Wirtschaftsinformatik Proceedings 2015

Wirtschaftsinformatik

3-5-2015

Bid-Price Control for Energy-Aware Pricing of Cloud Services

Marc Premm

Bid-Price Control for Energy-Aware Pricing of Cloud Services

Marc Premm

University of Hohenheim, Department of Information Systems 2, Stuttgart, Germany
marc.premm@uni-hohenheim.de

Abstract. The amount of electrical energy consumed by Cloud computing resources keeps rising continuously. To exploit the full potential of reducing the carbon footprint, technical optimization of data center load and cooling distribution is not sufficient. We propose a method that motivates Cloud service providers to invest in energy-efficient infrastructure, which then allows for increasing revenue. The differentiation between conventional and green services offers the possibility to apply price discrimination approaches known from Revenue Management literature. Applying bid-price controlled pricing for the provider's decision on accepting an incoming request bears the potential of increased revenue. We demonstrate the efficacy of the developed artifact through an experimental evaluation for various settings of supply and demand.

Keywords: Cloud Computing, Revenue Management, Energy-efficiency, Bid-Price Control

1 Introduction

Since the amount of IT devices and their electrical energy consumption are continuously increasing [1], a lot of efforts dedicated to *GreenIT* have been carried out to reduce the carbon footprint [2]. Energy-aware computation has spread further since the advent of Cloud computing with its data centers epitomizing the huge amount of energy consumed by today's IT. The literature already provides methods to increase energy efficiency on a technical level [3]. From an economic perspective, however, there is additional potential of reducing the carbon footprint by bringing the optimal set of contractors together.

Business practitioners are beginning to realize the importance of sustainability [4] and recent studies show strong evidence that the willingness to pay for green labeled products is higher than for conventional ones [5]. An increasing number of Cloud service providers, e.g. green.ch and GreenQloud, are propagating the carbon free resource consumption of their offered services. Various levels of willingness to pay for contractually guaranteed energy-efficient Cloud service provision bear potentials to increase the providers' revenues by applying price discrimination methods. As these Cloud service providers don't reveal their internal pricing strategy, we found no evidence that this *green* aspect is also used for price discrimination. Green resources

12th International Conference on Wirtschaftsinformatik,
March 4-6 2015, Osnabrück, Germany

Premm, M. (2015): Bid-Price Control for Energy-Aware Pricing of Cloud Services, in: Thomas. O.; Teuteberg, F. (Hrsg.): Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015), Osnabrück, S. 963-977

allow Cloud service providers to offer green services, but may also host services for which no special energy clause has been included in the service level agreement. For the Cloud service provider, this versatility leads to the decision problem whether to accept an incoming request that has to be hosted on green resources, though it wouldn't be necessary from a contractual perspective. Each new contract binds resources also in future time periods, and thus may hinder future requests with a higher willingness to pay for green services. Fig. 1 shows a brief example of two potential service consumers that request services from a provider, both with different prioritizations on the energy-efficiency. The provider on the other hand has two different kinds of resources at disposal to deliver contracted services.

For modeling and simulating the distributed nature of this scenario, approaches from multiagent research are suitable [6]. Drawing from Revenue Management (RM) literature, we propose a method for the decision over accepting an incoming request with the aim to maximize revenue as an extension of existing pricing mechanisms. Previous work has shown that the principles of RM can be purposefully adapted to model Cloud computing scenarios [7]. We extent the proposed model by including two types of resources and by supporting long-term requests with an apriori unknown contract length. In contrast to other approaches that require a trusted third party, the method can be implemented by a provider with different kinds of resources without changing the way of propagating its services, e.g. the usage of green or conventional resources may be promoted by different brands of the same corporation. Hence, price-discrimination based on energy-efficiency parameters is not necessarily observable by the consumer and, thus, avoids the risk of negative side-effects.

We show that rejecting requests for conventional services in certain demand situations increases the provider's revenue and, thus, the method motivates providers to rely on energy-efficient service provision from an economical perspective with the consequence of an increased ecological sustainability. In summary, the objectives of this research are (1) to develop a method that increases expected revenues through price discrimination based on energy-efficiency and (2) to demonstrate the efficacy of the method through an experimental evaluation.

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 introduces the formal framework. The proposed method is presented in Section 4 and evaluated in Section 5. Section 6 concludes.

