Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.054102

ARTICLE

AMAD: Adaptive Mapping Approach for Datacenter Networks, an Energy-Friend Resource Allocation Framework via Repeated Leader Follower Game

Ahmad Nahar Quttoum[1,*] and Muteb Alshammari[2]

[1]Department of Computer Engineering, Faculty of Engineering, The Hashemite University, Zarqa, 13133, Jordan

[2]Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Arar, 91431, Saudi Arabia

*Corresponding Author: Ahmad Nahar Quttoum. Email: quttoum@hu.edu.jo

Received: 19 May 2024 Accepted: 11 August 2024 Published: 12 September 2024

ABSTRACT

Cloud Datacenter Network (CDN) providers usually have the option to scale their network structures to allow for far more resource capacities, though such scaling options may come with exponential costs that contradict their utility objectives. Yet, besides the cost of the physical assets and network resources, such scaling may also impose more loads on the electricity power grids to feed the added nodes with the required energy to run and cool, which comes with extra costs too. Thus, those CDN providers who utilize their resources better can certainly afford their services at lower price-units when compared to others who simply choose the scaling solutions. Resource utilization is a quite challenging process; indeed, clients of CDNs usually tend to exaggerate their true resource requirements when they lease their resources. Service providers are committed to their clients with Service Level Agreements (SLAs). Therefore, any amendment to the resource allocations needs to be approved by the clients first. In this work, we propose deploying a Stackelberg leadership framework to formulate a negotiation game between the cloud service providers and their client tenants. Through this, the providers seek to retrieve those leased unused resources from their clients. Cooperation is not expected from the clients, and they may ask high price units to return their extra resources to the provider's premises. Hence, to motivate cooperation in such a non-cooperative game, as an extension to the Vickery auctions, we developed an incentive-compatible pricing model for the returned resources. Moreover, we also proposed building a behavior belief function that shapes the way of negotiation and compensation for each client. Compared to other benchmark models, the assessment results show that our proposed models provide for timely negotiation schemes, allowing for better resource utilization rates, higher utilities, and grid-friend CDNs.

KEYWORDS

Data center networks; energy-aware resource management; resource utilization; game-theory mechanisms

1 Introduction

Nowadays, the era of computing and digital telecommunication systems has changed almost every aspect of our lives, starting with education, business sectors, medical applications, and many other service examples that we may find everywhere. For organizations, to afford such services, the option of retaining physical network infrastructures and what is required of powerful computing and processing stations have phased out. The theme of cloud computing has proved to allow for a sufficient alternative that provides adequate virtual environments to host such sorts of services in an efficient, reliable, and cost-competitive way. Cloud-based Datacenter Networks (CDNs) are built with nodes allowing massive processing power and ultimate network capacities. However, service applications that are hosted on such data center networks are increasing every day, and so are the resource requirements of the hosting data center nodes and their associated networked traffic.

In CDNs, the providers are required to guarantee their continuous service availability and reliability. To do so, such cloud providers employ well-engineered resource management schemes that map their data center resources in efficient ways. Usually, Service Level Agreements (SLAs) that shape commitments of reliability, service availability, classes, and price units cover such service provision processes. To accommodate the rapid increase in the number of new services and their associated resource requirements, providers of CDNs need to adapt by expanding their resource fabrics to allow for far more capacities at both, nodes and network links. This option may help accommodate the required increase of resources; however, it also conveys other challenges. Indeed, such an option may add greater administrative overhead and higher service price units. Furthermore, expanding the scale of a data center may require major changes to the whole infrastructure of the data center, not to mention what it imposes on higher requirements of energy for such network nodes to run and cool.

Besides the other traditional resource management problems, the research work today is required to tackle the issue of power and energy consumption of such CDNs [1]. The number of computing devices that are connected to the power grids is increasing rapidly; what is more, the diffusion of other power-hungry devices like electric vehicles represents a non-stationary sort of energy consumer that adds sudden consumption hikes to the power grids. Such unexpected hikes may compromise the power grids' stability; therefore, besides the traditional objectives of resource utilization, the need today is to consider the energy consumption rates of the employed machines in such data center networks [2].

Services of the CDNs are provided with commitments to predefined SLAs that dictate the resource amounts to be allocated, and their assigned price units to be paid. Virtual Network Clients (VNCs) lease their required resources in the form of Virtual Machines (VMs) from cloud providers, these VMs vary in their computing and processing power based on the services they are leased to run. Therefore, VNCs requests of VMs vary according to their services and required specifications. Statistics show that the VNCs tend to lease extra resources compared to their real VM requirements [3], the real usage rates are mostly around 20% of the leased amounts of resources. Although the CDNs are paid for such leased but unused resources, still, in the terminology of resource management, this leads to low resource utilization rates as the resources are reserved but not used. For CDN providers, any commitments to new VNCs should consider the resource availability at their premises. Providers of CDNs are always motivated to expand their engagement in leasing their resources, which helps in maximizing their revenue and attaining higher reputations and market publicity. Accordingly, providers may expand their data center networks by adding more server racks and necessary networking resources to accommodate new VNC allocation requests in new hosting spaces. This may allow for new resource capacities indeed, however, such expansion approaches come with new costs to be counted, and also require more energy to run and cool these new racks and the other related network components. Such networks with huge energy consumption rates may represent heavy-loaded zones on the power grids, which may further lead to power failures and service disruptions. It is worth highlighting that such additions of server racks to the existing data center networks may necessitate major updates to the whole network infrastructure, which may also impose further physical changes with higher costs that exceed those of the newly added components and the anticipated revenue rates.

2 Problem Statement

Resource utilization plays an important role in delivering energy-friend data center network fabrics with competing service price units and satisfying SLA commitments. In the literature, the proposals of resource management and clients' mapping models in cloud-based networks are many, each with different methodology and management objectives. VM migration and the other techniques of resource adjustments and resource arbitration may help in accommodating new leasing requests of VNCs using the existing data center resources without incurring further costs [4,5], however, such complex techniques may result in service disruptions that violate the Quality of Service (QoS) commitments govern by the SLAs. QoS violations may require the cloud providers to pay the hosted VNCs compensation costs that exceed the anticipated revenue to be collected from the new VNCs leasing requests.

In [6], we proposed an allocation model that suppresses resource request exaggeration, in that work; the CDN providers motivate leasing fewer resources by adopting a non-linear pricing scheme [7] for their data center networks' resources. Therefore, for the clients, the more resource units they lease, the higher price-units they would expect to pay, and so, this may affect their anticipated utilities and profit objectives. In this work, we are extending our work of [6] and [8] by proposing another resource allocation methodology that helps mitigate the effects of VNCs' tendency of resource exaggeration, through which, we developed a negotiation model that allows for a reallocation scheme for those leased but unused resources. CDN providers can increase their utilities by accepting more VNC leasing requests; however, such new requests could be satisfied by increasing the availability of the existing resources in their networks without the need to upgrade their topologies to other ones with more new resources. Such upgrades may require major changes to their assets and their related configurations. Through this model, providers of CDN can retrieve their leased but unused resources from the existing VNCs and reallocate them to other new VNCs. This would provide the resources to the new VNCs with lower price units, increase the reputation of the leasing CDNs provider, avoid any SLA violations, and decrease the need for further energy consumption rates from the power grids.

In [9], the authors proposed a negotiation model that helps avoid resource underutilization at the service tenants' side of cloud service networks. In their work, they proposed a set of negotiation strategies between the service providers and their service tenants to return the leased but unused resources to the provider's premises. In one strategy, they proposed a kind of greedy pricing model that slightly increases the provider's offered price units at each negotiation round. The objective of such a strategy is merely to return the extra resources at the lowest possible prices, regardless of the tenants' utility objectives or the time spent in negotiation rounds. The outcome of the strategy will be a compensation price unit that is the same for all service tenants, such as a price unit that conserves the provider's utility function and maximizes its profit objectives. In another strategy of the same work, but with different price units being offered, the authors proposed considering the mean and standard deviation metrics to show how the increments in the offered price units at the previous negotiation rounds could affect the number of negotiation rounds in the coming sessions. This may reduce the number of negotiation rounds with maybe bigger jumps in the offered price-units indeed, though; it cannot motivate the cooperation of the tenants to accept the providers' offered price-units instead of waiting for the next rounds. From the perspective of the service tenants, those we called the VNCs, retaining unused resources might be an option that may cover their future needs or any unexpected changes in the service requirements. However, offering them reasonable return price units might be convincing as this may allow for higher utilities being earned by reducing the prices they pay for their chosen service resources. Moreover, we need to keep in mind that the whole resource retrieval process was originally motivated to satisfy the new VNCs resource allocation requests, such a process that needs to be processed in a fast manner with no long delays. Hence, the process of price-unit offerings and its negotiation rounds need to be timely. Therefore, we believe that besides the need to make it fast by reducing the number of negotiation rounds, such a process needs to be incentive-compatible for both, service providers and their tenants.

Accordingly, in this work, we are presenting AMAD, an Adaptive Mapping Approach for data center processing and network resources for energy-efficient CDNs. In AMAD, through a negotiation process with the existing VNCs, those leased but unused resources could be returned to the CDN providers' premises so they can use them to serve other new VNCs. The negotiation process is done periodically according to the dynamic resource requirements a CDN provider may find needed to satisfy the new leasing requests of new VNCs. In this way, the existing VNCs will be compensated for the resources they return, and accordingly satisfied, as they will be paying less for their running services. So do the CDN providers, they will be able to offer their hosting services to new VNCs using the existing resource expenditures without incurring further higher costs for the upgrades. What is more, with AMAD, such a resource management scheme allows for green and energy-friend CDNs that consume less energy rates for the number of clients and services they provide.

For such a negotiation-based resource retrieval process, in AMAD, we chose to employ a modified version of the Stackelberg Leadership model [10] to create a kind of strategic game between the CDN's providers and their existing VNCs, being the two players of the game. In economics, Stackelberg strategic games between two main entities, in which, the first entity (i.e., the leader) performs its move first, and then the other (i.e., the followers) does their move next. It is therefore called leader-follower games. In the terminology of game theory, these two entities (i.e., the leader and the followers) are called players, and they both compete to maximize their utilities. Accordingly, the leader is the CDN provider, which we denote as the first player, while the followers, which we denote as the second player, are the set of VNCs who have extra-unused resources to return. Running such a game is periodic so it periodically retrieves what the CDN may need to satisfy the new VNCs leasing requirements. Among the rules that we set for this resource-retrieval game are the following: neither the leader nor the followers should have prior knowledge of the opponent's action (i.e., the price-unit) that is willing to accept, though, the leader must have a committing power restricted by the model constraints to be discussed in the coming sections.

