Future Generation Computer Systems 41 (2014) 79–90

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Cost model based service placement in federated hybrid clouds

Jörn Altmann *, Mohammad Mahdi Kashef

Technology Management, Economics, and Policy Program, Department of Industrial Engineering, College of Engineering, Seoul National University, Seoul, South Korea

HIGHLIGHTS

- Comparison of cost models for cloud computing.
- Design of a comprehensive cost model for federated hybrid clouds.
- Design of an algorithm for service placements in federated hybrid clouds.
- Performed a sensitivity analysis with the help of the service placement algorithm.

ARTICLE INFO

Article history:
Received 11 September 2012
Received in revised form
17 August 2014
Accepted 19 August 2014
Available online 3 September 2014

Keywords:
Cloud federation
Service migration
Information technology cost factors
OPEX and CAPEX
Cloud economics
Cost model comparison

ABSTRACT

As cloud federation allows companies in need of computational resources to use computational resources hosted by different cloud providers, it reduces the cost of IT infrastructure by lowering capital and operational expenses. This is the result of economies of scale and the possibility for organizations to purchase just as much computing and storage resources as needed whenever needed. However, a clear specification of cost savings requires a detailed specification of the costs incurred. Although there are some efforts to define cost models for clouds, the need for a comprehensive cost model, which covers all cost factors and types of clouds, is undeniable. In this paper, we cover this gap by suggesting a cost model for the most general form of a cloud, namely federated hybrid clouds. This type of cloud is composed of a private cloud and a number of interoperable public clouds. The proposed cost model is applied within a cost minimization algorithm for making service placement decisions in clouds. We demonstrate the workings of our cost model and service placement algorithm within a specific cloud scenario. Our results show that the service placement algorithm with the cost model minimizes the spending for computational services.

1. Introduction

The cloud computing paradigm has been established as a promising model for using computational resources and software resources on-demand [1–3]. As one of its advantages, it shifts complex IT resource management tasks typically performed by a company that is in need of computational resources to service providers. It also increases the flexibility of companies to adapt their IT resources to changes in demand [4,5]. Assuming a cost-based pricing of cloud services, it can reduce infrastructure investment cost and cost for organizing IT resources. The capital expenditure (CAPEX) and operational expenditure (OPEX) are lower due to the economies of scale of the cloud providers' infrastructure.

A potential scenario for the foreseeable future is that of a hybrid cloud environment. Hybrid clouds refer to a composition of a private cloud (i.e., data center managed through cloud technology and owned by the company in need of computational resources) and at least one public cloud (i.e., resources owned by a cloud provider) [6]. Most enterprises, except for those with a very small or stable demand for computing resources, will employ a mix of in-house and outsourced computing resources [7]. The reason for keeping a private cloud, despite the economies of scale of public clouds, is the increased security risks in the public cloud and possibility of serving the base demand at lower cost [8]. Therefore, with the ability of a seamless migration of processing load between a private cloud and public clouds, the hybrid cloud is ideal for handling variable demand [9].

To truly fulfill the promise of cloud computing (i.e., its benefit from economies of scale), technology is needed to federate dis-

* Corresponding author. Tel.: +82 7076786676; fax: +82 15016415384.
E-mail addresses: jorn.altmann@acm.org (J. Altmann), mmkashef@temep.snu.ac.kr (M.M. Kashef).

http://dx.doi.org/10.1016/j.future.2014.08.014
0167-739X/

parate public clouds at the infrastructure level (IaaS). Only through federation and interoperability at this level, small cloud providers (i.e., providers of limited scale) can take advantage of their aggregated capabilities and provide a seemingly infinite service computing utility that can compete with the bigger cloud providers in the market. We refer to the infrastructure that supports this paradigm as federated clouds [10]. By definition, a federated cloud is a cloud, in which (competing) cloud providers have reached a cross-site agreement for cooperating regarding the deployment of service components in a way similar to electrical power providers who use capacity from each other to cope with demand variations among their own customers [11].

To take advantages of hybrid federated clouds, companies need to know the overall cost of running their services on this kind of cloud infrastructure [12–15]. A service runs on one or more virtual machines (VMs) and can be moved (with the VMs) between federated clouds. The reports on the cost savings of clouds vary, however. For example, a report by McKinsey states that moving to the cloud would actually cost 144% more than the current data center infrastructure [9,16]. West reports that cloud technology reduced a government agency's cost by 25%–50% [16]. This uncertainty about cost savings and about actual expenditure were also discussed by Kondo et al. [4]. These examples highlight the necessity for an overall cost model.

Apart from the cost model for federated hybrid clouds, the service placements on this kind of cloud infrastructure need to be considered by the company that is in need of computational resources, as each service placement incurs a different cost. Hence, minimizing the overall cost for using federated hybrid clouds requires an optimization model that considers all cost factors and cost values for each of the service placements [17,18]. Whenever the configuration and prices of clouds change over time, the optimization problem has to be solved again [11,14,19].

Summarizing this discussion, our research can be described with two research questions: (1) What are the cost factors specific to federated hybrid clouds and how can a comprehensive cost model for federated hybrid clouds be defined? (2) If such a cost model for federated hybrid clouds exists, how can it be used for a service placement optimization algorithm?

In particular, we intend to construct an overall cost model that can be used by enterprises to decide where to place their services on a federated hybrid cloud such that it minimizes cost. This overall cost model, which is based on the initial and less comprehensive version of the cost model given in [20], comprises all cost factors and cost functions, which are necessary to estimate the precise cost of running cloud services on an in-house cloud-enabled data center (private cloud) and on federated public clouds.

To answer the research questions, we conduct the following steps: First, we perform a systematic literature review of papers on cost factors and cost models in cloud computing. Second, we identify the gaps in the current research on federated hybrid cloud cost modeling. Based on this result, we design a comprehensive overall cost model for federated hybrid clouds. In the next step, we propose an optimization algorithm for service placement on federated hybrid clouds, using this cost model. Finally, we apply the proposed cost model and the service placement optimization algorithm to a case study to demonstrate its workings.