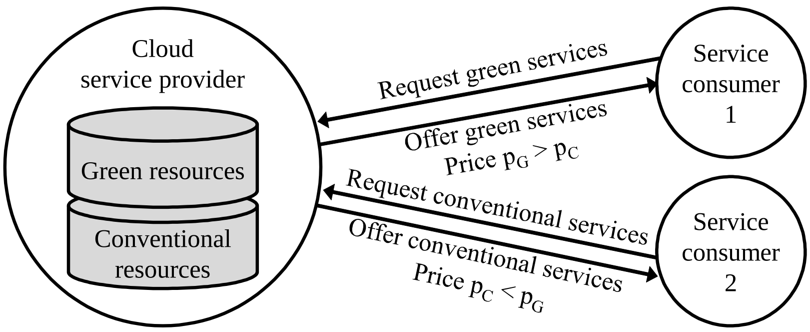

Fig. 1. Service requests for different kinds of energy efficiency

2 Theoretical Background

The theoretical background is three-fold: (i) Section 2.1 discusses related work on energy-ware pricing for Cloud computing, (ii) due to the conceptual similarities to Cloud scenarios, Section 2.2 introduces the paradigm of multiagent organizations for software agents, and (iii) Section 2.3 paves the way for the application of methods from RM.

2.1 Energy-Aware Pricing for Cloud Computing

We use the term Cloud computing in accordance with the NIST-definition [8] and require a Cloud service to satisify the following criteria: (i) automatization, (ii) standardization, and (iii) scalability. A distinction between Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) is not necessarily required for a provider that hosts the whole infrastructure, but may be relevant for a SaaS provider that needs to rent green infrastructure to offer energy-efficient services to its customers.

The terms *green* and *energy-efficient* are used synonymously throughout this work, and as the decision problem itself is independent from the various available metrics [9], we do not define or chose any. The expectations on Cloud computing are different for the main two groups of actors. Customers aim at moving fixed costs to variable costs and gain the ability of nearly unlimited resource scaling. This change in cost structure is achieved through short-term contracts with a typical duration of at least one hour. Providers on the other side focus on lower overall prices through standardization and automatization, i.e. there is a standardized set of offered services which are provisioned in an automated fashion.

Most existing approaches related to energy-efficiency in Cloud computing aim at reducing total energy consumption by optimizing load and cooling distribution within a single data centre [10], [11]. However, some already consider the contract negotiation phase as an essential step towards sustainability. Garg et al. [12] suggest introducing a Cloud broker that behaves on behalf of the service consumer and selects the most energy-efficient Cloud provider for each request. However, the use of a trusted third party is an obstacle for the successful implementation, as not every consumer is willing to give up sovereignty. An approach that can be applied to nearly any Cloud service scenario is provided by Haque et al. [13], who propose Green Service Level Agreements (SLA) with a quantifiable green rate. The proposed infrastructure enables the differentiation of green and conventional services based on electrical energy in a single datacenter, while a service provider has the possibility to accept or reject a request depending on the requested service level. The authors show that a price discriminating approach based on energy-efficiency promises increased revenues, but lack a method to handle an oversupply of green resources.

2.2 Multiagent Organizations

A multiagent system (MAS) serves as a platform for autonomous software agents that allow representing scenarios with distributed authority and, hence, fits into the context of Cloud service providers and consumers [14]. We consider autonomous software agents that are able to act and react within a MAS deputizing a natural person or another legal entity. There are conceptional similarities between the negotiations between a provider and a consumer, as well as between software agents within a MAS, also providing services to each other. The provider independent TOSCA standard (Topology and Orchestration Specification for Cloud Applications) enables the application of software agents that may autonomously search and contract Cloud services. Even if none of the actors in the Cloud computing context would actually apply agent-technologies, we may draw from MAS literature for the development of our decision model.

In a business context, however, it is not always suitable to have a totally flexible MAS, where software agents are only loosely coupled. Instead, research in the area of MAS takes the direction of stabilizing structures within MAS drawing from organizational theory. There are two main differences in the context and approaches in literature: (a) The aim of the organizational structure within the MAS is to represent the organizational structure of a social system with clearly defined boundaries [15], [16], e.g. a company with employees that are represented by a software agent, or (b) approaches from social science are used to adapt the behavior of software agents [17], e.g. to form coalitions or groups potentially across company borders.

Horling and Lesser [17] discuss the differences between different terms related to multiagent organizations, namely the most widespread ones coalitions [18], [19], teams [18], [20] and congregations [21]. They all have in common that software agents form some kind of organizational structure allowing an improved cooperation between the participants. We define a multiagent organization as a long term orientated structure that compensates software agents for providing services to the organization. Thus, we denote all software agents that have a contractual binding with a multiagent organization to provide services as its participants. However, we do not require that the goals or any subset of the goals of the multiagent organization and those of its participants are necessarily the same.

From the perspective of a single autonomous software agent a decision has to be made whether the participation in a certain multiagent organization is expected to be beneficial for the agent. In a Cloud computing context this corresponds to the decision of a provider respectively a consumer whether an offer of the counterpart will be accepted. For multiagent systems Kraus [22] suggests to adapt methods from game theory and Operations Research. While game theoretic concepts are broadly used for multiagent systems, there are fields of Operations Research that are paid much less attention. The next section presents the background on Revenue Management as part of Operations Research and its application to multiagent organizations in the context of energy-aware Cloud computing.