Therefore, in AMAD, the leader initiates the resource-retrieval game by revealing its offered price unit with no information about the amount of required resources. After that, the followers respond by either accepting the offer or not (i.e., waiting for higher offers in the next negotiation rounds), for those who accept, and have only one chance to do so, they declare the amount of resources they intend to return. The offered price-unit revealed by the leader is defined according to one of the following two strategies:

- The leader starts the negotiation game by allowing the followers to wait for a price unit that matches their desired objectives; however, it sets the final price unit to be paid to all the followers according to the highest price (FHP) that is accepted by any of the followers participating in the current retrieval game. This motivates the follower VNCs to accept early offered price units and not wait for further negotiation rounds, as this may lead to them being excluded from the retrieval game.

- The leader sets each follower's final price unit individually according to a belief function. Such a function is proposed to build a history of the follower's previous bidding behavior and negotiation tendency. Based on that, the model accepts/declines the follower's requests to wait for the next negotiation rounds. This strategy may help motivate better bidding behavior (i.e., accepting reasonable price units being offered) and reduce the number of negotiation rounds, which means faster processing for the new resource mapping requests of new VNCs.

Different from other proposals in the literature, the followers' cooperation in such resource-retrieval games is motivated through mechanisms that guarantee no utility losses for both parties of the game, neither the leader nor the followers. For the leader player, the CDN provider, its utility objective is to retrieve the extra-unused resources from the existing VNCs and use them to serve other requests of new VNCs. For such an objective to be satisfied, the retrieval price units need to be reasonable, and not exceed price thresholds being set according to its revenue goals and original costs already paid for the running resources. It also needs to be timely with no lengthy negotiation rounds. For the followers, those leasing VNCs with extra-unused resources, their utility is to satisfy their required services and reduce the cost receipts associated with such services. Therefore, for the followers, returning the extra-unused resources by any price could be considered profitable, though; the strategies followed in our model AMAD guarantee reasonable prices being offered for such resources from the CDN providers. Through such a framework, we can create a pool of unused resources that can be used to fulfill part of the new service requests without imposing any burdens of any structural upgrades that would certainly come with higher costs and higher power consumption rates.

2.1 Contribution

Hence, through the proposed model AMAD, our contribution can be summarized in developing a grid-friend resource management model for CDNs that allows for:

- Efficient resource utilization; where the network processing and computing resources are allocated in a well-engineered manner that guarantees both, higher *utilization* and *client satisfaction* rates.
- Green data center networks; given that any scale or further expansion plans for the datacenter nodes or network resources are only allowed after verifying the true usage and resource availability. Hence, limiting the need for structural updates or upgrades. This helps in reducing the number of physical machines in the data center to the least appropriate, and therefore, less power consumption by the data center to run and cool.
- Incentive-compatible resource retrieval; in the sense that the proposed resource retrieval model provides for strategies that serve the best options for both, followers and leaders. Indeed, bounding the price units to predefined thresholds that consider the interests of both parties in the game provides for an equilibrium framework in such a non-cooperative environment.
- Timely resource-retrieval model; which is achieved by motivating the cooperation of both parties, the CDN providers and their VNCs, to offer reasonable price units that encourage retrieval requests with less number of negotiation rounds.
- Behavior-based negotiation model; employs a belief function to record the history of the player in terms of its negotiation behavior in the previous resource retrieval rounds to give an insight into its pricing tendency.

2.2 Paper Organization

The rest of this paper is organized as follows: Section 3 presents related work from the literature. We discuss our proposed model in Section 4, its definitions, and mathematical modeling in Sections 4.1 and 4.3. Section 4.5 presents the benchmark model, and then the numerical results are presented and discussed in Section 5. Finally, Section 6 concludes this paper.

3 Related Work

Resource utilization in cloud-based service networks greatly affects the energy usage of such environments and the computing nodes they run. Green computing is emerging as an important criterion to consider when managing such networks. Indeed, the rising environmental concerns [11], besides the rapid increase in energy costs make such criteria emerging more than ever. This includes the infrastructures of CDNs and what follows of resource allocation processes, monitoring, and any future scaling and expansion plans. Migration of VMs has been proposed in many works in the literature to minimize the need to run new physical machines, which claim to help in reducing energy requirements. This may help, though, the migration process itself may impose other problems that contradict the SLAs [12,13]. Indeed, keeping up the commitments that are guaranteed by the SLAs is quite challenging in such migration scenarios. Not to mention other management objectives such as load balancing [14], network performance, service delays, and others. Optimal scheduling and VM migrations models proved to be NP-hard, though; metaheuristic models could deliver efficient allocation models. The authors of [15] modeled the resource allocation problem using an improved Grey Wolf Optimization chaotic binary model, to allow for load-balanced network fabrics and low traffic volumes. In [16], the authors proposed a metaheuristic workflow-scheduling model for VMs in green cloud computing platforms, with the goals of enhancing throughput while reducing costs and energy consumption [17]. Studies the problem of VM mappings and migration processes over physical machines from the perspective of energy efficiency using a decentralized auction-based management model. Such proposals are great, they present efficient models that ease the VM migration processes and provide for load-balanced networks. However, to allow for real resource utilization and reduce the energy consumption rates, we believe that VM migration might be an option to deploy only when we make sure that the allowed resources at the hosting cloud servers are truly utilized (i.e., fully used by their leasing clients), otherwise, deploying other management options might be more efficient. Those hosting cloud servers where resources are reserved but not efficiently used are considered candidate assets that could be considered to host other new clients before deciding to migrate a VM from one place to another or scale the data center fabric to spawn unnecessary new server racks which would impose both, further costs and energy requirements.

Stackelberg sequential games have been used in similar problems in the literature to draw a kind of interaction strategy between the providers and their clients. In [18], the authors proposed an energy-aware resource allocation model for virtual resource management in cellular networks, in their work; the allocation model is built as a Stackelberg game but with the assumption of ideal network link capacities. Moreover, the authors of [19] proposed deploying a Stackelberg-based model for load balancing compared to a benchmark model with random and Flow Shop scheduling algorithms. Their delivered results show better resource utilization and an enhanced throughput, with less number of errors and failures. In this work, the objective is extended to allow for efficient and true utilization of cloud resources in a way to deliver data center networks with less power consumption rates. Modeling the resource negotiation problems through Stackelberg and other Game Theoretic strategies has also been proposed in other works in the literature [20–22]. In such decision-making problems, strategies of game theory and mechanism design may help in building systematic allocation models. In this context of resource management and allocation in CDNs, there are several research proposals in the literature that we studied and reviewed. The goal of such proposals is mainly around enhancing the resource allocation models while maintaining the clients' satisfaction. The authors of [23] proposed a finite extensive form game in a backward induction strategy for better resource allocation and utilization rates compared to a benchmark model with a first fit allocation algorithm. In [24], the authors proposed deploying a repetitive non-cooperative game for a resource allocation model with partial information. Their model may allow for fast run times, lower violation rates for the SLAs, and higher utilities to the providers. However, their pricing strategy was static rather than dynamic which does not allow for considering the real-time updates in the network and the power grids' loads. Another work tackled the problem as a non-cooperative game to reach a Nash equilibrium in computational grids through deploying a proportional scheme algorithm, the model of [25] provided for a kind of load-balanced network. However, the balancing tasks are allocated according to the computing capacity of the hosting machines with no regard to the true utilization and power consumption rates.

The use of auctions to set dynamic price units for resources is not new, compared to those traditional inflexible pricing mechanisms, such dynamic mechanisms allow for competitive price negotiation environments that could be utilized to manage the resource allocation process. In this context, the authors of [26] proposed Bazaar-Extension, a negotiation model built on the CloudSim framework to allow for direct cost negotiations between service providers and consumers. In the Haizea resource management framework, the authors of [27] developed another model that tackles the problem of dynamic monitoring for the leased VMs' resources, to consider the utilization rates and the dynamic changes in clients' resource requirements. In terms of resource management, authors [28] considered allowing cloud service providers to create a kind of cloud federation to pool their unused resources. These resources are then allocated to the spot market clients. While this model may help attain better resource utilization rates and provide revenue maximization for the federation, individual service providers may find no incentive to participate in the federation's pool of resources. Instead, they may choose to lease their unused resources in their spot markets.

The authors of [29] studied the problem of resource allocation in cloud networks through a model that combines genetic algorithms with state-action reward and state-action learning. In their model, they investigated how to maximize the resource usage rates by choosing the appropriate set of activities. The results of their proposal show satisfying resource utilization and load balancing rates. Few other research proposals for resource management in data center networks suggest a third-party cloud broker to negotiate [30] or collect the resource availability readings from several CDN providers and match it with the received clients' leasing requests. Such resource leasing or service provision services are covered by SLAs that go through a management cycle that starts with the SLA creation state which includes the discovery of the service provider, and may later evolve and adapt according to the changing requirements of both parties after the first agreement [31]. Again, this helps in mapping the received resource requests in a way that could be considered satisfying for both: providers and clients, but not achieving the goal of efficient resource utilization and reduced energy requirements.

Different from the other models in the literature, our proposed model AMAD allows for a novel negotiation model for resource retrieval between the cloud network providers and their clients. To the best of our knowledge, our AMAD model is the first to employ a dynamic negotiation framework that considers the electrical power consumption rates of the running machines besides their true resource utilization rates. Moreover, AMAD also allows for a novel win-win resource retrieval model that returns the unused-leased resources through a pricing mechanism that satisfies both: providers and their clients. What is more, this proposed model combines the Stakelberg leader-follower strategy with a game-theoretic approach to incentivize timely resource retrievals. In addition, AMAD proposes using a belief function that records the players' retrieval behavior to motivate truthful resource requirements in the end.

4 The AMAD Model

Resource utilization in such cloud-based networks can take place before and after the allocation processes. Before the allocation, called the initial phase, CDN providers can utilize their resources better by setting policies to discourage resource exaggeration and motivate the service clients to reveal their true resource requirements instead. This can be done in many ways through various models. In [6], we proposed a model that adopts the Vickrey Clarke Groves pricing mechanism to reduce such exaggeration actions. Having the resources being allocated, the providers may still monitor the resource usage of the leasing clients and accordingly find a way to utilize unused resources even if their clients already lease them. However, once the allocation is set, call it the second phase, any amendments to the allocation decisions need to respect the contracted QoS levels and its associated SLAs. Hence, the utilization processes at this phase might differ compared to those of the first phase, as the leasing clients need to approve any amendments. To do so, in this work, we propose AMAD, a negotiation model between the providers and their clients to formulate a way to utilize the resources better while keeping both parties satisfied.

