i = 0, \dots (3)$$

In addition, from Equation (1), the probability that a job newly arriving is sent to the public cloud can be expressed as follows:

$$p_L = \frac{(1-\rho_1)\rho_1{}^L}{1-\rho_1{}^{L+1}} \ (4)$$

In the above, we have obtained various performance measures about the system. Based on the above results, We will further analyze how the managers can set optimal system capacity of the private cloud under the L-control strategy in the future.

LOS$_1$-control strategy. Under the LOS$_1$-control strategy, we also adopt an M/M/1/L queueing model to describe the private cloud. Similar to the model for L-control strategy, we can derive the steady state probability of having i jobs/requests in the private cloud, namely.

$$p_i = \frac{(1-\rho_1)\rho_1{}^i}{1-\rho_1{}^{L+1}}, i = 0, \dots (5)$$

Based on the steady state probability distribution, the mean number of jobs newly arriving in the private cloud can also be obtained as follows:

$$N_i = \sum_{i=0}^{L} ip_i = \frac{\rho_1 - \rho_1{}^{L+1} - (1-\rho_1)L\rho_1{}^{L+1}}{(1-\rho_1)(1-\rho_1{}^{L+1})}, i = 0, \dots (6)$$

W_i can be computed using Little's law.

$$W_i = \frac{N_i}{\lambda}, i = 0, \dots (7)$$

In addition, the number of jobs in the orbit space can be obtained by the balance equations. First, we use a pair (I (t),N (t)) to represent the state of the system at time t, where I (t) and N (t) denote the number of jobs in the private cloud and the number of jobs in the orbit space, respectively. Obviously, {(I (t),N (t)),t≧0} is a continuous time Markov chain and its state space Ω={(i,j):0≤i≤L, j≥0}. Let p (i,j) be

Table 2. Application Scenarios for Four Strategies

Strategy	Application Scenarios
L-control strategy	Low-security compliance requirements and adequate computing resources in the private cloud
LOS$_1$-control strategy	Highest security compliance and lower service efficiency requirements for jobs
LOS$_2$-control strategy	A lower quantity of jobs with high-security requirements and higher security compliance
(L,q)-control strategy	Higher service efficiency requirements for jobs and higher security compliance

the probability that (I (t),N (t))stays at the state (i,j). Then we will get the balance equations as follows:

$$\lambda p(0,0)=\mu_1 p(1,0) \qquad (8)$$

$$(\lambda+\mu_1)p(i,0)=\mu_1 p(i+1,0)+\theta p(i-1,1) + \lambda p(i-1,0), i=1,\dots \qquad (9)$$

$$(\lambda q+\mu_1)p(L,0)=\theta p(L-1,1)+\lambda p(L-1,0) \qquad (10)$$

$$(\lambda+j\theta)p(0,j)=\mu_1 p(1,j), j=1,\dots \qquad (11)$$

$$(\lambda+\mu_1+j\theta)p(i,j)=\mu_1 p(i+1,j) + (j+1)\theta p(i-1,j+1)+\lambda p(i-1,j), i=1,\dots, j=1,\dots \qquad (12)$$

$$(\mu_1+\lambda q)p(L,j)=\lambda q p(L,j-1) + (j+1)\theta p(L-1,j+1)+\lambda p(L-1,j), j=1,\dots \qquad (13)$$

The calculation for solving the balance equations is complicated and may require the high-performance computers and advanced programming procedures. We ignore the corresponding mathematical calculation and focus on several risk control strategies and model construction for hybrid cloud. We derive various performance measures based on the steady-state probability distribution, such as the mean number of jobs in orbit space, N$_{orbit}$, can be expressed as Equation (14):

$$N_{orbit}=\sum_{i=0}^{L}\sum_{j=0}^{\infty} ip(i,j) \ \ (14)$$

LOS$_2$-control strategy. The hybrid cloud system with LOS$_2$-control strategy has been studied by Zhu et al.[22]. They derived relevant performance measures, such as the probability of having i jobs/requests in the private cloud, the mean waiting time of a job newly arriving in the private cloud, the mean number of jobs in the orbit, the mean waiting time of jobs in the orbit and so on. Based on these performance measures of the system model, Zhu et al.[22] further explored the conditional equilibrium joining probability of entering the private cloud when the public cloud is not open, and they determined the cooperatively optimal retrial rate and the noncooperatively optimal (L,q)-controletrial rate for a given queue-length information.

Under the (L,q)-control strategy, the hybrid cloud system can be modelled as a two-tier queueing service system. In the system, the private cloud is characterized as an M/M/1/L queueing system with arrival rate λ and service rate μ_1, while the public cloud is characterized as an M/M/∞ queueing system with service rate μ_2. The arrival rate of jobs in the public cloud can be expressed as λ and λ(1-q). When the queue length in the private cloud is less than L, the arrival rate of jobs in the public cloud can be expressed as λ(1-q). If not, the arrival rate of jobs in the public cloud can be expressed as λ.