The contributions of this paper are threefold: First, it comprises a comprehensive cost model that cannot only be used for cost calculation of federated hybrid clouds (as in this paper) but also to guide and compare investment decisions in private clouds. Second, the cost model considers cost for the migration between clouds which includes the deployment cost and the transmission of data between public clouds (and not just between private clouds and public clouds). Third, a brute-force service placement algorithm that considers the cost of service placements.

The remainder of the paper is organized as follows. The next section gives an overview about cost factors and cost models for cloud computing, which have been proposed in literature. Section 3 introduces our cost model for federated hybrid clouds. In Section 4, our cost-model-based service placement optimization algorithm is presented. Furthermore, an application of our cost model as part of the service placement algorithm to a cloud computing scenario is shown. Section 5 concludes the paper with a brief discussion.

2. Background

2.1. Cloud computing, hybrid clouds, and federated clouds

Although there are many definitions of cloud computing, the NIST definition seems to have captured the commonly agreed cloud computing aspects that are mentioned in most of the academic papers [5]. The NIST definition states that cloud computing is "A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [5]. For the purpose of our paper, the definition given by NIST suffices.

Furthermore, we assume that a cloud provider can own several clouds [3]. Each of these clouds is assumed to be at different geographical locations (e.g., different regions of Amazon AWS are considered to be different clouds) and could use the same cloud standard. Besides, from an economic perspective, there is no difference between a cloud provider owning the data center and a cloud provider that rents hardware from a data center provider [3]. The only difference from the perspective of a hardware-renting cloud provider is that this provider outsources the maintenance of the data center hardware. The cloud provider's profit comes from the difference in selling cloud services on demand and the fixed renting cost paid to the data center provider. The data center provider benefits from a 100% utilization of his hardware services and does not need expertise on cloud computing.

The research of this paper is related to two cloud properties, namely, hybrid clouds and federated clouds. The definition of hybrid clouds used in this paper follows the one of Metzler and Taylor as well as Van den Bossche et al. [18,21], in which organizations use public clouds to cover their demand for computational resources in excess of the capacity of their private cloud.

With respect to federated clouds, technology is needed to combine disparate public clouds, including those owned by different organizations. Only through federation (including its interoperability requirement) can a single cloud provider take advantage of the aggregated capabilities to provide a seemingly infinite service computing utility. We refer to this category of clouds as federated clouds [10]. In detail, federated clouds comprise clouds of (competing) cloud providers, who have reached a cross-site agreement for cooperating regarding the deployment of service components (e.g., through a marketplace of standardized goods [22,23]). The concept is similar to electrical power providers, who use capacity from each other to cope with demand variations among their own customers [11].

Fig. 1 illustrates the above definition of cloud federation by showing an example of a cloud customer (company) that is in need of computational resources. This cloud customer runs its private cloud (i.e., its own data center with cloud technology) to host its security-critical services. Beside the private cloud, the company uses two different clouds (i.e., public cloud 1 and public cloud 2) for services that are needed at times of peak demand. The two public clouds offer the same cloud interfaces to each other and the cloud customer, following the cloud federation agreement between the two public clouds.

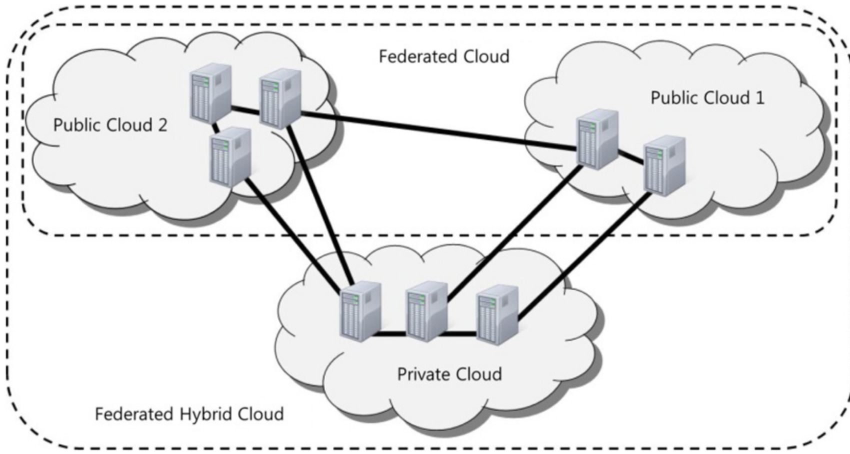

Fig. 1. Schematic view of a federated hybrid cloud, which consists of a private cloud and two public clouds (owned by two different providers). The arrows represent the communication between the services.

Table 1
Cloud categories.

			Cloud ownership		
			Clouds are owned by the same provider	Clouds are owned by different providers	One cloud is owned by the cloud customer and the remaining clouds are owned by different providers
Number of clouds	One cloud	One cloud standard is used	Category 1: Public clouds		Category 2: Private clouds
	Two or more clouds	Different cloud standards are used	Category 6: Differentiated clouds (clouds are not interoperable)		Category 3: Hybrid clouds (private cloud can access public clouds using their cloud standards)
		One cloud standard is used	Category 7: Distributed clouds	Category 4: Federated clouds (standard used is based on agreement)	Category 5: Federated hybrid clouds

Clarifying the differences between these definitions of public cloud, private cloud, hybrid cloud, and federated clouds, we categorize these clouds with respect to the ownership of the clouds, the number of clouds involved, and the number of cloud standards used (Table 1).