4.1 Model Definitions

A CDN hosts several server machines with massive processing and computing power and these server machines are interconnected through a set of network links. Therefore, in our model, using the graph theory, we define the data center as a graph D that consists of a set of nodes S and edges L. The server machines and what they allow for processing power and storage spaces are represented by the graph nodes S. The communications between such nodes are carried out through the set of edges L that represent the CDN's links, which provide the required bandwidth capacities to interconnect the data center nodes together, and to other networks. Each data center node s, $s \in S$, is equipped with predefined physical resource capacities of processing power ρ_{s_k} and storage space ζ_{s_k}. The tenant clients of the CDN are denoted as C, where each client c, $c \in C$, is leasing an amount of processing and storage resources given by ρ_{c_i} and ζ_{c_i}, respectively. For each client c_i, the leased node resources of processing and storage come with the bandwidth resources l_{s_k} required to get it interconnected with other nodes and networks. Therefore, in this work, we assume that resource retrieval at the nodes' side would implicitly include the bandwidth resources required to reach the tackled nodes' resources.

In AMAD, the resource retrieval process is modeled as a negotiation game between the CDN providers and their tenant clients. Hence, resources are first allocated to the CDN clients after setting the resource amount to be reserved, price units, and the corresponding lease times. Next, any later resource reservation adjustments are done via a *repeated leader-follower negotiation game* framework, through which, the CDN provider (i.e., the leader) initiates the resource negotiation game with its clients (i.e., the followers) to retrieve those leased unused resources according to predefined policies. In such a game, both players, the provider and the clients are assumed to be *rational*, and so they both aim to maximize their *utilities*. Therefore, each leasing client c_i aims to maximize its utility function U_{c_i} defined in (1). This can be achieved by maximizing the difference between the gain $\Im_{c_i}$ it can attain from the services being created using the leased resources, and the aggregate cost paid for the reserved resources calculated according to the reserved set of resource $s\left(\rho_{c_i}.\zeta_{c_i}\right)$ at the price-unit τ_{c_i} while considering the electricity tariff-unit e_{c_i}.

$$U_{c_i} = max\{\Im_{c_i} - (\rho_{c_i} + \zeta_{c_i})(\tau_{c_i} + e_{c_i})\} \tag{1}$$

Accordingly, to satisfy the objective of (1), a client c_i needs to lease its resources from a CDN service provider that provides reliable services that maximize the gain $\Im_{c_i}$ with the minimum summation of costs that depends on the amounts of resources being reserved at a price unit τ_{c_i} and electricity-unit of e_{c_i}. As for the CDN provider P_D, the utility function is presented in (2).

$$U_p = max \sum\nolimits_{\forall c_i.\, c\in C} (\rho^u_{c_i} + \zeta^u_{c_i}) \tag{2}$$

To preserve the resource utilization objective, providers' utility U_p is built in a way that only considers the utilized resources $(\rho^u_{c_i}.\ \zeta^u_{c_i})$ among those allocated to the leasing set of clients C. Hence, for a CDN provider P_D, it's the truly utilized resources that are counted in the utility function, and therefore, maximizing its utility would come through maximizing the utilization rates of its pool of resources.

4.2 The Resource-Retrieval Mechanism

Having the processing and storage resources being allocated to the leasing set of clients C, $\rho^a_{c_i}$ and $\zeta^a_{c_i}$, as demonstrated in Fig. 1, the provider needs to keep monitoring the real utilization rates of the allocated processing and storage resources to each client c, $\rho^u_{c_i}$ and $\zeta^u_{c_i}$, respectively. For any new resource allocation requests received by the provider P_D from new clients, it checks for the available but unleased resources $\rho^v_{s_k}$ and $\zeta^v_{s_k}$ at its premises first, if it suffices, the new allocations may proceed without the need for any retrieval processes. Otherwise, it runs the resource-retrieval mechanism to help attain what is required from those allocated unused extra resources, $\rho^r_{c_i}$ and $\zeta^r_{c_i}$, that are already leased to the existing leasing clients.

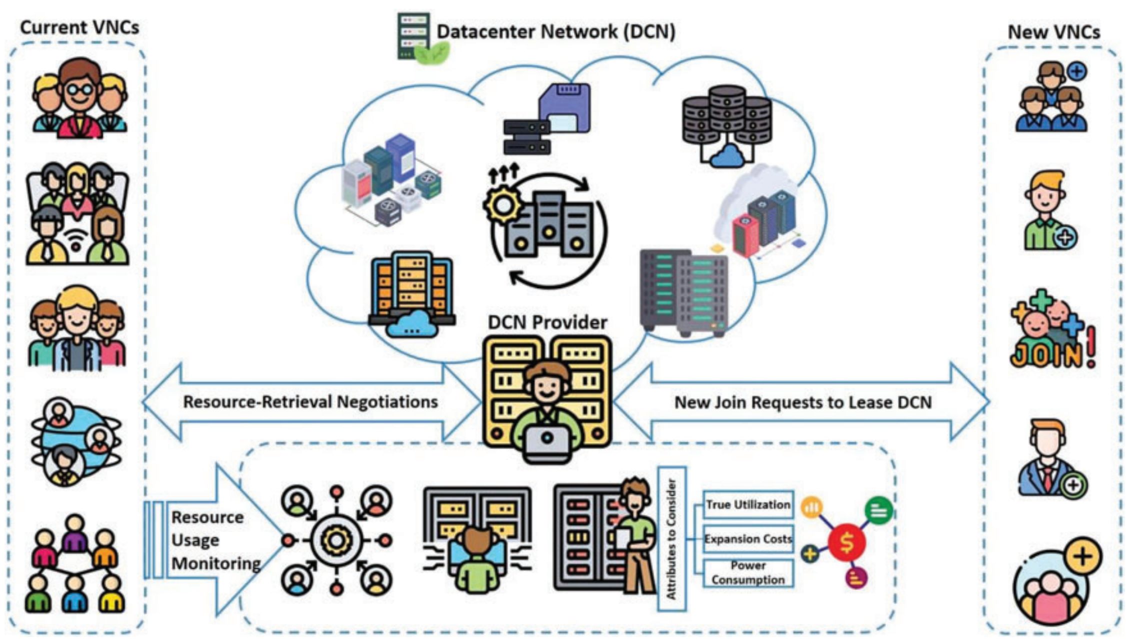

Figure 1: AMAD's model demonstration

For each physical node s, $s \in S$, of the CDN, the provider P_D uses the formulas defined in Eqs. (3) and (4) to keep its pool's resource availability information up to date.

$$\rho^{v}_{s_k} = \left(\sum\nolimits_{\forall s \in S} \rho_{s_k} - \sum\nolimits_{\forall c \in C} \rho_{c_i}\right) \tag{3}$$

$$\zeta^{v}_{s_k} = \left(\sum\nolimits_{\forall s \in S} \zeta_{s_k} - \sum\nolimits_{\forall c \in C} \zeta_{c_i}\right) \tag{4}$$

Yet, for the true resource availability, $\rho^{r}_{c_i}$ and $\zeta^{r}_{c_i}$, those rates that reflect the real usage among the allocated resources at the clients' reserve-box side, the model uses Eqs. (5) and (6) to keep such availability rates up to date.

$$\rho^{r}_{c_i} = \left(\sum\nolimits_{c_i} \rho^{a}_{c_i} - \sum\nolimits_{c_i} \rho^{u}_{c_i}\right) \tag{5}$$

$$\zeta^{r}_{c_i} = \left(\sum\nolimits_{c_i} \zeta^{a}_{c_i} - \sum\nolimits_{c_i} \zeta^{u}_{c_i}\right) \tag{6}$$

Accordingly, if the true available resources at the provider's premises, $\rho^{r}_{s_k}$ and $\zeta^{r}_{s_k}$ defined in Eqs. (7) and (8), can't cover the resource requirements of the new leasing requests $\rho^{n}_{s_k}$ and $\zeta^{n}_{s_k}$ of the new clients C_n, then the provider P_D needs to start the resource-retrieval negotiation process to motivate those clients who have extra leased unused resources to return.

$$\rho^{r}_{s_k} = \rho^{v}_{s_k} + \sum\nolimits_{\forall c \in C} \rho^{r}_{c_i} \tag{7}$$

$$\zeta^{r}_{s_k} = \zeta^{v}_{s_k} + \sum\nolimits_{\forall c \in C} \zeta^{r}_{c_i} \tag{8}$$

The utilization ratios, $\rho^{u'}_{c_i}$ *and* $\zeta^{u'}_{c_i}$, for all existing clients in C are calculated as shown in Eqs. (9) and (10) to mark the target clients to be contacted by C_x. So, the clients with utilization ratios below 80% are asked to retrieve part of their extra unused resources while being offered a compensation price-unit $\tau^{r}_{c_i}$. In AMAD, we proposed two different strategies to shape the way of negotiating such retrieval price units.

$$\rho^{u'}_{c_i} = \frac{\sum\nolimits_{c_i} \rho^{a}_{c_i} - \sum\nolimits_{c_i} \rho^{u}_{c_i}}{\sum\nolimits_{c_i} \rho^{a}_{c_i}} \tag{9}$$

$$\zeta^{u'}_{c_i} = \frac{\sum\nolimits_{c_i} \zeta^{a}_{c_i} - \sum\nolimits_{c_i} \zeta^{u}_{c_i}}{\sum\nolimits_{c_i} \zeta^{a}_{c_i}} \tag{10}$$

4.3 Model Constraints

In AMAD, we set the following constraints to keep the resource-retrieval model motivated and maintain the utility objectives of both: providers and clients.

4.3.1 Multiple Participation Is Not Allowed

In each resource-retrieval game, N_y, the price-unit negotiation process may take several rounds. In the expression N_y, N refers to a new negotiation session, with the reference y. A client in C_x has only one chance to participate in the retrieval game, no matter how long the negotiation rounds last. This is to motivate the clients to reveal their intentions earlier, accept the early offered price units, and participate with the extra resources they have. To verify that, in our proposed model AMAD, we introduced a binary variable named $N^{c_i}_{y}$ for each candidate client c_i at the resource retrieval game N_y, such that:

$$N_y^{c_i} = \begin{cases} 1 & if\ client\ c_i\ returned\ resources\ at\ any\ round\ r\ of\ game\ N_y \\ 0 & otherwise \end{cases} \tag{11}$$

This is restricted by the following condition to guarantee only one participation per client c_i, so multiple participations are not allowed:

$$\sum\nolimits_{\forall r \in N^y} N_y^{c_i} \leq 1 \quad ; N_y^{c_i} \in [0.1] \tag{12}$$

4.3.2 Offered Price-Units Threshold

To motivate leasing reasonable resource amounts from the beginning, and reduce the motivations of requirement exaggeration, in AMAD, we constrained the provider's offered price-units for resource retrieval with a threshold point that considers the original leasing price-units τ_{c_i}, and the leasing price-unit for the new clients τ_{c_n} as well. Accordingly, the offered price-unit τ_x^* is bounded as follows:

$$\tau_x^* \leq \tau_{c_i} \quad ; \forall \left(\rho_{c_i}^a . \zeta_{c_i}^a\right) . c_i \in C \tag{13}$$

$$\tau_x^* \leq \tau_{cn} \quad ; \forall \left(\rho_{cn}^n . \zeta_{cn}^n\right) . c_n \in C_n \tag{14}$$

4.4 AMAD Strategies

In AMAD, we proposed two different strategies to shape the way of negotiating such retrieval price units. The main goal of both strategies is to motivate the leasing clients' cooperation with their providers to help them utilize their network resources better. In parallel, such strategies need to be incentive-compatible for both parties, clients and providers. Hence, besides the resource utilization objective, the return price units need also to be satisfying while being constrained to revenue and time objectives. As mentioned in the previous section, the process of choosing those clients that are targeted by the negotiation process C_x is *dynamic* and dependent on their utilization ratios $\rho_{c_i}^u$ and $\zeta_{c_i}^u$. Hence, those with extra resources being leased but not used at their reserve box and who satisfy the 80% condition at the current round are contacted, others are not.