2.3 Revenue Management

RM has its origins in the airline industry, where due to regulations in the late 1970s, companies were searching for new ways to increase revenues, e.g. by overbooking existing resources [23]. Since then, RM has been extensively developed and has a firm place in pricing of various other goods and services, including hotel rooms as well as rental cars.

All applications of RM share some common characteristics: The provided good or services is (i) perishable, (ii) can be booked in advance and (iii) resource are scarce. The base example in an airline scenario is one scheduled flight. Empty seats on the flight cannot be stored for later use, so the revenue for this service is lost (i). Customers are able to book the flight in advance (ii), but the totally available amount of seats on the airplane is limited (iii). RM now aims at maximizing revenue by extending the set of potential customers with price discrimination, e.g. offering the same flight in various booking classes with different cancellation policies. However, the main challenge arises with the risk of selling too many low priced services and, thus, the need for an appropriate procedure to select the offered booking classes.

An approach to enhance pricing in the context of formation and extension of multi-agent organizations has been part of previous work [24]. The method uses principles of RM aiming at optimizing profits and, thus, utility for each agent applying the method. Another previous work [6] addresses the application of RM for Cloud services. Both methods assume that the requested Cloud service or organizational membership is set for a specific time horizon. However, in many Cloud computing scenarios a Cloud service usually is requested without a long lead time and doesn't have a specified end point at the time requesting the service. Instead, a Cloud service is generally requested immediately and contracted for an uncertain amount of time.

Another requirement for the application of RM methods is the presence of a possible discrimination criterion. For a Cloud service, nearly all aspects of a SLA may serve as a discrimination criterion [6]. However, pricing according to energy-related variables are already common for some industries like logistics. A higher willingness to pay of a certain set of customers for *green* services offers the possibility to skim off additional profit and, thus, sets the basis for the application of RM methods [13]. Hence, we draw from literature on RM, to design a pricing method for an energy-aware Cloud computing context.

3 Formal Framework

Conventional RM methods are focused on a certain type of product or service, like one single flight or the provision of a hotel room for one night. These products or services have a fixed time span between the start and the end. It is quite uncommon, if not entirely impossible, to book a hotel room for an unlimited time. However, this marks a difference between conventional RM applications and Cloud computing: A Cloud service is usually requested immediately for an uncertain amount of time. The applicability in general of RM methods is not affected by this difference as the main requirements pointed out in Section 2.3 are still satisfied. However, models commonly used in RM must be adapted to the energy-aware Cloud computing context. We rely on previous work connecting RM with Cloud services [7] as well as multiagent organizations [24], and adapt the models to cover requests for an unlimited time span.

3.1 Services and Resources

We consider software agents that represent Cloud providers and that provide services to a multiagent organization while receiving compensation in return. We assume that the Cloud provider has a fixed set of services $S = \{s_1,\ldots,s_N\}$, while the exact type of service is not further specified and, thus, each service might be SaaS, PaaS or IaaS. The discrete time axis is represented by $t \in \{0,\ldots,T\}$, while each t represents a certain time span depending on the type of service.

For each service, the provider can guarantee a certain minimum level of energy-efficiency for the devices used for service provision. We draw from RM literature and, thus, call the different categories *booking classes*. It is also possible to combine energy-related discrimination criteria with other contractual side conditions usually used in service level agreements, e.g. availability, response time, etc. These additional criteria could be combined to increase the discriminating distance between two consecutive booking classes. Despite the general possibility for an arbitrary number of discrimination levels, for the sake of simplicity, we only consider two booking classes $b \in B = \{Green, Conv\}$, while *Conv* represents a conventional operation without any energy-considerations and *Green* the sole execution on energy-efficient devices.

Each booking class has a fixed price $p_{s,b}$ for each service s and booking class b that is constant for all $t \in \{0,\ldots,T\}$. There are no withdrawal restrictions, so every service consumer may choose to withdraw anytime without a penalty.

The service providing agent has to rely on resources to be able to fulfill its contractual obligations. We assume that there are two types of resources: Conventional resources that are only able to be used for booking class *Conv* as well as energy-efficient resources that comply to both booking classes and, hence, are more flexible for the provider. The capacity left for both resources at time t is represented by the vector

$$\boldsymbol{x}_t = \begin{pmatrix} x^t_{Green} \\ x^t_{Conv} \end{pmatrix}.$$

The concrete type of differentiation between the two resource pools is not relevant for the presented method. There are numerous possible criteria to measure the energy-efficiency of IT resources, e.g. PUE for a whole computation center or SPECpower for a single server [9]. The expected capacity left in t for some future time $\hat{t}$ with $t < \hat{t} \leq T$ is represented by the function $E_t\left(x^t_b, \hat{t}\right)$.