The first category (public clouds) describes clouds, which cloud customers do not own but use to fulfill their computing service needs. The cloud provider uses a cloud infrastructure standard, which can be proprietary. The second category (private clouds) defines the company's private data center as a private cloud. The cloud customer, who owns the data center, uses the data center with cloud computing technology to meet all of its computational needs. The third category (hybrid clouds) describes interconnected clouds, in which one of the clouds is owned by the cloud customer and the second cloud is a public cloud. This category defines a combination of a private cloud and a public cloud. The fourth category (federated clouds) represents public clouds that use the same cloud infrastructure standard. Therefore, VMs can easily be migrated between the federated clouds owned by different cloud providers. A cloud federation requires, at least, an agreement between cloud providers to commit to a specific cloud infrastructure standard. The fifth category (federated hybrid clouds) represents clouds that are the focus of this paper. A cloud customer, who uses federated hybrid clouds, runs some of its services on its private cloud and some others on a federation of public clouds (federated clouds). The sixth category (differentiated clouds) represents the overall cloud market, in which cloud providers offer their services using different standards [22]. Consequently, the different public clouds are not interoperable. A cloud customer, who wants to use several public clouds, would need to use the standards of each cloud. The seventh category comprises clouds, which use the same cloud standard and are owned by the same provider. Those clouds are located at different geographical locations (e.g., Amazon AWS).

2.2. The need for cost models for clouds

The majority of participants of a survey, which Rayport and Andrew conducted [24], indicated that their company already used public clouds, or discussed, planned, trialled, or implemented the use of a cloud infrastructure. Their objective for using cloud technology has been the potential reduction of the company's data center costs. Even though there is a lot of mentioning about the benefits and even inevitability of migrating to a cloud [25], the exact costs are still unknown. This lack of cost details (and its accounting within a cost model) makes any decision on a migration to clouds uncertain [2,26–29].

Many other statements about the need for cost models for clouds have also been made [4,26,27,30,31]. Using these cost models, one may investigate economic factors like Return-on-Investment (ROI), Net-Present-Value (NPV), Benefit-to-Cost-Ratio (BCR), and Discounted-Payback-Period (DPP). These measures are essential to decide when and under which conditions it is better to use private or public clouds [26,27].

However, to perform this cost analysis, detailed knowledge about applications, hardware, and load levels is required [1]. The difficulty of obtaining this knowledge makes it non-trivial to estimate short-term or long-term costs. As Ali et al. pointed out, it is difficult to know "the actual resources consumed by a system", "the deployment option used by a system, which can affect its costs as resources", and the "probable changes in the cloud service provider's pricing scheme" [30].

2.3. Existing research on cost factors and cost models

In order to find all major papers addressing our research objectives, we searched research databases with a combination of keywords from two groups. The first group comprises the keywords: cloud computing, elastic computing, utility computing, Infrastructure-as-a-Service, IaaS. The second group of keywords includes cost model, business model, cost estimation, cost parameters, and economics. The research databases used are IEEE Explore, ACM Digital Library, ISI Web of Knowledge, SpringerLink, ScienceDirect, and Google scholars.

In total, we have found 42 papers pointing to both fields of cloud related and cost or economy related subjects. Those works can be categorized into three groups according to the cloud level considered: (1) transaction level; (2) application level (e.g., educational system); and (3) VM level, which is independent of the application of VMs. As an example of the first group, Buell and Collofello have addressed transaction level cost factors: processing, storage, bandwidth, and services costs [32]. Many of the works can be categorized into the second group as they focus on a particular topic [32–36]. Among the papers of the third group, only 13 papers have proposed cost factors and detailed cost models. These papers are introduced in the following paragraphs in more detail.

The first paper by Risch and Altmann evaluates some claims about the financial advantages that companies would gain from using a commercial cloud [37]. The authors have calculated costs in four cloud scenarios. For that, a comparison of the costs for a company, which purchases resources from Amazon EC2 or from a hardware vendor for its in-house data center with cloud technology (private cloud), were pointed out.

For deciding when and under which situation it is recommended to move to a cloud service, a very good coverage of cost factors has been given by Tak et al. [1]. The authors have also classified cost factors into quantifiable cost factors and less quantifiable cost factors (which are more difficult to estimate). In addition, their classification includes a grouping into direct or indirect costs, which eases the understanding of all aspect of the costs.

Even though Kondo et al., the authors of [4], have not proposed any cost model, they conducted an economic comparison. For that, they covered all cost aspects of a cloud project. In total, they considered 11 cost factors, which will be discussed in detail in Section 2.4.

Besides identifying six cost factors, Armbrust et al. also compared cloud and in-house cases, using a trade-off formula [38]. The authors did not focus on cost modeling but gave a general understanding of various aspects of cloud computing. They have pointed to the cost-benefit tradeoff of migrating to the cloud, and have proposed a simple cost formula to support decision making.

The challenges, which enterprises face when applying hybrid cloud models, were discussed by Hajjat et al. [39]. Component placement is one of the challenges, addressing the decision on which of the components must be kept local and which components can be migrated. Another challenge is the specification of cost factors with respect to computing components, storage components, and wide-area communication. As the cost savings from migration depend on the placement of compute-intensive, storage-intensive, and communication-intensive components, graph theory has been used to model the network of components and their data transfer relations.

Truong and Dustbar present a service for estimating and monitoring costs [26]. The service distinguishes three situations: the complete use of on-premise resources, the partial use of cloud resources, or the complete use of cloud resources. In addition, they discuss the composability of cost models and, for this, have introduced seven basic cost models. The data used for the comparison of the models came from cloud providers' pricing specifications.

Alford and Morton present an economic analysis to investigate the potential savings from migrating services to the cloud. They focus on IT data centers and use a proprietary cost model [27]. The study takes into consideration transition costs, life-cycle operations, and migration schedules. Although they did not specify the cost model, they emphasize the existence of the cost model that is owned by the Booz Allen Hamilton Inc. The cost model is used to compare the three scenarios: public cloud, hybrid cloud, and private cloud.

Khajeh-Hosseini introduced a cloud adoption toolkit [30]. The toolkit provides a collection of tools that support decision making with respect to the adoption of cloud computing in an enterprise. However, the cost model mentioned here has not been described. Only the cost factors have been described.

Opitz et al. performed a very detailed analysis of costs [40]. The authors conducted an analysis of different cost factors that have to be considered by a provider of a computational Grid. This analysis uses realistic values. In addition to this, the authors have shown an estimation of the total costs for resource providers in two real-world examples.