Then the steady state probability of having i jobs in the private cloud can be obtained as follows:

$$p_i=\frac{(1-\rho_2 q)(\rho_2 q)^i}{1-(\rho_2 q)^{L+1}}, i=0,\dots (15)$$

Where $\rho_{1=}\lambda_q/\mu_1$. We represent the state of the system at time t by a pair (I (t),Q (t)), where I (t) and Q (t) denote the number of jobs in the private cloud and the number of jobs in the public cloud, respectively. Then we will get the following balance equations as follows:

$$\lambda p(0,0)=\mu_1 p(1,0)+\mu_2 p(0,1) \qquad (16)$$

$$(\lambda+\mu_1)p(i,0)=\mu_1 p(i+1,0)+\mu_2 p(i,1) + \lambda q p(i-1,0), i=1,\dots,L-1 \qquad (17)$$

$$(\lambda+\mu_1)p(L,0)=\mu_2 p(L,1)+\lambda q p(L-1,0) \qquad (18)$$

$$(\lambda+\mu_2)p(0,j)=\mu_1 p(1,j)+\mu_2 p(0,j+1) + \lambda(1-q)p(0,j-1), j=1,\dots \qquad (19)$$

$$(\lambda+\mu_1+\mu_2)p(i,j)=\mu_1 p(i+1,j)+\mu_2 p(i,j+1) + \lambda(1-q)p(i,j-1)+\lambda q p(i-1,j), i=1,\dots, j=1,\dots \qquad (20)$$

$$(\lambda+\mu_1+\mu_2)p(L,j)=\mu_2 p(L,j+1) + \lambda p(L,j-1)+\lambda q p(L-1,j), j=1,\dots \qquad (21)$$

It should be feasible to obtain the performance measures of the hybrid cloud based on the balance equations. After solving the Equations (16-21), we can get the steady state probability of the system, but it is complex to calculate these equations. We can use the matrix analytical method to solve these equations, which is our future work.

4.3 Application Scenarios

For different types of businesses in terms of quantity, security compliance, and service efficiency requirements, the strategies proposed in this paper are suitable for different scenarios. When a company has a low level of job security requirements or when the private cloud computing resources are sufficient, it can use the L-control strategy. The system does not classify jobs and selects to use public cloud computing resources only when private cloud computing resources are not available. The company uses the LOS$_1$-control strategy when it needs the highest security compliance and lower service efficiency requirements for high-security jobs. All jobs with high-security requirements are sent to the private cloud, so this is the safest strategy. However, some jobs need to stay in orbit space, so their services are less efficient. Because the higher the percentage of jobs with high security requirements, the fewer jobs are sent to the public cloud. In order to achieve higher security compliance, the company can choose to adopt LOS_2-control strategy. If the company has a high requirement for service efficiency, it should select the (L,q)-control strategy that does not set orbit space. The following table is a summary of the scenarios where the different strategies are applicable (Table 2).

5 CONCLUSIONS AND FUTURE WORKS

In this paper, we have summarized the advantages and disadvantages of public, private and hybrid clouds. By contrast, hybrid clouds are better suited to meet efficiency and security needs. We focused on risk control strategies for hybrid cloud service systems. Most of the previous literature has not considered specific risk control strategies. This work proposed four risk control strategies for different situations to help companies with risk control, namely, L-control strategy, LOS_1-control strategy, LOS_2-control strategy, and (L,q)-control strategy. We compared the advantages and disadvantages of these control strategies and provided corresponding application scenarios. This paper did not investigate the specific computation of each control strategy but simply gave several ideas for risk control for hybrid service systems. For future works, we list the following issues.

Based on our proposed models, we will explore various performance measures of these systems to obtain optimal strategies. How should the volume of the private cloud be set? How to control the arrival rate in the public cloud? In addition, we will consider the behavioral strategies of customers and explore the optimal strategies of customers in cooperative and non-cooperative cases, respectively.

Customers and service providers are different stakeholders. There will be market competition between them. The equilibrium after the dynamic game between customers and service providers is worth studying.

Acknowledgements

The authors were thankful to the anonymous reviewers for their insightful comments which significantly improved the quality of our papers. This work was supported in part by the key scientific and technological research projects in Henan province of China (Grant No. 242102210137), in part by the fundamental research funds for the universities of Henan province (Grant No. NSFRF220442), in part by the doctoral fund projects of Henan polytechnic university (Grant No. B2020-36), and in part by the “Double First Class” discipline creation project of surveying and mapping science and technology of Henan polytechnic university (Grant No. GCCRC202307).

Conflicts of Interest

The authors declared no conflict of interest.

Author Contribution

Zhao Y and Zhu S completed the theoretical analysis, drafted the manuscript, contributed to the writing of the article, read and approved the article for submission.