We assume that each service consumes a certain amount of resources independently of the chosen booking class. The vector

$$\boldsymbol{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix}$$

expresses how much resources each service consumes. One service s_1 in booking class b = *Green* will, thus, reduce the capacity left x^t_{Green} by $s_1 m_1$. In booking class b = *Conv* the same service would be able to reduce x^t_{Green} or x^t_{Conv} by the same amount.

3.2 Requests

The value network of a service consumer requesting a Cloud service may be referenced to a multiagent organization that requests services offered by another agent. The service can be requested in both possible booking classes, while the amount of requested services is represented by the vector

$$\boldsymbol{r}_t = \begin{pmatrix} r^t_1 \\ \vdots \\ r^t_N \end{pmatrix}.$$

In accordance with RM literature, we assume that there is at most one request per time t. However, this restriction does not limit the applicability as the time span between two time points can be arbitrary small. The acceptance of an incoming request will reduce the available resources of the service provider and, hence, limit his scope for future usage. The vector $\boldsymbol{c}_t$ serves as an indicator whether and what kind of a contract has been signed in t with $\boldsymbol{c}_t = (1,0)^T$ for green services, respectively $\boldsymbol{c}_t = (0,1)^T$ for conventional services and $\boldsymbol{c}_t = (0,0)^T$ if no contract has been signed. The capacity left for the next time period is recursively calculated by

$$\boldsymbol{x}_{t+1} = \begin{cases} \begin{pmatrix} x^t_{Green} \\ x^t_{Conv} \end{pmatrix} - \boldsymbol{c}_t \left[\boldsymbol{m}\boldsymbol{r}_t \right] & \boldsymbol{x}_t - \boldsymbol{c}_t \left[\boldsymbol{m}\boldsymbol{r}_t \right] > 0 \\ \begin{pmatrix} x^t_{Green} \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left[\boldsymbol{m}\boldsymbol{r}_t - x^t_{Conv} \right] & \textit{otherwise.} \end{cases} \tag{1}$$

The second alternative in Equation (1) formalizes the possibility of the service provider to use green resources also to cover conventional service contracts. The total capacity of the provider is given in $t = 0$ with $\boldsymbol{x}_0$.

4 Pricing Mechanism

This section presents the pricing mechanism that allows a software agent representing a Cloud service provider to increase its revenue by improving the decision whether to accept or decline an incoming request. We propose a bid-price approach that calculates some sort of reservation prices for each resource and that does not affect the commonly used process of requesting a service and a response within a narrow time frame.

4.1 Decision Model

The decision that has to be made is whether to accept an incoming request or to reject it. Most Cloud service providers will accept any incoming request for a fixed price level as long as resources are available to serve the request. However, the differentiation in two different booking classes as presented in the preceding section enables more complex decision models with the aim to increase revenue. Bid-price control from RM literature uses a reservation price for each of the resources to control booking class availability. If conventional resources are available that cannot be used for providing green services, a request for the corresponding booking class should always been accepted. The more advanced green resources, however, might be used to provide conventional or green services and, thus, are the focus of the decision that has to be made. The bid-price π_t at time t represents the reservation price for green resources under which it is expected to be suboptimal to sell the lower priced conventional booking class.

For an incoming requests that involves the usage of green resources, the offering condition at time t for booking Class *Conv* is

$$\pi_t\left[\boldsymbol{mr}_t\right] \leq \sum_{s=0}^{N} r_s^t p_{s,Conv} \tag{2}$$

where the vector multiplication $\boldsymbol{mr}_t$ maps the requested services to the amount of required resources and the right part of the inequation represents the price for the incoming request in the conventional booking class.

4.2 Decision Mechanism

The decision about refusing a request even though resources would have been available has a huge difference compared to traditional RM applications: The resources saved temporally by a refused flight request, might be used to provide higher priced booking classes later on. Of course, there is the risk that the resources remain unused, but in the present case, the rejection of a request has the consequence of immediate revenue loss. This revenue has to be gained later on, and this is only possible with higher priced booking classes.

For a differentiable valuation function $V_t(\boldsymbol{x}_t)$ the bid-price is defined as the partial differentiation with respect to the relevant resource variable. As the only relevant resource for the decision in time t is x_{Green}^t, we can adapt Littlewood's rule [25] for the present case with long-term contracts and immediate service provision:

$$\pi_t := \frac{\partial}{\partial x_{Green}^t} V_{t+1}\left(\boldsymbol{x}_t\right) \approx \overline{p}_{Green} P\left(D_{Green}\left(\Delta t\right) > \hat{E}_t\left(x_{Green}^t, \Delta t\right)\right) \tag{3}$$

with the average price per resource $\overline{p}_b$, the expected demand $D_b(\Delta t)$ and the lower limit of the expected capacity left based on the current contracts $\hat{E}_t\left(x_b^t, \Delta t\right)$ each for a future time span Δt and booking class *Green*.