Although the main contribution of Altmann and Rohitratana is in the area of Software-as-a-Service only, some of the considered cost factors for the software selection process can be used in any infrastructure cloud environment [41]. The authors have focused on the software license selection support. Their model helps users to select a SaaS Licensing (SaaSL) model or a Perpetual Software Licensing (PSL) model.

Hwang et al. introduce an algorithm for service providers to minimize their service provision cost when offering reserved resources and on-demand resources [33]. For their calculation, they consider server usage, data transfer, and storage usage in data centers.

In the work by Patel and Shah, the focus is on a cost model for data centers [6]. In particular, the cost factors for electricity, network devices, labor, software, and business premises are analyzed. The authors also present some economic models for decision support.

Amazon, as a cloud provider, supports its customers with a simple web-based cost calculation tool for its AWS services [42]. In addition to the cost for storage, servers, and data transfer, cost for middleware services (PaaS) can also be specified. Within this work, we do not consider middleware services.

2.4. Cost factors

The 13 articles presented in the previous section identified 21 different cost factors that are relevant for IaaS services of clouds. The columns labeled 'Literature' in Table 2 show the cost factors that have been used by a specific article. The paper identifiers used in Table 2 are the same as the literature references. Furthermore, the cost factors are structured and categorized into six main groups: electricity, hardware, software, labor, business premises, and service. Each of these main groups is further subdivided into subgroups. It is notable that these groups overlap with a categorization into public clouds and private clouds. More specifically, cost factors that are important for private clouds are listed in rows a to d and row f1 of the Table. The remaining rows show cost factors that are important for public cloud use. The meaning of the cost factors is as follows:

Electricity: The power usage of in-house electronic devices like servers, gateways, routers, and other network devices is grouped in one subgroup [1,4,6,27,38,40,37]. In another subgroup, the electricity consumed through cooling is categorized separately from other electronic devices [1,6,27,38,40]. Additionally, for achieving accurate estimates, two kinds of values can be considered for all

Table 2
Contributions of existing research articles to cost factors.

Cost type	Cost factor	Cloud category applied to	Expected impact of cost factor	Literature												
				Number of factors considered in specific articles												
				6	11	8	5	5	4	10	4	10	4	4	9	5
(a) Electricity	(a1) Cooling	Private cloud	+++									[40]			[6]	
	(a2) Electronic devices (idle)	Private cloud	+++									[40]				
	(a3) Electronic devices (use)	Private cloud	+++	[37]	[1]	[4]	[38]			[27]		[40]			[6]	
(b) Hardware	(b1) Server	Private cloud	+++	[37]	[1]	[4]				[27]		[40]	[41]			
	(b2) Network device	Private cloud	+		[1]					[27]			[41]			
(c) Software	(c1) Basic server software license	Private and public cloud	++		[1]					[27]			[41]		[6]	
	(c2) Middleware license	Private and public cloud	++									[40]			[6]	
	(c3) Application software license	Private and public cloud	++		[1]					[27]		[40]	[41]			
(d) Labor	(d1) Software maintenance	Private and public cloud	+		[1]					[27]					[6]	
	(d2) Hardware maintenance	Private cloud	+		[1]					[27]		[40]			[6]	
	(d3) Other support	Private cloud	+		[1]	[4]				[27]		[40]				
(e) Business Premises	(e1) Rack, air conditioner	Private cloud	+							[27]					[6]	
	(e2) Cabling	Private cloud	+							[27]					[6]	
	(e3) Facility	Private cloud	+				[38]								[6]	
(f) Cloud Service	(f1) Internet connectivity of private cloud	Private and public cloud	+			[4]		[39]								
	(f2) Cloud server use	Public cloud	+++	[37]		[4]		[39]	[26]		[30]			[33]		[42]
	(f3) Data transfer into cloud	Public cloud	+++	[37]	[1]	[4]	[38]	[39]	[26]		[30]	[40]		[33]		[42]
	(f4) Data transfer from cloud	Public cloud	+++	[37]	[1]	[4]	[38]	[39]	[26]		[30]	[40]		[33]		[42]
	(f5) Cloud storage use	Public cloud	++	[37]	[1]	[4]	[38]	[39]	[26]		[30]			[33]		[42]
	(f6) Data transfer between clouds	Federated cloud	+++													[42]
(g) Deployment	(g1) Number of deployments	Federated cloud	+++													

devices: the power consumption, when the system is idle, and the power consumption, when the system is heavily used [40].

Hardware: Hardware cost refers to the acquisition of hardware resources. In particular, two subgroups can be distinguished. One comprises the purchasing cost of computing hardware needed in-house [1,4,27,40,41,37] and the purchasing cost of network devices (e.g., switches, routers) needed in-house [1,27,41]. In order to determine the actual annual costs of the resources, their economic lifetime has also to be considered, as indicated by Opitz et al. [40]. A depreciation period of three years is widely used for this kind of technology.

Software: The purchasing price of licenses of software, which is used in-house, is considered in this class. There are three subgroups of software, which should be considered separately: basic server software licenses, middleware software licenses, and application software licenses. Basic server software license refers to operating system software and back-up software [1,6,27,41]. Middleware software licenses refers to licenses for commercial middleware software that is needed for running applications [6,40]. Application software licenses comprise any kind of software that directly contributes value to users (e.g., enterprise resource planning software) [1,27,40,41]. Although there are many types of pricing of software licenses possible [43], we consider a simple perpetual licensing model in this paper.

Labor: This category comprises three subgroups. The subgroups are salaries for technicians, who work on maintaining software [1,6,27], salaries for hardware technicians [1,6,27,40], and salaries for technicians, who provide support [1,4,27,40]. These cost factors are impacted by the region or country, in which the cloud is located.

Business Premises: This category includes the basic costs, which are essential to establish an in-house data center with cloud technology (private cloud). These basic costs are grouped into three subgroups, namely according to the cost for renting or purchasing the data center facility [6,38], the cost for racks and other non-electronic instruments, which are required for the safety and the reliability of the data center [6,27], as well as the cost of cabling of the data center [6,27].