4.4.1 Further-Highest Price Strategy (FHPS)

The negotiation game starts with the CDN provider P_D (i.e., player 1) initiating the resource retrieval call $N_y . y \in Y$, for all those clients in C_x (i.e., player 2) satisfying the utilization ratio threshold at time t. In-game N, each player in C_x has only one chance to participate in the retrieval process, and therefore, once it finds an acceptable return price-unit being offered by P_D, it announces its acceptance with the number of resource units to release. A breakdown of the FHPS model is presented in Algorithm 1, accordingly, the negotiation rounds continue as long as the required amounts of resources, $\rho_{s_k}^n$ *and* $\zeta_{s_k}^n$, are not yet retrieved. Once retrieved, to set the final price-unit to be paid for the retrieving clients, the provider proceeds with one further negotiation round N_{y+1} pretending a need for more resources and records the next acceptable price-unit $\tau_{c_x}^*$. This price-unit, which we call the *further-highest*, is now paid for all those clients who accepted the previous calls, but not the current clients which will be excluded.

Algorithm 1: The FHPS algorithm: further-highest price strategy for resource retrieval	
1: ***input:***	*At time t, CDN provider P_D checks the status of its data center nodes S, through which, it:*
2:	**reads:** (1) The real-time processing power availability at its nodes $\rho_s{}^v{}_k$;
3:	(2) The real-time storage space availability at its nodes $\zeta_s{}^v{}_k$;

(Continued)

Algorithm 1 (continued)

4: (3) The resource requirements of the new clients' requests C_n to be hosted, $\rho_s{}^n{}_k$ and $\zeta_s{}^n{}_k$;
5: (4) The real-time utilization rates of the leased resource, $\rho_c{}^u{}_i$ and $\zeta_c{}^u{}_i$, by each client c_i.
6: **updates** C_x: it appends each leasing client c_i with either $\rho_c{}^u{}_i$ or $\zeta^u_{ci} \leq 80\%$ to list C_x;
7: **if** the CDN's node's availability of $\rho^v_{sk} \geq \rho_s{}^n{}_k$ and $\zeta^v_{sk} \geq \zeta_s{}^n{}_k$, then:
8: **allocate** $\rho_c{}^n{}_n$ and $\zeta_c{}^n{}_n$ to each new client c_n of the set of clients C_n.
9: **else;**
10: **find:** (1) the required processing power $\rho_s{}^d{}_k = \rho_s{}^n{}_k - \rho_s{}^v{}_k$ to retrieve from C_x;
11: (2) the required storage space $\zeta_s{}^d k = \zeta_s{}^n{}_k - \zeta_s{}^v{}_k$ to retrieve from C_x;
12: **start** the first round r the resource-retrieval game N_y with the clients in C_x, so it:
13: (1) sends the retrieval request to all clients in C_x, with $N_y{}^{ci} = 0$;
14: (2) reveals the offered price unit τN∗y to compensate resource-unit retrievals;
15: (3) collects the retrieved resources from the retrieving clients $c_i \in C_x$;
16: **update** $\rho_s{}^v{}_k$, $\zeta_s{}^v{}_k$, and C_x;
17: **while** $\rho_s{}^v{}_k \leq \rho_s{}^d{}_k$ or $\zeta_s{}^v{}_k \leq \zeta_s{}^d{}_k$, and C_x /= φ, do;
18: (1) resend a new resource-retrieval request again to the updated C_x;
19: (2) start a new negotiation round with a higher price unit being offered $\tau_{N^*{}_{y++}}$;
20: (3) wait for new retrievals;
21: (4) return to line 16;
22: **else;**
23: (1) resend one further resource-retrieval request again;
24: (2) start a new negotiation round with a new price unit being offered $\tau_{N^*{}_{y++}}$;
25: (3) if the offered price unit is accepted, skip to line 27;
26: else, return to 24;
27: (4) set the same compensation price unit for all as $\tau_c{}^*{}_x$;
28: **allocate** $\rho_c{}^a{}_n$ and $\zeta_c{}^a{}_n$ to each new client c_n of the set of clients C_n;
29: **compensate** the retrieving clients of C_x with price unit $= \tau^*_{cx}$;
30: ***output:*** *updated readings of $\rho_s{}^u k$, $\zeta_s{}^u k$ and C_x;*

With FHPS, even non-cooperative clients are expected to show more *cooperative* behavior with the CDN provider at the early stages of the negotiation rounds. Indeed, to avoid being excluded from the retrieval game, they will tend to accept early provider offers. Such clients would be motivated by expecting higher price units to be paid by the provider if they accept early offered ones. This would promote their willingness to accept lower price units (i.e., early offers), and accordingly, retrieve the required resources with less number of negotiation rounds and price units.

4.4.2 Belief Function-Based Strategy (BFBS)

To motivate timely negotiation rounds, in AMAD, we proposed building a kind of behavior-based score for each client participating in the resource-retrieval games, through which, the provider may adapt its negotiation policy with each client according to its negotiation history in the previous retrieval games. Hence, those clients who have a repetitive tendency to stall their retrievals, and wait for higher price units are flagged with a score that reflects such behavior. Accordingly, in BFBS, the model records the behavioral history of each client calculated as presented in Eq. (15) which considers the number of accepted offers $\mathfrak{R}^A$ to the total number of negotiation rounds the client participated in, $\mathfrak{R}^T$.

$$\beta_{c_i} = \frac{\mathfrak{R}^A}{\mathfrak{R}^T}. \quad ; \beta_{c_i} \in [0.1] \tag{15}$$

Referenced to the value of β_{c_i}, the CDN provider may motivate the cooperation of its clients by adapting their offered price units as presented in Eq. (16). Hence, the client players will receive different price-unit offers according to their previous history of negotiation. Those who showed cooperative behavior (i.e., accepting offers in relatively short negotiation rounds) are rewarded by receiving new offers with higher price units $\tau_{N_y}^{\triangleright}$, compared to others who showed non-cooperative behavior with a lengthy number of negotiation rounds. Those non-cooperative clients would only receive a relatively small increase in the offered price units depending on their belief scores given by β_{c_i}.

$$\tau_{N_y}^{c_i} = \begin{cases} \tau_{N_y}^{\triangleright} & if\ c_i{'}s\ belief\ score\ \beta_{c_i} \geq 90\% \\ \tau_{N_y}^{\triangleright} . {\beta_{c_i}}^2 & otherwise \end{cases} \tag{16}$$

Based on that, and according to the belief score each client accumulates as defined in (15), the model calculates the price units to offer. Being a value that is dependent on the clients' retrieval history, each client would expect different compensation price units that are higher for cooperative clients (i.e., who showed a tendency to accept early retrieval offers) and lower for those with less cooperation history (i.e., those who showed a tendency to wait for further negotiation rounds). In this way, the model pushes the behavior of the clients to a better level of cooperation.

Indeed, having the return price-unit $\tau_{N_y}^{c_i}$ linked to the client's belief value β_{c_i} would incentivize early retrievals. True, as those who accept early retrieval offers would accumulate higher belief values which would result in higher compensation offers. The integration of the retrieval process with such belief-dependent compensation price units allows for a win-win retrieval model that satisfies its resource utilization attempts in timely negotiation rounds.

Therefore, as shown in Eq. (16), clients with a belief score that is 90% or more are rewarded by receiving new offered price units (full value of $\tau_{N_y}^{\triangleright}$) that are higher if compared to what is offered to others with lower belief scores.The compensation price-unit for any client with a β_{c_i} value that is less than 90% is calculated in a direct relationship with the accumulative belief values they built in the previous negotiation games. A detailed representation of the BFBS is presented in Algorithm 2.

Algorithm 2: The BFBS algorithm: belief function-based strategy for resource retrieval

input: *at each resource allocation time t, if the P_D node's availability $\rho_s{}^v k < \rho_s{}^n k$ or $\zeta_s{}^v k < \zeta_s{}^n k$:*
    find: (1) the required processing power $\rho_s{}^d{}_k = \rho_s{}^n{}_k$ - $\rho_s{}^v{}_k$ to retrieve from C_x;
        (2) the required storage space $\zeta_s{}^d k = \zeta_s{}^n{}_k$ - $\zeta_s{}^v{}_k$ to retrieve from C_x;
        (3) the accumulative belief score β_{ci}, for each c_i *in* C_x;
    start the resource-retrieval game N_y with the first negotiation round r by:
        (1) sending resource retrieval requests to all clients in C_x with $N_y{}^{ci} = 0$;
        (2) announcing the offered price-unit $\tau_{N^*{}_y}$ to compensate resource-unit retrievals;
        (3) collecting the returned resources from the retrieving clients $c_i \in C_x$;
    update the availability records ($\rho_s{}^v{}_k$, $\zeta_s{}^v{}_k$), and the set C_x;
        while $\rho_s{}^v{}_k$ is still $< \rho_s{}^d{}_k$ or $\zeta_s{}^v{}_k$ is still $< \zeta_s{}^d{}_k$, and $C_x =/ \varphi$, **do**;
            (1) resend a new resource-retrieval request again to the updated C_x;
            (2) for 16 each $c_i \in$ C_x, based on the belief dependent price unit formula;

(Continued)

Algorithm 2 (continued)

13: (1) offer a new, belief-dependent, price-unit $\tau_N{}^{ci}y$;
14: (2) wait for return decision;
15: (3) update the belief-score β_{ci};
16: (3) return to line 10;
17: **else;**
18: **allocate** $\rho_c{}^a{}_n$ and $\zeta_c{}^a{}_n$ to each new client c_n of the set of clients C_n;
19: **compensate** each retrieving client a "β_{ci} dependent" price-unit $= \tau_N{}^{ci}y$;
20: ***output:*** *updated readings of* $\rho_s{}^u k$, $\zeta_s{}^u k$, *and* β_{ci};

4.5 Benchmark Model

Resource utilization strategies that deploy such negotiation methodologies are many, though, each has its objectives with different outcomes to deliver. As a benchmark strategy, we chose to consider a model that shares part of the objectives our model has. In [9], the authors proposed a model that emulates the behavior of a CDN provider that aims to retrieve the unused leased resources while satisfying the client's preferences and time concerns, and not being sensitive to the risk of its utility. Consequently, for those leased unused resources, the resource-retrieval process starts by offering a return price-unit that is lower than that paid by the clients at the time they leased their resources. Clients who seek higher-priced units have the option to wait for better offers in the coming negotiation rounds. With no implicit threats of being deprived of participating in the resource return game, the leasing clients will not be motivated to return their unused resources by low-price units as long as they can attain higher ones with a kind of tolerance waiting for the next negotiation rounds. Indeed, such a strategy helps the clients collect higher utilities, while it is less for their providers. What is more, with the absence of motivations for early returns, such negotiation rounds might be lengthy, taking respectively long time to finish.