4.3 Forecasting Strategy

The forecasting strategy is based on the ideas of Littlewood [25], but has to be fundamentally adapted for the present case as withdrawals may occur continuously. The decision mechanism presented in the previous section is mainly determined by the quality of the demand and withdrawal forecasting, as Equation (3) is dependent on these two functions. The probability that the demand for green services will be higher than the expected resources left can provide is settled by multiple variables: (i) the distribution of both functions, (ii) their parameters, and (iii) the considered time span Δt.

We assume that the expected withdrawals follow an exponential distribution with expected mean contract length λ_b^t in *t*. As the actual contract length is not previously known, the expected mean contract length λ_b^t is determined by the average contract length of already terminated contracts. With the assumption that these past events can be used to forecast future withdrawal events, we can reformulate $E_t\left(x,\hat{t}\right)$ in *t* for all upcoming periods $\hat{t}$ to

$$E_t\left(x,\hat{t}\right) = x\left(1-e^{-\lambda_b^t\left(\hat{t}-t\right)}\right).$$

As stated before, the rejection of an incoming request for a lower priced booking class can only be beneficial for the Cloud service provider if he manages to create enough revenue with future higher priced contracts instead. With an expected contract length for green services of λ_{Green}^t and a price relation between green and conventional booking classes of $\frac{p_{Conv}}{p_{Green}}$ the time frame in which a higher priced contract should be signed is set to $\Delta t = \lambda_{Green}^t\left(1-\frac{p_{Conv}}{p_{Green}}\right)$. If the provider has to wait longer than Δt for a higher priced request, the time span is not long enough to gain more revenue despite the higher price for green services.

The demand is expected to be normally distributed with mean μ_b and standard deviation σ_b for booking class *b*. Both variables can be determined – analogously to the mean contract length λ_b^t – by the observation of past requests. For the calculation of π_t, the relevant variables μ_{Green} and σ_{Green} for booking class *Green* are observed for each time period and the convolution of multiple independent distributions has to be calculated. With the assumption that requests are equally distributed among the involved number of service providers $N_{Provider}$, the convolution of the demand distribution functions for the time interval Δt has the following two parameters:

$$\hat{\mu} = \frac{\mu_{Green}\Delta t}{N_{Provider}}$$

$$\hat{\sigma} = \sigma_{Green}\sqrt{\Delta t}.$$

With the simplification that the capacity left in the interval Δt is regarded as the mean of $E_t\left(x_{Green}^t,\hat{t}\right)$ in the corresponding interval, the probability that the expected demand in $\Delta t = \hat{t}-t$ exceeds the expected capacity left can be calculated by

$$P\left(D_{Green}\left(\Delta t\right) > \hat{E}_t\left(x_{Green}^t, \Delta t\right)\right) = 1 - \int_{-\infty}^{\frac{E_t\left(x_{Green}^t, \hat{t}\right)}{\hat{\mu}}} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2} dx \,. \tag{4}$$

With Equations (3) and (4), the bid-price π_t can be calculated in each time interval t and provide the input for Equation (2) to decide whether an incoming request will be accepted or rejected.

5 Evaluation

This section presents an experimental evaluation of the proposed artifact. We describe the experimental setup, report the simulation results, and discuss the findings.

5.1 Experimental Setup

For the simulation of the Cloud computing scenario, we used the agent-based simulation framework Repast Simphony [26]. Service consumer agents sent service requests to service providers, which may have made an offer depending on their decision model. There were two types of service providers: (1) A *regular* provider that offered services if the remaining capacity allowed it and (2) a *bid-price* provider that used the proposed method and that beside capacity restrictions also involved the forecasting of future requests. We simulated a competitive duopoly Cloud service market with one regular and one bid-price provider.

To minimize the amount of relevant variables that influence results, providers only offered one type of service that exactly used one type of resource when contracted implying $m_1 = 1$ for s_1. As only the more versatile green resources are of interest to price discrimination purposes, we considered one energy efficient computation centre and set $x_{Conv}^0 = 0$ respectively $x_{Green}^0 = 1{,}000$. The planning horizon of the provider agents was $T = 5{,}000$.

The service consumer agents requested the service in one of the two booking classes and chose the cheapest proposal respectively chose randomly when receiving two or more equivalent proposals. Service consumers of type *C1* requested only lower priced conventional services while those of type *C2* requested green services. Demand had been generated with the following parameters for the probability distribution:

- Every consumer agent generated one request in every time period with a probability of 10 percent.
- The amount of services requested was normally distributed with mean 10 and standard deviation 5.
- The contract length that was not apriori known to the providers was normally distributed with mean 1,000 and standard deviation 500.