Service: This group of cost factors, which is categorized as services, comprises in-tangible items. Seven subgroups can be distinguished. Some of the service costs may incur independent to whether a company uses in-house data center services or cloud-based services. The charge for Internet connectivity (i.e., Internet access charges for the enterprise), which is such a cost factor, comprises the first subgroup [4,39]. Another cost factor, which defines the second subgroup, is the usage cost for servers (CPU hours) [4,26,30,33,39,37]. The cost for incoming data transfer (e.g., cloud data transfer in) is another factor, defining subgroup three [1,4,26,30,33,38–40,37]. Correspondingly, there is the cost for the amount of outgoing data transfer (i.e., data that is transferred from the cloud) [1,4,26,30,33,38–40,37]. The cloud storage factor, which is the fifth subgroup, describes the amount of storage, which is used in the cloud [1,4,26,30,33,38,39,37]. Subgroup 6 comprises the cost of data transfer between clouds. This cost is a specific cost that needs to be considered for federated clouds. This cost factor, which did not get attention in literature yet, could be a substantial cost item in federated clouds. Amazon prices data transfer between its clouds at different geographical locations at a price higher than between clouds at the same location [42].

Deployment: The last cost factor considers the number of service deployments in clouds. For calculating the service deployment cost, the number of service deployments per cloud is multiplied with the cost for migrating a VM to this cloud. For simplification without losing generality, the cost of migration is assumed to be constant for a cloud. This cost factor did not get attention in earlier literature, as the number of deployments in hybrid cloud scenarios is negligible due to the rare events that could trigger a new migration. In federated clouds, distributed clouds, and federated hybrid clouds, however, where the probability of changing cloud prices, performance changes, and changing input conditions is much higher, the total deployment cost can become substantial if any of those events triggers a new service placement. Therefore, a cost model needs to consider the number of service deployments. The number of service deployments has an impact on whether an optimized service placement is economically efficient to be implemented.

2.5. Shortcomings of existing cost models

Among these 13 papers, which have been identified in this review to cover cost factors and cost models, just three have presented their own sophisticated quantitative cost model. These are the works by Truong and Dustbar [22], Armbrust et al. [38], and Hajjat et al. [39]. Although there are three more papers, which mentioned to have their own cost models, they did not provide any details.

Moreover, the three papers by Truong and Dustbar, Armbrust et al., and Hajjat et al. center only on either an analysis of certain scenarios or provide a very generic economic solution that has not been adapted well to the cloud area. For example, Armbrust et al. have proposed a very general economic formula, which needs significant adaptation to be applicable in practice in the cloud industry [38]. Although the paper of Hajjat et al. provides details about an optimization application, no cost formula has been specified [39]. Clearly, any kind of cost optimization should be based on the total cost. Truong and Dustbar have suggested seven formulas, each one of them addressing a certain activity in clouds [26]. In case of hybrid clouds, those formulas are not sufficient. The formulas lack information about in-house data center costs. Moreover, with respect to cloud federation, data transfer cost between clouds needs to be considered as well.

For these reasons, none of the three works by Armbrust et al. [38], Hajjat et al. [39], and Truong and Dustbar [26] can provide a comprehensive estimation of costs for using federated hybrid clouds. Most probably, using those models leads to inaccurate estimates. Consequently, any optimization based on these estimates will be of little help for cloud consumers. This uncertainty is an obstacle for applying the proposed cost models in practice for federated hybrid clouds.

2.6. Optimization of service placements

Even though knowing the estimated overall cost of services of an enterprise is essential, a further step towards finding the optimal allocation of services in the federated hybrid cloud is required. This is necessary as different service placements may incur different costs. For example, the different cost might come from different prices set by cloud providers, geographical differences in electricity prices, or from different fees for Internet connectivity [44,15,17,19,33,45,18,46].

Besides, the rate of data transfer between services might be different as different types of applications interconnect. Therefore, for example, keeping services with high traffic within one cloud decreases the traffic cost. To predict the traffic rate, Koch et al. proposed a "workload-aware method of provisioning", which is effective for the case of an educational institution [45]. If the workload characteristics and parameters of the domain are known, then an accurate prediction of future traffic among services is possible. In the general case, such an optimization model is not possible.

Hwang et al. proposed a cost optimization model for service provisioning, considering two types of pricing plans, namely on-demand server pricing and reserve-server pricing [33]. The objective of this work has been to support cloud providers in lowering their service provisioning cost. Consequently, the authors have neither addressed hybrid clouds nor federated clouds.

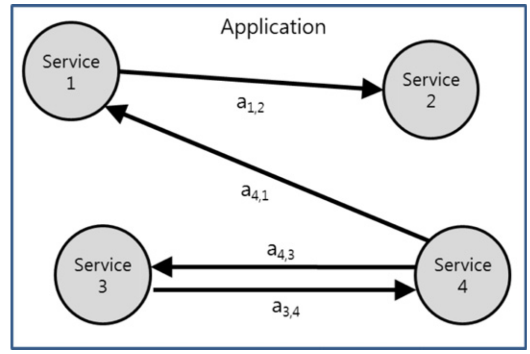

Fig. 2. Example of a user application with $M = 4$ services, that exchange data (black arrows).

Another approach is the one of Kessaci et al. [19]. They optimize the scheduling of tasks over a geographically distributed data centers. Their algorithm considers three objectives for the optimization, namely energy consumption, CO2 emission, and profit.

Bjorkqvist et al. analyze the total cost and performance of running services on hybrid clouds [17], based on the cost model of [20]. Their focus is on a framework for allocating services to clouds, considering the performance–cost tradeoff between nodes of private clouds and nodes of public clouds.

Tran and Agoulmine proposed an algorithm for service placement that considers the network topology, resource availability, and customer demand [44]. Therefore, Tran and Agoulmine's solution can deal with changes in the network environment. In addition to this, the efficiency of the new locations for the services versus the necessary modifications, which have to be made to obtain them, is compared.