As long as the negotiation strategy allows return price units that may equal those paid by the leasing clients at the time they leased their resources, they will never be motivated to return their unused resources earlier, instead, they will always tend to wait for higher offers to be released. An efficient resource utilization model needs to suppress any motivation for resource exaggeration. However, such a strategy can not help suppress clients' exaggeration, indeed, knowing that any extra unused resources could be returned with the same price-units paid at the lease time holds no risk for the client. Instead, they would always keep asking for more resources to reserve for any unexpected needs or requirements, and then simply return it once found unnecessary at the same price-unit.

5 Simulation and Numerical Results

To assess the proposed resource retrieval model, in this section, we are presenting part of the simulation results obtained from the test-bench environment being developed. In which, we assumed have five CDNs, each coming with different CPU and storage resource availability profiles. Based on the dynamic resource availability readings, a CDN can serve both current and new VNCs. In the simulated experiments, for each CDN, we assumed having 30 different VNC profiles with varying resource requirements asking for resources to lease from the assigned CDNs' providers. Not only in their resource requirements, have VNCs also varied in the time their leasing requests are submitted to the CDNs' providers.

In Fig. 2, the model records its resource availability in the providers' pool of resources. The model uses a monitoring module that dynamically reads the resource utilization rates of CPU resources being leased to the current VNCs. With the real utilization rates of these leased resources in the current VNCs, the CDN providers will attempt to fulfill new VNCs' leasing requests using the proposed AMAD resource retrieval models. They will then compare the results with those of the benchmark model, as shown in Fig. 3.

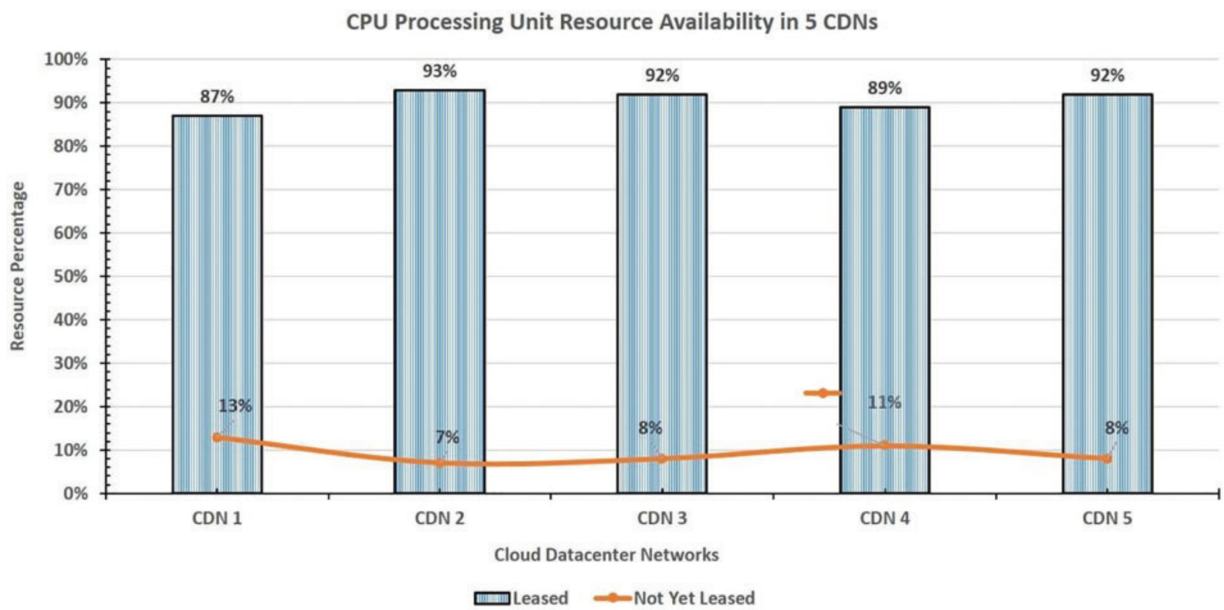

Figure 2: CPU resource availability of 5 CDNs after leasing its resources to VNCs

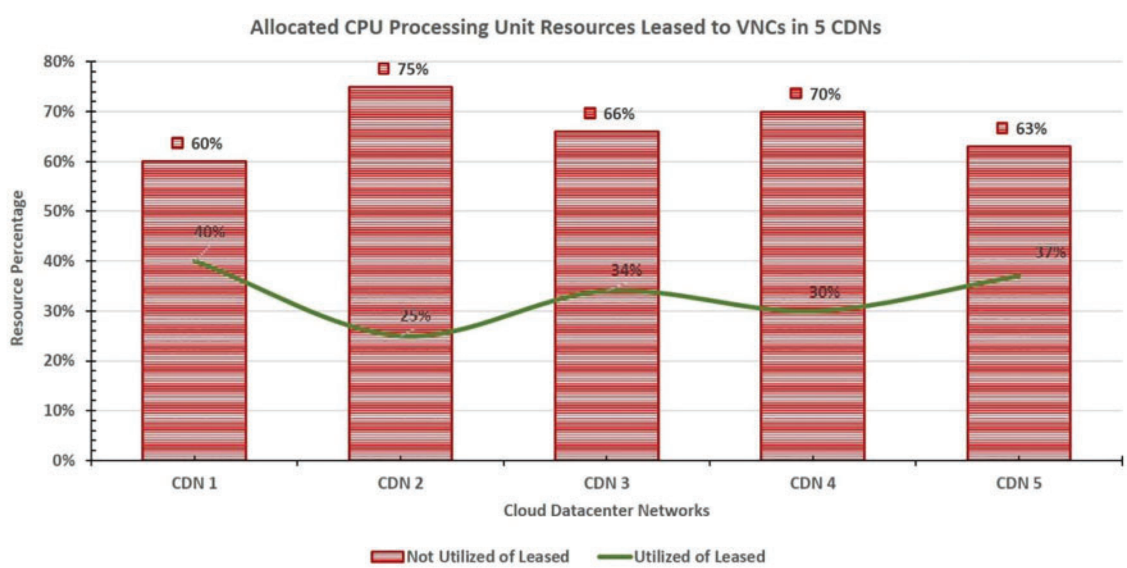

Figure 3: Utilization readings of the CPU resource units leased to 5 VNCs

As for the storage resource units, Figs. 4 and 5 show the reading of both resource availability and the utilization rates for the VNCs' leased resources.

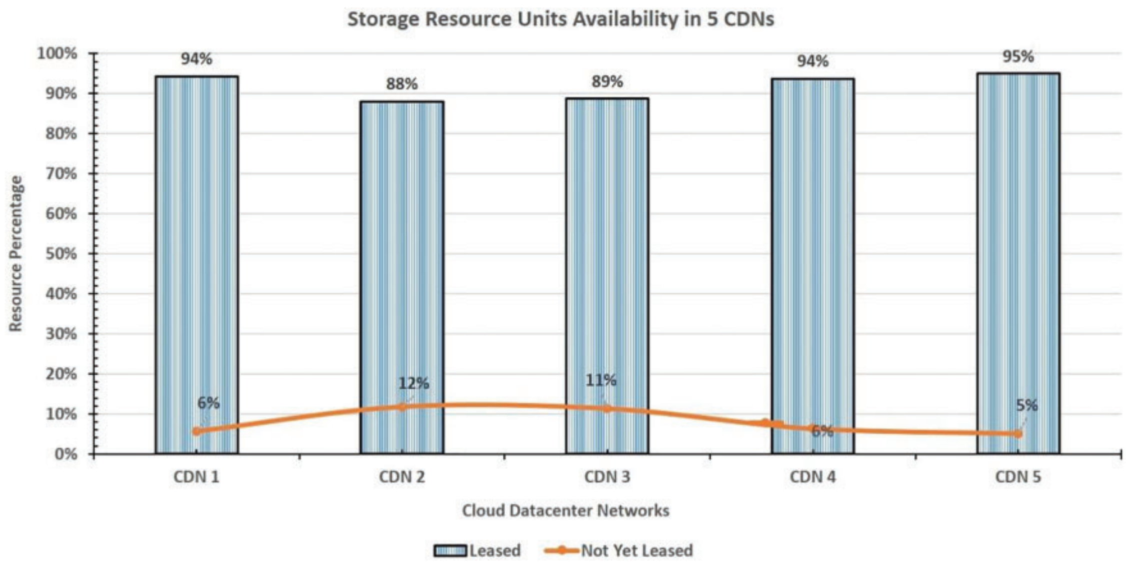

Figure 4: Five CDNs storage resource availability after leasing resources to 5 VNCs

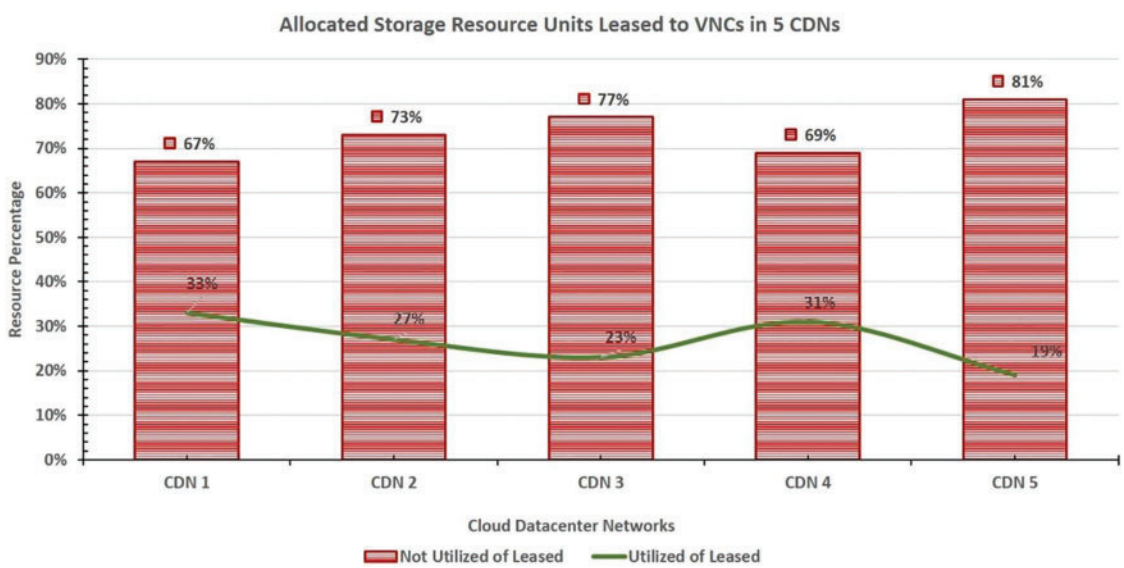

Figure 5: Utilization readings of the storage resource units leased to five VNCs

For CDN providers, the price-unit of the leased resources changes concerning both: the data center network structure, and the corresponding electrical energy consumption rates. Therefore, with more resources being required, the price-unit is expected to rise as this may require upgrades to the current data center network's structure to allow for higher resource capacities. So do the energy consumption rates, it increases in a direct relationship with network and server devices to run. Table 1 lists the resource price units being followed in the experiments we run in this work.