The price for one unit of service s_1 in booking class *Green* was set to 10 monetary units per time period while the one for booking class *Conv* was reduced by 5 % with respect to the first one.

5.2 Results

The simulation was performed 100 times and for each run the values of the time period between ticks 1,000 and 3,000 have been used for evaluation to sort out fading in and out effects. Average revenue per tick was the first value we investigated. Fig. 2 shows the percentage of average revenue that the bid-price provider exceeded the average revenue of the regular provider depending on the demand setting. The number of both types of service consumer agents has been varied on an exponential scale from 1 to 1,024. Negative values represent a surplus of the regular provider and only occured if requests for the green booking class were very rarely. In all other settings the bid-price provider gained a surplus over the regular provider. The highest surplus of 3.91% has been gained with 1,024 consumers of type C1 and 64 of type C2. Fig. 3 shows the percentage of the standard deviation of the periodically gained revenue in relation with the mean value. The bid-price approach had a significantly higher standard deviation of revenue.

While in the first scenario demand heights and contract length were generated using a normal distribution, we examined a second scenario where those two variables have been generated using a Poisson distribution with the same mean parameters. Hence, the expected distribution by the bid-price provider and the actual one differed and a lower performance of the proposed approach was expected. However, the results showed the same structure as Fig. 2, given Poisson distributed demand variables and even exceeded the optimal setting for the first scenario with a surplus of 4.40%.

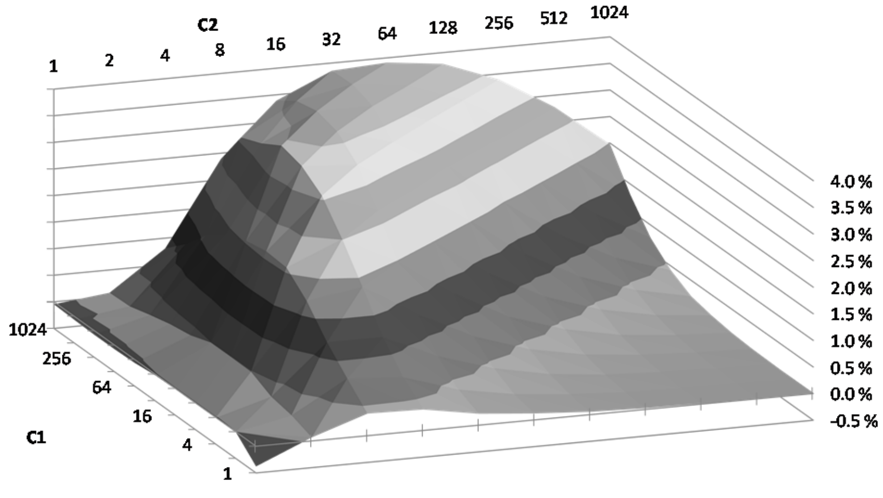

Fig. 2. Surplus of bid-price provider in relation to regular provider with respect to revenue

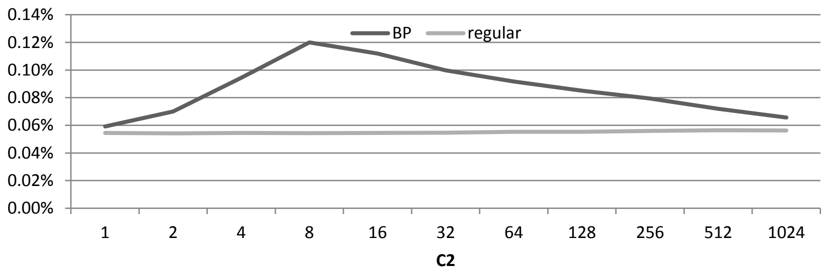

Fig. 3. Standard deviation in relation to mean for C1 = 1024

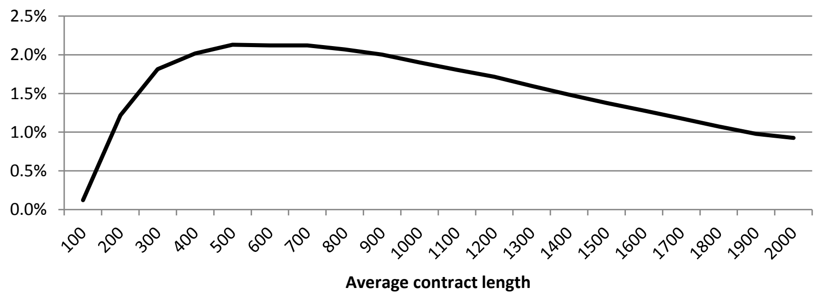

Fig. 4. Average revenue surplus of bid-price provider depending on average contract length

We also examined the impact of the relation of contract length and the amount of services per request on the average revenue of the provider. The providers faced 10 consumers of each type and the average contract length has been varied negatively correlated to the amount of requested services to remain the average demand unchanged. Fig. 4 shows the percentage of average revenue of the bid-price provider exceeding the average revenue of the regular one.