Van den Bossche et al. present service placement algorithms to cost-efficiently schedule deadline-constrained bag-of-tasks applications on hybrid clouds. Their algorithms use computation-related cost, data transfer costs, different pricing models, and network capacity limitations [18].

Bittencourt and Madeira propose an algorithm for hybrid cloud customers to decide which of the services should run on the public cloud and which on the private cloud [15]. Their algorithm considers budget aspects and service requirements.

3. A comprehensive cost model for federated hybrid clouds

3.1. User application model

Our user application model assumes a service-oriented architecture. A user application, which represents a business process, comprises the execution of M services. Services are assumed to be basic units, which can communicate with each other by transferring a specific amount of input data and output data. Each service runs on a VM. An example of this user application model is shown in Fig. 2.

The representation of services and the data transfer relationships between them through graph theory allows the use of network analysis methods. The resulting graph of an application is a directed and weighted graph. Vertices represent services, while edges show data communications. That means, the average monthly data transfer a_{ij} from service i to j, is represented through an edge from vertex i to vertex j with a_{ij} as edge label.

3.2. Cost functions

To simplify the representation of the overall cost function, the following four subsections define parts of the total cost function. The total cost function is presented in Section 3.2.5.

3.2.1. Traffic-related cost function

For expressing the traffic-related cost function, the variables *SP, DTM,* and *TCR* are introduced. They are defined as follows:

1. *Service Placement Vector (SP)*: The service placement vector is a vector with M elements, which identifies the clouds that host the M services. For example, the service placement vector $SP = (0, 0, 1)$ represents that service 1 and service 2 are executed on cloud 0, while service 3 is executed on cloud 1.
2. *Data Traffic Matrix (DTM)*: This asymmetric $M \times M$-dimensional matrix shows the data transfer between all M services. The matrix elements denote the traffic a_{ij} from service i to service j. The elements a_{ij} of *DTM*, which do not represent an arrow in the corresponding graph, get assigned *0* (Fig. 2). Depending on the required accuracy, the average data transfer or the exact amount of data transfer might be used.
3. *Traffic Cost Rate Matrix (TCR)*: The traffic cost matrix is a $P \times P-$ dimensional matrix, in which the cost rate per unit traffic (l, k) from cloud l to cloud k is shown. That means, it is the sum of the cost for one unit outgoing traffic from cloud i to cloud j and the cost for one unit incoming traffic to cloud j from cloud i. As clouds usually charge only for in-bound and out-bound traffic of a service, all main diagonal elements of TCR are zero (i.e., there is no charge for traffic that stays within a cloud). However, if a cloud also charges users for their internal data traffic, the corresponding main diagonal element will be unequal zero. Considering the current practice, the TCR matrix is symmetrical.

Based on the above variables, we propose the following equation to capture the total cost of traffic, T^{IO}, that is incurred by inbound data traffic and outbound data traffic between clouds within time period m (f3, f4, and f6 of Table 2). The equation represents a volume-based pricing scheme:

$$T_m^{IO} = \sum_{i=1}^{M} \sum_{j=1}^{M} DTM(i, j) * TCR(SP(i), SP(j)). \tag{1}$$

DTM(i, j) represents the element of the data traffic matrix that indicates the traffic from service i to service j. *SP(i)* and *SP(j)* determine the clouds, on which services i and j run, respectively. Besides, $TCR(SP(i), SP(j))$ represents the traffic cost rate charged by cloud *SP(i)* and cloud *SP(j)* that host services i and j, respectively.

For our calculation (Eq. (1)), we considered a volume-based pricing scheme, in which the amount of traffic is multiplied with a specific price per unit traffic. However, this formula can easily be adapted to other pricing schemes that are common in the market [47]. Examples of those pricing schemes are tiered pricing schemes, volume-based pricing scheme with buy-out option, and 95 percentile pricing scheme [48].

3.2.2. Deployment-related cost function

To capture the deployment cost that is incurred by changes in the placement of services, we propose the *Deployment Cost Vector (DCV)*, which captures the average cost per deployment for each cloud. Deployment costs are incurred when a virtual machine with the service is migrated from one cloud to another. During the duration of the migration, the resources cannot be used for any task. The deployment cost vector holds the cost of each cloud in Euro. For example, the cost could be calculated by multiplying the price of a virtual machine of a cloud with the time that a VM cannot be used during the migration. The cost should also consider the amount of data that needs to be transferred to the new public cloud.

The overall deployment costs are incurred at the source cloud (i.e., the cloud, from which the service is migrated) and at the destination cloud (i.e., the cloud, to which the service is migrated to).

Following this definition of deployment cost, the cost factor *number of deployments* (i.e., the number of installations and un-installations) for each cloud i can be used. For this, the number of deployments, which are needed for a migration from the current service placement to a new service placement, needs to be counted. The number of deployments for each cloud is listed in vector Q^d. Following a simple cost function, the total deployment cost C_m^D for the time period m can be calculated by multiplying the number of deployments Q^d for the time period m per cloud with the deployment cost vector DCV:

$$C_m^D = Q^d * DCV. \tag{2}$$

By considering this cost factor, an optimization algorithm (based on this cost model) can capture the trade-off between incurring some cost due to changes in the service placement and the gains through an optimal service placement.