Abbreviation List

IT, Information technology

References

[1] Chowdhury NM, Boutaba R. A survey of network virtualization. *Comput Net*, 2010; 54: 862-876.[DOI]

[2] Bende S, Shedge R. Dealing with small files problem in hadoop distributed file system. *Procedia Comput Sci*, 2016; 79: 1001-1002.[DOI]

[3] Mei J, Li K, Ouyang A et al. A profit maximization scheme with guaranteed quality of service in cloud computing. *IEEE Trans Comput*, 2015; 64: 3064-3078.[DOI]

[4] Mell P, Grance T. The NIST Definition of Cloud Computing. Accessed May 17 2024. Available at:[Web]

[5] Liu F, Tong J, Mao J et al. NIST cloud computing reference architecture. *NIST Spec Pub*, 2011; 500: 1-28.[DOI]

[6] Hu H, Wang H. A prediction-based aco algorithm to dynamic tasks scheduling in cloud environment: Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications, Chengdu, China, 14-17 October 2016.[DOI]

[7] Achar S. An Overview of Environmental Scalability and Security in Hybrid Cloud Infrastructure Designs. *Asia Pac j energy environ*, 2021; 8: 39-46.[DOI]

[8] Li J, Han Y. A hybrid multi-objective artificial bee colony algorithm for flexible task scheduling problems in cloud computing system. *Cluster Comput*, 2020; 23: 2483-2499.[DOI]

[9] Malik SUR, Khan SU, Srinivasan SK. Modeling and analysis of state-of-the-art VM-based cloud management platforms. *IEEE Trans on Cloud Comput*, 2013; 1:1.[DOI]

[10] Kayalvili S, Selvam M. Hybrid SFLA-GA algorithm for an optimal resource allocation in cloud. *Cluster Comput*, 2019; 22: 3165-3173.[DOI]

[11] Ben Alla H, Ben Alla S, Touhafi A et al. A novel task scheduling approach based on dynamic queues and hybrid meta-heuristic algorithms for cloud computing environment. *Cluster Comput*, 2018; 21: 1797-1820.[DOI]

[12] Beegom ASA, Rajasree MS. Integer-pso: a discrete pso algorithm for task scheduling in cloud computing systems. *Evol Intell*, 2019; 12: 227-239.[DOI]

[13] Oddi G, Panfili M, Pietrabissa A. A resource allocation algorithm of multi-cloud resources based on markov decision process: Proceedings of the 2013 IEEE 5th International Conference on Cloud Computing Technology and Science. Bristol, UK, 02-05 December 2013.[DOI]

[14] Teng F, Magoules F. Resource pricing and equilibrium allocation policy in cloud computing: Proceedings of the 2010 10th IEEE International Conference on Computer and Information Technology. Bradford, UK, 29 June 2010-1 July 2010.[DOI]

[15] Huang D, Zhang X, Kang M. MobiCloud: building secure cloud framework for mobile computing and communication: Proceedings of the 2010 Fifth IEEE International Symposium on Service Oriented System Engineering. Nanjing, China, 04-05 June 2010.[DOI]

[16] Brender N, Markov I. Risk perception and risk management in cloud computing: Results from a case study of Swiss companies. *Int J Inform Manage*, 2013; 33: 726-733.[DOI]

[17] Rimal BP, Choi E, Lumb I. A taxonomy and survey of cloud computing systems: Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC, Seoul, Korea (South), 25-27 August 2009.[DOI]

[18] Abdulsalam YS, Hedabou M. Security and privacy in cloud computing: technical review. *Future Int*, 2021; 14: 11.[DOI]

[19] Goswami V, Patra SS, Mund GB. Performance analysis of cloud with queue-dependent virtual machines: Proceedings of the 2012 1st International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 15-17 March 2012.[DOI]

[20] Nan X, He Y, Guan L. Optimal resource allocation for multimedia cloud based on queuing model: Proceedings of the 2011 IEEE 13th International Workshop on Multimedia Signal Processing, Hangzhou, China, 17-19 October 2011.[DOI]

[21] Vilaplana J, Solsona F, Teixidó I et al. A queuing theory model for cloud computing. *J Supercomput*, 2014; 69: 492-507.[DOI]

[22] Zhu S, Wang J, Li WW. Cloud or In-House Service? Strategic Joining and Social Optimality in Hybrid Service Systems With Retrial Orbit. *IEEE Syst J*, 2023; 2: 1-12.[DOI]

[23] Burnetas A, Economou A. Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Syst*, 2007; 56: 213-228.[DOI]

[24] Burnetas A, Economou A, Vasiliadis G. Strategic customer behavior in a queueing system with delayed observations. *Queueing Syst*, 2017; 86: 389-418.[DOI]

[25] Guo P, Hassin R. Strategic behavior and social optimization in Markovian vacation queues. *Oper Res*, 2011; 59: 986-997.[DOI]

[26] Xu B, Xu X. Equilibrium strategic behavior of customers in the M/M/1 queue with partial failures and repairs. *Oper Res*, 2018; 18: 273-292.[DOI]

[27] Zhu S, Wang J, Li WW. Optimal pricing strategies in cognitive radio networks with multiple spectrums. *IEEE Syst J*, 2020; 15: 4210-4220.[DOI]

[28] Shortle JF, Thompson JM, Gross D et al. Fundamentals of queueing theory, 3th ed. Wiley-Blackwell Press: New Jersey, USA, 2018.