5.3 Discussion

When looking at the results of the previous section, one has to keep in mind that the price difference between the two booking classes was only 5%. This difference marks an upper limit for the proposed method and its advance is not able to exceed this limit. Fig. 2 shows that the advantage of the approach is strongly dependent on the relation between consumers with a higher and those with a lower willingness to pay. The highest surplus was gained when the amount of C1 consumers was high and the amount of C2 consumers was above a certain limit, approximately 8. In this case, most requests sent to the providers aim at the conventional booking class and blocked the resources of the regular provider, while the bid-price approach denied most of those requests. When additionally the number of C2 consumers increased, the probability that a request achieved a higher revenue increased analogously and, hence, reduced the advantage of the proposed method. Even in non-optimal settings the bid-price approach achieved a positive surplus in nearly all cases.

Even with a demand distribution function that the bid-price provider did not apriori anticipated, he gained a significant surplus in many settings. The standard deviation of the periodical revenue was significantly higher, but on an extremely low level with respect to the mean value.

The average revenue was of course highly dependent on the average contract length. If all or most contracts only last one period, there would be no need to reserve capacity for future requests. The same holds for the number of requests per time period, as a rejected request has to be compensated by another higher priced request within a relatively short amount of time. The reasonable applicability of the proposed method, thus, depends on the relation between contract length and request frequency. However, we did not observe significant drawbacks in settings that do not meet these requirements.

6 Conclusion

This paper presents a method to increase expected revenues of a Cloud service provider using price discrimination. We adapted and applied approaches from Revenue Management and used the varying willingness to pay for ecologically sustainable services as the discrimination criterion. The model successfully introduced a rolling time horizon that is not common for Revenue Management methods, but necessary to represent real-world Cloud computing scenarios. We evaluated the proposed method in a multi-agent simulation and showed its efficacy depending on different settings of supply and demand. The method gains an advantage on regular approaches by rejecting lower priced requests when higher future revenue is expected.

The demand for an enhanced sustainability of corporate IT keeps increasing continuously and Cloud service provider cannot ignore this trend [4]. Our research has three implications for practice. First, the proposed method increases revenues for Cloud service providers that are hosting green computing resources. Second, the possibility of higher revenues additionally motivates providers to invest in more versatile energy-efficient infrastructure fed with carbon free produced electrical energy and incorporating price discrimination possibilities. Third, the individual optimizing of Cloud service providers in terms of revenue leads to a substitution of old inefficient resources by new versatile ones and, thus, a society-wide improvement of ecological sustainability.

The approach is limited to markets where demand exceeds supply as the scope is restricted to accepting or rejecting a request, while there are no significant drawbacks in other situations and may thus also be used in scenarios with apriori unknown demand-supply ratio. Another prerequisite is that a new contract reserves sufficient resources in relation to the request frequency, while otherwise a compensation of a rejected request is rarely possible. The discrimination criterion energy-efficiency may even be combined with other aspects that are usually part of SLAs and, hence, justify a higher price difference than 5% with a simultaneously increased revenue potential.

Extensions that involve decommitment penalities into the negotiation phase of Cloud service provision show major similarities to cancelation respectively withdrawal fees known from traditional RM applications and may further enhance results [6]. The bid-price control approach and thus the proposed model is easily extendable and adaptable, e.g. by adding additional discrimination levels. The proposed method is highly depending on the quality of demand forecasting, hence, advanced forecasting methods [27] promise to further increase revenue especially in non-optimal settings of supply and demand. Selling more services than the available resources may deliver is practiced in traditional RM domains as overbooking. Findings from RM indicate that combining the proposed method with overbooking positively affects revenue [28].

Acknowledgements. This work has been supported by (1) the project MIGRATE!, funded by the German Federal Ministry of Economics and Technology (BMWi, FKZ 01ME11052), and (2) the eHealthMonitor project (http://www.ehealthmonitor.eu), funded by the European Commission under contract FP7-287509.