3.2.3. Fixed cost function

Based on the cost factors listed in Section 2.4, we propose a comprehensive formula that covers all aspects of fixed costs. The costs considered are the purchasing cost for all servers needed ($\sum_i^S C_i^{Se}$; b1 of Table 2), the purchasing cost for all network devices needed ($\sum_i^N C_i^{Ne}$; b2 of Table 2), the costs for all basic server software licenses ($\sum_i^B C_i^{BSS}$; c1 of Table 2), the costs for middleware licenses ($\sum_i^K C_i^{MS}$; c2 of Table 2), the costs for application software licenses ($\sum_i^A C_i^{AS}$; c3 of Table 2), the cost of facility space ($C_{Fa}F$; e3 of Table 2) that the company needs for its data center, the cost of non-electronic equipment (C_{Nee}; e1 of Table 2) that is needed for a data center, and the cost for cabling (C_{Ca}; e2 of Table 2). These costs are fixed costs for the duration of the depreciation period. The sum of these fixed costs *FC* is shown in Eq. (3):

$$FC = \sum_i^S C_i^{Se} + \sum_i^N C_i^{Ne} + \sum_i^B C_i^{BSS} + \sum_i^K C_i^{MS} + \sum_i^A C_i^{AS} + C_{Fa}F + C_{Nee} + C_{Ca}. \tag{3}$$

The parameters used in Eq. (3) are as follows: S is the number of servers, which are needed to run services. S is fixed. This assumption of a fixed number of servers S is justified, if a private cloud is used to run mission-critical services, for which the number of needed servers can be estimated beforehand. Assuming a permanently cost-minimizing utilization of servers, economies of scale also works for large cloud customers as for cloud providers. In this case, the number of servers S can also be calculated beforehand for large cloud customers. Therefore, the number of servers S calculated can be assumed to be the most economical solution. C_i^{Se} is the cost of server i. N is the number of network devices that are needed. C_i^{Ne} is the cost of network device i. B denotes the number of basic server software licenses. C_i^{BSS} is the cost of a basic server software license i. K denotes the number of middleware software licenses, which are needed. C_i^{MS} is the cost of middleware software license i. A is the number of application software licenses, which are used. C_i^{AS} is the cost of application software license i. F denotes the size of the facility in square meters. C_{Fa} is the cost of facility space in square meters. The facility cost can easily be obtained from rent contracts.

As the fixed cost needs to be adjusted with respect to the depreciation period of each cost factor, each cost item is multiplied with a certain factor (e.g., D_{IT}, D_F, D_U). Considering that the depreciation period is $D_{IT} = 36$ months for IT devices, $D_F = 600$ months for facilities, $D_U = 120$ months for cabling and utilities, the fixed cost per one time period, FC^D, can be calculated as:

$$FC^D = \frac{1}{D_{IT}}\left(\sum_i^S C_i^{Se} + \sum_i^N C_i^{Ne} + \sum_i^B C_i^{BSS} + \sum_i^K C_i^{MS} + \sum_i^A C_i^{AS}\right) + \frac{1}{D_F}(C_{Fa}F) + \frac{1}{D_U}(C_{Nee} + C_{Ca}). \quad (4)$$

3.2.4. Variable cost function

The following variable cost factors, which are calculated over a certain time period, are usually summed up for a time period of a month or a year. For this, simple usage-based pricing schemes are assumed. However, if needed, the simple usage-based pricing schemes can be replaced with more complicated pricing schemes easily. They comprise the cost of electricity usage from cooling and from other electric devices ($C_m^{Elec}\left(E_m^C + E_m^{EIdle} + E_m^{EUse}\right)$; a1, a2, and a3 of Table 2), the cost of Internet connectivity of the private cloud ($C_m^{Int} I_m$; f1 of Table 2), the cost of labor for maintaining software ($C_m^{LS} L_m^S$; d1 of Table 2), for maintaining hardware ($C_m^{LH} L_m^H$; d2 of Table 2), for other work ($C_m^{LO} L_m^O$; d3 of Table 2), the cost of deploying services (C_m^D; g1 of Table 2), the cost of data transfer between different clouds (T_m^{IO}; Eq. (1); f3, f4, and f6 of Table 2), the cost for cloud storage ($C_m^{Sto} H_m$; f5 of Table 2), and the cloud server usage cost ($\sum_i^{SerType} C_{m,i}^{Ser} S_m^i$; f2 of Table 2). The sum of all these cost factors, as shown in the following equation, determines the total variable cost $VC(m)$ in period m:

$$VC(m) = C_m^{Elec}\left(E_m^C + E_m^{EIdle} + E_m^{EUse}\right) + C_m^{Int} I_m + C_m^{LS} L_m^S + C_m^{LH} L_m^H + C_m^{LO} L_m^O + C_m^D + T_m^{IO} + C_m^{Sto} H_m + \sum_i^{SerType} C_{m,i}^{Ser} S_m^i. \quad (5)$$

The parameters used in Eq. (5) are as follows: C_m^{Elec} is the cost of electricity per unit in time period m (usually, the time period is one month). E_m^C specifies the amount of electricity used for cooling in time period m. E_m^{EIdle} specifies the amount of electricity used for all other devices in time period m when they were idle. Similarly, E_m^{EUse} specifies the amount of electricity used for all other devices in time period m when they were used. C_m^{Int} is the cost for Internet connectivity usage per unit. I_m denotes the amount of use of the Internet connectivity in time period m. (Note: the Internet cost calculation should be adjusted to the Internet contract, i.e., the pricing scheme used.) C_m^{LS} is the cost of labor for maintaining software per unit (usually, the unit is one hour) and L_m^S is amount of labor in time period m. C_m^{LH} is the cost of labor for maintaining the hardware per unit and L_m^H is the amount of labor for maintaining hardware in time period m. C_m^{LO} is the cost of labor for other tasks per unit and L_m^O is the amount of this labor in time period m. C_m^D is the cost of deployment to clouds (Eq. (2)) within time period m. T_m^{IO} is the total cost of traffic between clouds in month m (Eq. (1)). C_m^{Sto} is the cost of cloud storage per unit in time period m. H_m is the usage of storage in time period m. Note, if different types of storage (with respect to the storage access speed) are used and priced differently, the pricing scheme $C_m^{Sto} H_m$ need to be replaced with a quality-of-service based pricing scheme. $C_{m,i}^{Ser}$ is the cloud server usage cost for server type i per unit in time period m. S_m^i is the amount of server usage of type i in time period m.