Table 1: CPU and storage resource price-units according to capacity being required at the CDN

Resource units leased to VNCs	0–250	251–500	501–625	626–750	751–875	876–100	1001–1125
CPU unit	\$0.8	\$0.8	\$1.0	\$1.2	\$1.4	\$1.6	\$2.0
Storage unit	\$0.4	\$0.4	\$0.5	\$0.6	\$0.7	\$0.8	\$1.0
Electricity unit	\$[0.0–0.25]	\$[0.251–0. 50]	\$[0.51–0.625]	\$[0.626–0.75]	\$[0.76–0.875]	\$[0.876–1.00]	\$[1.01–1.125]

If we consider the allocation map of the CPU resources of the first CDN at 8 different consecutive time units, (t_0, t_1, t_2,...,t_7), the values presented in Table 2 could be considered as an example of the resource readings (leased, utilized, unused) and their corresponding price-units.

Table 2: Accumulative CPU resource units being leased at the first CDN over time-slots t_0 to t_7

Reading time slot	t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7
Accumulative leased resources $\rho^a_{s_k}$	150 units	275 units	435 units	500 units	610 units	750 units	875 units	1020 units
Resource price-unit τ_{c_i}	\$0.8	\$0.8	\$0.8	\$0.8	\$1.0	\$1.2	\$1.4	\$1.6
Electricity price-unit e_{c_i}	\$0.15	\$0.275	\$0.435	\$0.50	\$0.61	\$0.75	\$0.875	\$1.02
Total price-unit	\$0.95	\$1.075	\$1.235	\$1.30	\$1.610	\$1.950	\$2.275	\$2.62
Accumulative utilized leased resources $\rho^u_{c_i}$	60 units	110 units	174 units	200 units	244 units	300 units	350 units	408 units
Potential resource price-unit $\tau_{c_i}*$	\$0.8	\$0.8	\$0.8	\$0.8	\$0.8	\$0.8	\$0.8	\$0.8
Potential electricity price-unit $e_{c_i}*$	\$0.06	\$0.11	\$0.174	\$0.20	\$0.244	\$0.30	\$0.35	\$0.408
Potential total price-unit $*$	\$0.86	\$0.91	\$0.974	\$1.0	\$1.044	\$1.10	\$1.15	\$1.208

Note: $*$ These values show what the CDN provider could offer if the resource leasing behavior considers the true resource requirements from the beginning.

Before serving any new leasing request from those requests received at the time slot t, and before running any of the proposed resource utilization models, CDN providers may check their monitoring records to check the true resource availability and its corresponding price-unit τ_{ci}. Energy-wise, referring to the readings in the table, we can notice the relation between price-unit hikes and the energy consumption rates. Compared to the reading in rows 2 to 4 of Table 2, the reading of rows 6 to 8 shows the price units a CDN provider could offer if the VNCs' resource leasing behavior considers the true resource needs from the beginning. Analyzing the readings reveals an average reduction of 40% in the electricity price units. This would be an optimal allocation scenario if prevailed. However, part of the VNCs would still tend to exaggerate their true requirements, and consequently, we propose the models of FHPS and BFBS to tackle such issues and motivate a resource retrieval framework that could serve all, resource tenants and providers.

5.1 VNCs' Price-Unit

For VNCs, maximizing the utility function defined in (1) mainly comes through (1) reducing the amounts of resources being leased, and (2) reducing their corresponding resource and electricity price units, τ_{ci} and e_{ci}, respectively. Therefore, compared to the utilization-oblivious and the benchmark resource management models, our proposals in AMAD, the FHPS, and BFBS models, are expected to motivate (1) resource utilization better and accordingly (2) allow for lower price units. Unlike the utilization oblivious model, with the FHPS resource monitoring and retrieval algorithm being deployed, those exaggerating VNCs would have the chance to return their leased but unused resources to their CDN provider's pool of resources. Hence, according to the FHPS algorithm, the compensation price unit for the returned resources is calculated according to the *further highest* price unit being found through the resource-retrieval game. Such a compensation algorithm of FHPS allows for resource return chances, though, its offered return price units are bounded with thresholds that are always less than the original price-unit paid at the leasing process. This allows for better utilization of those leased unused resources that come with price units that motivate truthful resource reveals at the next leasing sessions. In BFBS, the compensation methodology is different; it links the return price units to the negotiation belief records of the negotiated VNCs. Consequently, those VNCs with belief records that are equal to or above 90% would expect to receive higher compensation price units compared to others whose compensations are calculated concerning their belief records (i.e., being < 90%) as presented in (16). BFBS allows for resource return chances that help in better resource utilization, however, it links the return price units to the VNCs historical behavior records being developed which motivates better negotiation behavior by the VNCs.

In the benchmark model, the negotiation process emulates the behavior of CDN providers that have tolerance to long negotiation rounds, which increases the VNCs' utilities but not their ones. Accordingly, it allows longer rounds, and so, the compensation price units are defined in a way that does not motivate any better behavior in the coming resource leasing requests. With more negotiation rounds, the offered return price units could reach those paid earlier in the initial leasing process. Indeed, with no threat of price unit losses, a VNC would never be motivated to lease its true resource requirements! What is left unused could be returned with the same price units. On the contrary, such a policy would motivate resource exaggeration instead, absolutely, as long as a VNC has the chance to return what it finds extra at the same leasing prices, it would always ask for more resources to guarantee its service requirements first, and then return the extra ones later.

Fig. 6 shows the resultant price units for the utilized resources (i.e., those resources a VNC would expect to pay with the FHPS and BFBS models compared to the utilization-oblivious and benchmark models. With the benchmark model, the CDN provider's return policy sets no limit on the number of negotiation rounds. Hence, for patient VNCs, this allows for return price-units that could be equal to the original leasing price-units, and so for VNCs, this means *no losses* to worry about, leading to VNCs' utilities that are mostly the highest compared to the three other models. When reading the figure, the results indicate that few VNCs (those with IDs: 1, 4, 3, 7, and 11) accepted early offers with return price-units that are less than what they paid before at the lease time, and so, their resultant resource price-units are higher compared to other VNCs with other utilization models in the game. However, as we discussed before, such a policy can never help in motivating truthful leasing requests that contradict the resource utilization goal of the work. However, our proposed FHPS and BFBS models allow for low price-units that are very close to those of the benchmark ones, while maintaining the truthful resource leasing behavior motivated.

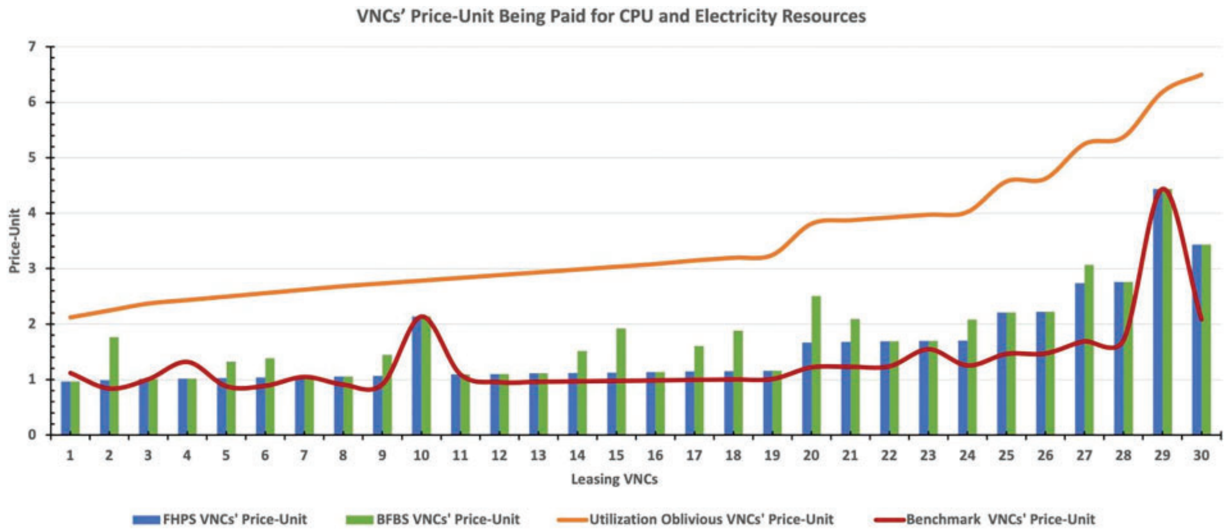

Figure 6: The VNCs price-unit with the different resource allocation models

It is worth highlighting that the resulting price-unit for the VNCs (id: 10 & 29) are the same for the three models of FHPF, BFBS, and the benchmark as those two VNCs show no cooperation with the system and did not return any of their unused resources, and therefore, their price-unit stayed the same. However, their price units are still less than those found with the utilization-oblivious model; as the cooperation of the other VNCs in the system helped in reducing the leasing price units for all from the beginning.

5.2 Number of Negotiation Rounds

CDNs' providers expect to receive new resource allocation requests at any moment, though; such resource requests might be bounded with time-window frames for the resource allocations to be processed. Therefore, relying on such resource-retrieval models to retrieve unused resources in a way to satisfy the new resource requests needs to be timely with time limits to meet, otherwise, such retrieval efforts may lose their intended efficiency. Accordingly, to assess the behavior of the examined models in this context, Fig. 7 shows the negotiation behavior of the proposed models of AMAD, both FHPS and BFBS, compared to that of the benchmark model.