References

1. Cook, G., Dowdall, T., Pomerantz, D., Wang, Y.: Clicking Green: How Companies are Creating the Green Internet. Greenpeace, Washington (2014)
2. Murugesan, S.: Harnessing Green IT: Principles and Practices. IT Pro 01-02, 24-33 (2008)
3. Berl, A., Gelenbe, E., Girolamo, M., Giuliani, G., Meer, H., Dang, M.Q., Pentikousis, K.: Energy-Efficient Cloud Computing. The Computer Journal, vol. 53, no. 7, 1045-1051 (2010)
4. Brooks, S., Wang, X., Sarker, S.: Unpacking Green IS: A Review of the Existing Literature and Directions for the Future. In: Brocke, J., Seidel, S., Recker, J. (Eds.), Green Business Process Management - Towards the Sustainable Enterprise, pp. 15-37. Springer (2012)
5. Cronin, J.J., Smith, J.S., Gleim, M.R., Ramirez, E., Martinez, J.D.: Green marketing strategies: an examination of stakeholders and the opportunities they present. Journal of the Academy of Marketing Science 39, 158-174 (2011)
6. An, B., Lesser, V., Irwin, D., Zink, M.: Automated Negotiation with Decommitment for Dynamic Resource Allocation in Cloud Computing. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, pp. 981-988, Toronto (2010)
7. Anandasivam, A., Premm, M.: Bid Price Control and Dynamic Pricing in Clouds. In: Proceedings of the 17th European Conference on Information Systems, pp. 238-241, Verona (2009)
8. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. National Institute of Standards and Technology (2011)
9. Daim, T., Justice, J., Krampits, M., Letts, M., Subramanian, G., Thirumalai, M.: Data cener metrics – An energy efficiency model for information technology managers. Management of Environmental Quality, vol. 20, no. 6, 712-731 (2009)
10. Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. In: Future Generation Computer Systems 28, 755-768 (2012)
11. Baliga, J., Ayre, R.W.A., Hinton, K., Tucker, R.S.: Green Cloud Computing: Balancing Energy in Processing Storage and Transport. Proceedings of the IEEE 99, 149-167 (2010)

12. Garg, S.K., Yeo, C.S., Buyya, R.: Green Cloud Framework for Improving Carbon Efficiency of Clouds. In: Jeannot, E., Namyst, R., Roman, J. (eds.) Euro-Par 2011. LNCS, vol. 6852, pp. 491-502. Springer, Heidelberg (2011)
13. Haque, M.E., Le, K., Gori, I., Bianchini, R., Nguyen, T.D.: Providing Green SLAs in High Performance Computing Clouds. In: Proceedings of International Green Computing Conference, Arlington (2013)
14. Huhns, M. N., Singh, M. P. (Eds.): Research Directions for Service-Oriented Multiagent Systems. IEE Internet Computing Nov.-Dec., 65–70 (2005)
15. Kirn, S., Gasser, L.: Organizational Approaches to Coordination in Multi-agent Systems. Journal it+ti 4, 23-29 (1998)
16. Hübner, J.F., Boissier, O., Kitio, R., Ricci, A.: Instrumenting multi-agent organisations with organisational artifacts and agents. Autonomous Agents and Multiagent Systems 20, 369-400 (2010)
17. Horling, B., Lesser, V.: A survey of multiagent organizational paradigms. The Knowledge Engineering Review 19:4, 281-316 (2004)
18. Wooldridge, M.: An Introduction to MultiAgent Systems (2nd Ed.). Chichester: John Wiley & Sons (2009)
19. Sandholm, T. W.: Distributed Rational Decision Making. In: G. Weiss (Ed.), Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence, Chapter 5, pp. 201-258. MIT Press (1999)
20. Jennings, N. R.: Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. Artificial Intelligence 75, 195-240 (1995)
21. Brooks, C., Durfee, E.: Congregation formation in multiagent systems. Autonomous Agents and Multiagent Systems 7 (1-2), 145-170 (2003)
22. Kraus, S.: Negotiation and Cooperation in Multi-Agent Environments. In: Artificial Intelligence 94, 79-97 (1997)
23. Talluri, K.T., van Ryzin, G.J.: The Theory and Practice of Revenue Management. Springer, New York (2004)
24. Premm, M., Widmer, T., Karaenke, P.: Bid-Price Control for the Formation of Mulitagent Organisations. In: Klusch, M., Thimm, M., Paprzycki (eds.) MATE 2013. LNAI, vol. 8076, pp. 138-151. Springer, Heidelberg (2013)
25. Littlewood, K.: Forecasting and control of passenger bookings. In: Proceedings of the Twelfth Annual AGIFORS Symposium, Nathanya (1972)
26. North, M., Collier, N., Ozik, J., Tatara, E., Macal, C., Bragen, M., Sydelko, P.: Complex adaptive systems modeling with repast simphony. Complex Adaptive Systems Modeling 1(1) (2013)
27. Syntetos, A.A., Boylan, J.E., Disney, S.M.: Forecasting for inventory planning: A 50-year review. The Journal of the Operational Research Society 60, 149-160 (2009)
28. Chiang, W.C., Chen, J., Xu, X.: An overview of research on revenue management: Current issues and future research. International Journal of Revenue Management 1(1), 97-128 (2007)