The figure shows how the number of negotiation rounds is affected by the return price-unit policies of the three models. To retrieve 500 CPU resource units, the results show that FHPS was the most motivating model, followed by BFBS, and then lastly, the benchmark model. This can be justified by the pricing model being considered by the FHPS model. Through this, it (1) motivates early resource returns by allowing higher price units according to the next coming negotiation rounds while (2) depriving those VNCs who delayed their acceptance decisions of return offers after collecting the required amounts of resources of the running round (which is kept anonymous from VNCs).

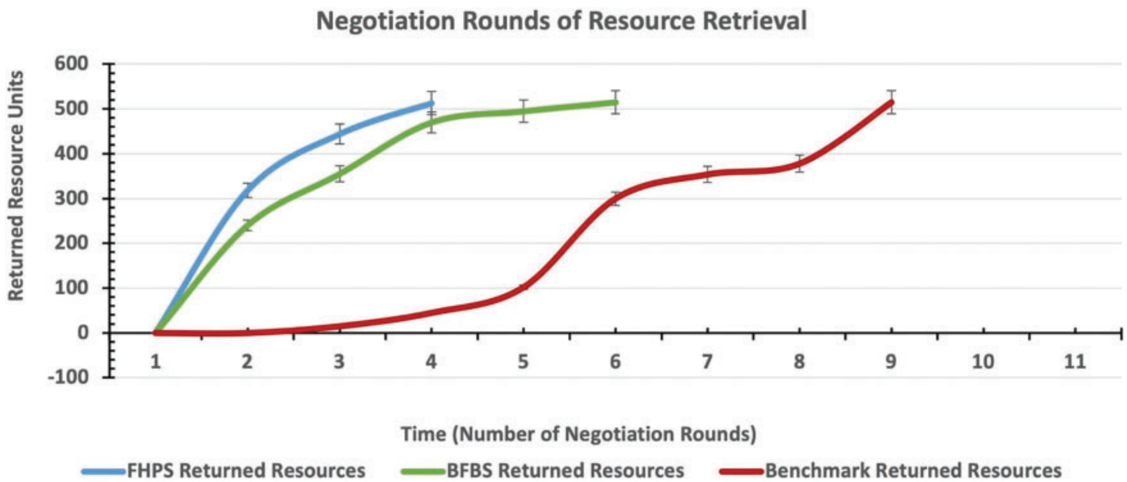

Figure 7: CDN provider's number of negotiation rounds for resource-retrieval

In BFBS, early resource returns are motivated by the belief function being proposed to record the VNC's negotiation behavior, which has a direct impact on the offered price units. So, early returns are incentivized, knowing that the higher belief record (i.e., fast return decisions), a VNC accumulates is the higher price-units it would expect. With the absence of return price-unit limits or cooperation records, according to the benchmark model, a VNC would not be motivated to return its unused resources earlier; on the contrary, it would prefer to defer it for later rounds to achieve higher return units.

5.3 Electricity Price-Unit

In a CDN, the consumed electrical energy is dependent on the number of computing and network machines it runs. Structures of CDNs vary, and so do their energy requirements to run and cool. In AMAD, the goal of resource utilization intersects with the energy consumption concerns. A management model that delivers better resource utilization rates serves all: CDN providers, VNCs, and the environment. Indeed, besides serving the utility objectives of both providers and their clients, it allows for green computing zones that run with lower energy rates. Moreover, a management model that motivates truthful reveals about the true resource requirements and encourages better utilization of the cloud network resources would certainly help the cloud service providers avoid those early updates on their data center structures. Traditionally, a big part of the data center machines might be running consuming electrical energy but not truly utilized. True, those resources are reserved by VNCs but not used. Not only a waste of resources but also the problem may extend when the cloud providers go with expansion plans that are built according to false usage requirements. This does not only mean new assets and structural upgrades costs to be paid, but more electrical consumption to consider which means higher energy bills and loads to the power grids. In this context, the results in Fig. 8 show how AMAD proposed models of FHPS and BFBS could help in reducing the electrical energy requirements of such data center networks when compared to other resource allocation and management models. The results are presented in terms of the electrical price units paid by the clients for their leased resources (calculated as a ratio of the accumulative CDN resources being leased at any moment), which reflects how AMAD models allow for lower price units indicating the utilization rates it achieved through the motivation policies being employed. Besides, it allows for a higher level of stability to the power grids, which otherwise might be compromised affecting other consumers in their homes and business sectors.

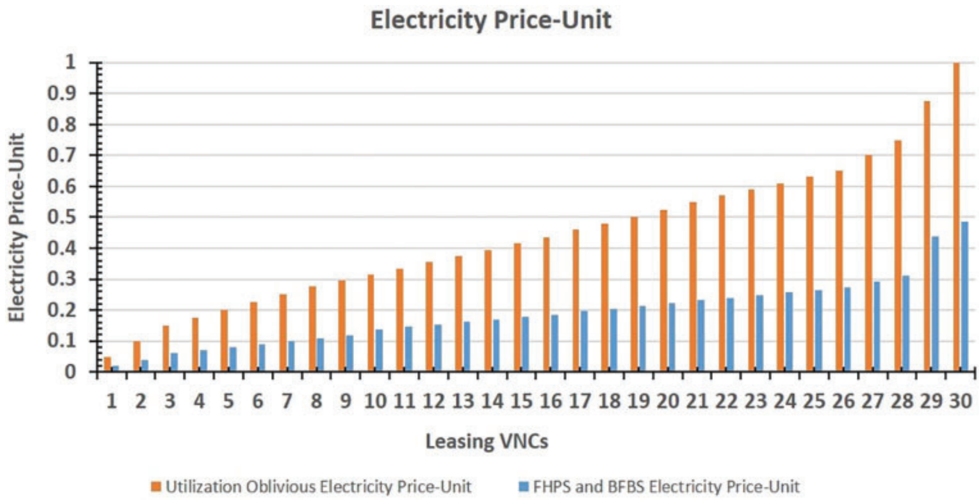

Figure 8: Price-units for the electrical energy being consumed by the datacenter network

5.4 CDN Providers' Utility

For providers, the utility function presented in (2) considers the utilities collected from the resources being utilized, those used by the leasing VNCs, and through the resource-retrieval models being deployed. Comparing the accumulative utilities of the studied models of FHPS, BFBS, and the benchmark one came with the results depicted in Figs. 9 and 10. The results show that both FHPS and BFBS models returned close accumulative utilities, though this is done faster with the FHPS (less number of negotiation rounds) as presented in Fig. 9. On the contrary, the accumulated utilities are 54% less for the benchmark model with almost twice the number of negotiation rounds compared to FHPS. This is due to the motivated return policy being deployed in both FHPS and the BFBS that encourages the VNCs' truthful reveals for their true resource requirements at the initial of the resource allocation stages. Even if extra or unused resources exist, the proposed models motivate early resource returns that are rewarded by competing return price units.

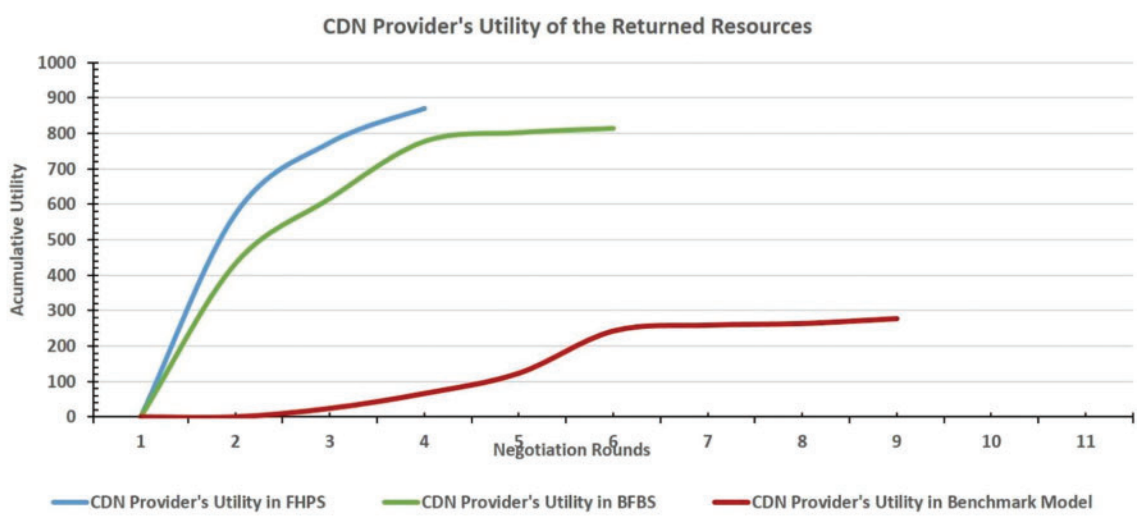

Figure 9: Providers' utilities from the retrieved resources

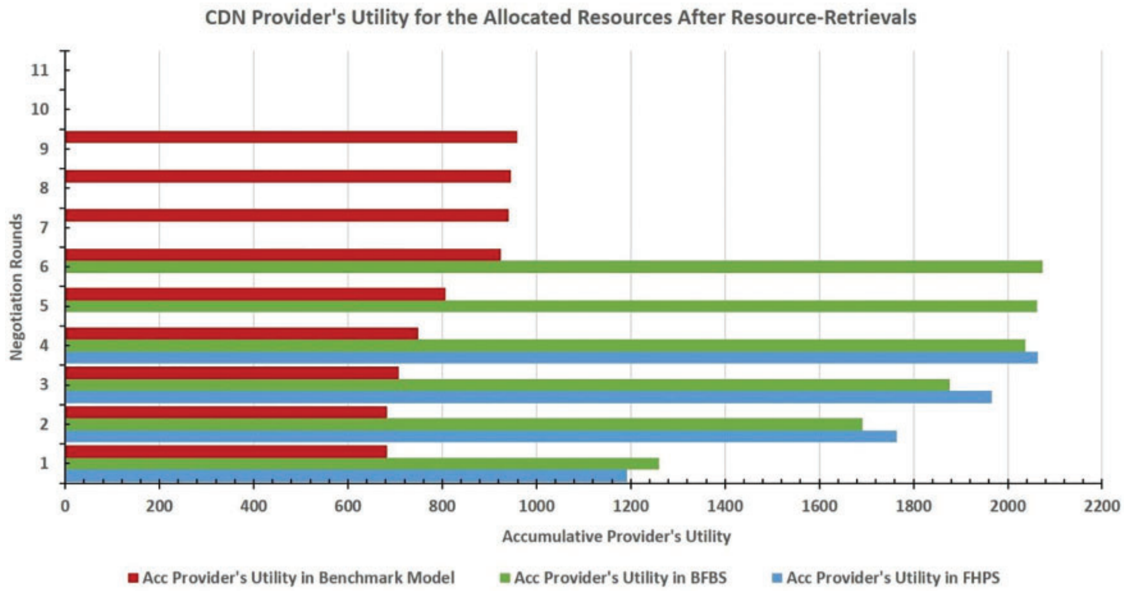

Figure 10: CDN provider's utility from the whole resources being allocated

6 Conclusions

CDNs have emerged as a promising alternative to those physical network platforms, through which, service providers offer cost-effective themes of services that host almost all technical aspects that we use every day. Resources of these cloud-based networks are offered to the clients in the form of services that are paid based on predefined tariffs. Such tariffs are defined in advance according to the amount of resources being reserved based on the client's resource requirements. Such CDNs are equipped with massive amounts of resources, though they still have limits. Therefore, managers need to allocate their resources in an efficient and well-engineered manner to keep their networks as healthy and agile as possible. Today, there are many cloud service providers in the market, and so do the services they offer. However, those who offer reliable service forms in competing tariffs compete better and gain a higher market share and reputation. Marketwise, to keep the offered price units competing, CDN providers need to utilize their resources better to avoid unnecessary updates on the structures of their data center networks. Such updates that may require major amendments to the existing parts of the network fabrics could increase the service costs on the providers' side, which makes it hard for them to maintain their competing offered price unit. At the same time, any scale on the network size would come with more power requirements. Indeed, those new servers and what follows of new network and computing devices would consume more electrical energy to run and cool which may also increase the service costs from the providers' perspective. Therefore, this work presents a resource-retrieval model that seeks to retrieve those leased but unused resources from the VNCs. A retrieval model that follows the Stackelberg negotiation strategy of repeated leader-follower game. For such retrieval processes to be timely and to keep the clients' cooperation motivated, both FHPS and BFBS models are proposed. Compared to the benchmark model, results show that both FHPS and BFBS can help in attaining better resource utilization rates, competing price units, higher utilities for the CDN providers, and timely negotiation rounds.

Acknowledgement: The authors would like to acknowledge the resources and research facilities allowed by the departments of Computer Engineering and Information Technology at their institutions. The authors also extend their acknowledgments to the editors and their reviewers for their efforts and valuable feedback and suggestions that contributed to making this work better.

Funding Statement: The Deanship of Scientific Research at Hashemite University partially funds this work. The authors extend their appreciation to the Deanship of Scientific Research at the Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2024-1580-08".

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Ahmad Nahar Quttoum; data collection: Ahmad Nahar Quttoum, and Muteb Alshammari; analysis and interpretation of results: Ahmad Nahar Quttoum, and Muteb Alshammari; draft manuscript preparation: Ahmad Nahar Quttoum. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data presented in this study are available on request from the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest, and the funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] S. Zhang *et al.*, "Toward green metaverse networking: Technologies, advancements, and future directions," *IEEE Network*, vol. 37, no. 5, pp. 223–232, Sep. 2023. doi: 10.1109/MNET.130.2200510.

[2] A. Katal, S. Dahiya, and T. Choudhury, "Energy efficiency in cloud computing data centers: A survey on software technologies," *Cluster Comput.*, vol. 26, no. 3, pp. 1845–1875, 2023. doi: 10.1007/s10586-022-03713-0.

[3] M. Armbrust *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010. doi: 10.1145/1721654.1721672.

[4] T. Miyazawa, M. Jibiki, V. P. Kafle, and H. Harai, "Autonomic resource arbitration and service-continuable network function migration along service function chains," presented at the IEEE/IFIP Netw. Oper. Manage. Symp., Taipei, Taiwan, Apr. 23–27, 2018, pp. 1–9.

[5] J. Simao and L. Veiga, "Partial utility-driven scheduling for flexible SLA and pricing arbitration in clouds," *IEEE Trans. Cloud Comput.*, vol. 4, no. 4, pp. 467–480, 2016. doi: 10.1109/TCC.2014.2372753.

[6] A. N. Quttoum, "AFAM: A fair allocation model for cloud-datacenter networks," in *Proc. of the 2017 Int. Conf. on Cloud and Big Data Comput.*, London, UK, Sep. 17–19, 2017.

[7] W. K. Tan, D. M. Divakaran, and M. Gurusamy, "Uniform price auction for allocation of dynamic cloud bandwidth," in *IEEE Int. Conf. on Commun. (ICC)*, Sydney, NSW, Australia, Jun. 10–14, 2014, pp. 2944–2949.

[8] A. N. Quttoum, A. Alsarhan, and A. Mohammad, "ARAAC: A rational allocation approach in cloud data center networks," *Future Internet*, vol. 9, no. 3, 2017, Art. no. 50. doi: 10.3390/fi9030050.

[9] A. H. A. Muktadir, M. Jibiki, P. Martinez-Julia, and V. P. Kafle, "Repeated leader follower game for managing cloud networks with limited resources," *IEEE Access*, vol. 7, pp. 108174–108188, 2019. doi: 10.1109/ACCESS.2019.2933031.

[10] L. Tao and S. Sethi, "A review of dynamic Stackelberg game models," *Discrete Cont. Dyn. Syst. Ser. B*, vol. 22, pp. 125–159, 2016.

[11] D. Kumar and Z. Raza, "A PSO based VM resource scheduling model for cloud computing," in *IEEE Int. Conf. Comput. Intell. Commun. Technol.*, Ghaziabad, India, 2015. doi: 10.1109/CICT.2015.35.
[12] S. K. Mondal, A. S. Sabyasachi, and J. K. Muppala, "On dependability, cost and security trade-off in cloud data centres," in *IEEE 22nd Pacific Rim Int. Symp. on Dependable Comput. (PRDC)*, Christchurch, New Zealand, Jan. 22–25, 2017, pp. 11–19.
[13] A. S. Sabyasachi and J. K. Muppala, "Cost-effective and energy-aware resource allocation in cloud data centers," *Electronics*, vol. 11, no. 21, 2022, Art. no. 3639. doi: 10.3390/electronics11213639.
[14] M. Xu, W. Tian, and R. Buyya, "A survey on load balancing algorithms for virtual machines placement in cloud computing," *Concurrency and Comput. Pract. Experience*, vol. 29, no. 12, 2017, Art. no. 238. doi: 10.1002/cpe.4123.
[15] A. Mohammadzadeh, M. Masdari, F. S. Gharehchopogh, and A. Jafrian, "Improved chaotic binary grey wolf optimization algorithm for workflow scheduling in green cloud computing," *Springer J. Evol. Intell.*, vol. 14, no. 4, pp. 1997–2025, 2021. doi: 10.1007/s12065-020-00479-5.
[16] A. Mohammadzadeh, M. Masdari, F. S. Gharehchopogh, and A. Jafrian, "A hybrid multi-objective metaheuristic optimization algorithm for scientific workflow scheduling," *Cluster Comput.*, vol. 24, no. 2, pp. 1479–1503, 2021. doi: 10.1007/s10586-020-03205-z.
[17] W. Wang, Y. Jiang, and W. Wu, "Multiagent-based resource allocation for energy minimization in cloud computing systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 2, pp. 205–220, Feb. 2017.
[18] G. Liu, F. R. Yu, H. Ji, and V. C. Leung, "Virtual resource management in green cellular networks with shared full-duplex relaying and wireless virtualization: A game-based approach," *IEEE Trans. Vehicular Technol.*, vol. 65, no. 9, pp. 7529–7542, 2016. doi: 10.1109/TVT.2015.2497360.
[19] R. Swathy, B. Vinayagasundaram, G. Rajesh, A. Nayyar, M. Abouhawwash and M. Abu Elsoud, "Game theoretical approach for load balancing using SGMLB model in cloud environment," *PLoS One*, vol. 15, no. 4, 2020, Art. no. e0231708. doi: 10.1371/journal.pone.0231708.
[20] Z. Ullah *et al.*, "Negotiation-based combinatorial double auction mechanism in cloud computing," *Comput. Mater. Continua*, vol. 69, no. 2, 2021, Art. no. 21232140. doi: 10.32604/cmc.2021.015445.
[21] S. Deochake and D. Mukhopadhyay, "An agent-based cloud service negotiation in hybrid cloud computing," in *ICT Syst. Sustain. Adv. Intell. Syst. Comput.*, Springer, Singapore, vol. 1270, pp. 563–572, 2020. doi: 10.1007/978-981-15-8289-9_55.
[22] A. Kannaki, V. Azhagu, and J. M. Gnanasekar, "A novel multi-agent approach to control service level agreement violations in cloud computing," *TURCOMAT*, vol. 12, no. 12, pp. 1432–1438, 2021.
[23] X. Xu and H. Yu, "A game theory approach to fair and efficient resource allocation in cloud computing," *Math. Problems Eng.*, vol. 14, 2014, Art. no. 915878.
[24] A. Nezarat and G. Dastghaibifard, "Efficient Nash equilibrium resource allocation based on game theory mechanism in cloud computing by using auction," in *Proc. of the 1st Int. Conf. on Next Gener. Comput. Technol. (NGCT)*, Dehradun, India, Sep. 2015, pp. 1–5.
[25] R. Subrata and A. Y. Zomaya, "Game-theoretic approach for load balancing in computational grids," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 1, pp. 66–76, 2008.
[26] B. Pittl, W. Mach, and E. Schikuta, "A negotiation-based resource allocation model in IaaSMarkets," in *IEEE/ACM 8th Int. Conf. on Utility and Cloud Comput. (UCC)*, Limassol, Cyprus, 2015, pp. 55–64.
[27] P. Chokhani and G. Somani, "Dynamic resource allocation using auto-negotiation in Haizea," in *Sixth Int. Conf. on Contemp. Comput. (IC3)*, Noida, India, 2013, pp. 232–238.
[28] N. Samaan, "A novel economic sharing model in a federation of selfish cloud providers," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 12–21, Jan. 2014. doi: 10.1109/TPDS.2013.23.
[29] L. Wu, R. Ding, Z. Jia, and X. Li, "Cost-effective resource provisioning for real-time workflow cloud, Complexity" *Wiley*, vol. 15, 2020, Art. no. 1467274. doi: 10.1155/2020/1467274.

[30] R. Rajavel and M. Thangarathanam, "Agent-based automated dynamic SLA negotiation framework in the cloud using the stochastic optimization approach," *Appl. Soft Comput.*, vol. 101, 2021, Art. no. 107040. doi: 10.1016/j.asoc.2020.107040.

[31] S. Mubeen, S. A. Asadollah, A. V. Papadopoulos, M. Ashjaei, H. Pei-Breivold and M. Behnam, "Management of service level agreements for cloud services in IoT: A systematic mapping study," *IEEE Access*, vol. 6, pp. 30184–30207, 2018. doi: 10.1109/ACCESS.2017.2744677.