3.2.5. Total cost function

The total cost of running a federated hybrid cloud over d time periods is the sum of the fixed cost, FC^D, and the variable cost, VC(), which is incurred during these d time periods. The total cost is shown in Eq. (6):

$$TC(d) = d\, FC^D + \sum_{m=1}^{d} VC\,(m)\,. \quad (6)$$

If only one time period, $d = 1$, is considered, then the total cost TC (1) represents the cost that is occurred within a single time period.

3.3. Discussion and comparison

All eight terms of the fixed cost formula FC (Eq. (4)) can easily be populated, since it requires only capital expenditure (CAPEX) information. This information can be obtained from vendors or IT consultancy companies that design data centers for enterprises. For calculating the variable cost VC (Eq. (5)), the expected base utilization and the peak resource utilization need to be estimated. This is difficult, as the services deployed may have highly variable peaks in resource demand, the practical limits on real-world utilization of purchased equipment are not widely known, and the operational costs differ widely, depending on the type of cloud environment [7]. In addition to this, estimating the variable cost factors on data transfer, cloud storage usage, and server usage requires further effort. For example, gathering the real trend of data transfer among services, one needs special preparation to monitor, aggregate, and keep data transfer records. Then, econometric methods need to be applied to detect trends of data transfers between services. Understanding peaks, which force companies to increase cloud expenditure, is a key to get good estimates of future costs.

Furthermore, the importance of each cost factor should be considered when estimating the total cost. Accurately focusing on major cost factors is suggested. Any error in their estimates has a large impact on the accuracy of the total cost estimation. For instance, in many cases, 31% of cost of data centers comes from labor, 30% from servers, and 25% from cooling [40]. Therefore, we indicated in Table 2 the expected impact of each cost factor for private clouds, public clouds, and federated clouds. This information helps users of the cost model to identify the cost factors that are worth finding accurate data about them.

Finally, the major difference of the presented cost model to existing works on cost model comprises three items: (1) The presented cost model can be applied to any type of cloud (i.e., private cloud, public cloud, federated cloud, and federated hybrid cloud), as it comprises all cost factors that have been considered for clouds; (2) The cost model considers data transfer of services between federated clouds, which has not been discussed in literature on federated clouds. The consequence of this cost factor is that, if the cost for data transfer between clouds gets too large, any federation of clouds will not be economically efficient; and (3) The cost model also considers the deployment cost as a cost factor. The deployment cost limits the number of optimizations that can be performed to lower the cost. Any migration of a service to another cloud comes with the cost of terminating the service on one cloud and moving it via the Internet to the new cloud, where it has to be installed. Although some of the tasks can be performed in parallel and, therefore, can be performed with low time cost, this process incurs cost for the virtual machines that need to be terminated and installed before they can start executing the services and the cost for transferring the checkpoint data of the service from one cloud to another cloud via the Internet.

```
1:  ***initialize all input variables***
2:  Set trafficCostRateMatrix; ***prices for traffic between clouds***
3:  Set dataTrafficMatrix; ***data transfer between all M services***
4:  Set deploymentCostVector; ***average cost of deployment per cloud***
5:  Set numberOfDeploymentsVector; ***variables listed in Table 2***
6:  ***initialize output variables***
7:  lowest_cost = max_real_number;
8:  ***check all service placement options***
9:  WHILE (anotherServicePlacementVectorPossible) {
10:       ***generate one service placement option***
11:       Generate servicePlacementVector;
12:       ***calculate cost formulas***
13:       Calculate dataTransferCost; ***equation 1***
14:       Calculate deploymentCost; ***equation 2***
15:       Calculate fixedCost; ***equation 4***
16:       Calculate variableCost; ***equation 5***
17:       Calculate totalCost; ***equation 6***
18:       IF (totalCost < lowestCost) THEN {
19:              lowestCost = totalCost;
20:              lowestServicePlacementVector = servicePlacementVector;
21:       }
22: }
23: ***return the best service placement option as solution***
24: RETURN lowestServicePlacementVector;
```

Fig. 3. Cost model based service placement optimization algorithm (COMBSPO Algorithm).

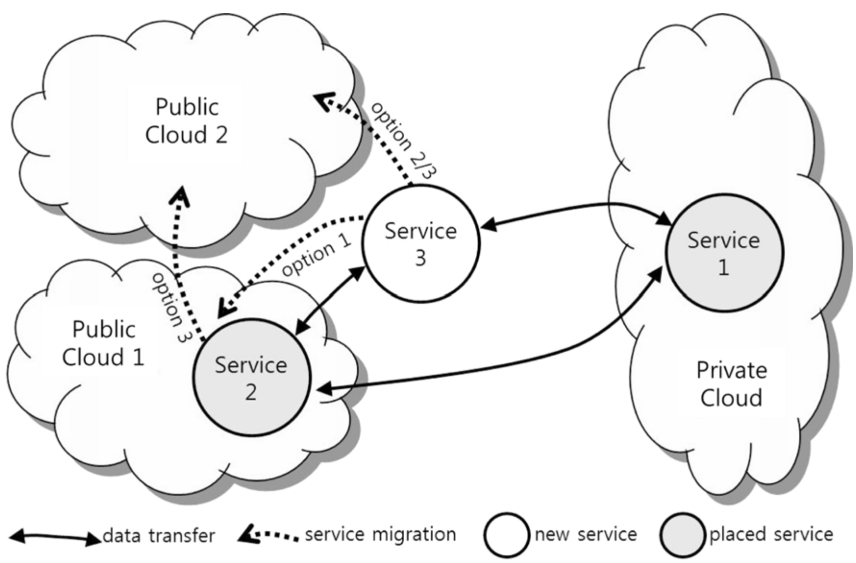

Fig. 4. Example of a service placement decision: it is assumed that, initially, service 1 and service 2 run on the private cloud and public cloud 1, respectively. When the demand for service 3 emerges, the company has to decide on a new service placement. There are three service placement options: service 2 and service 3 run on public cloud 1 (option 1), service 2 and service 3 run on public cloud 2 (option 3), or service 2 and service 3 run on public cloud 1 and public cloud 2, respectively (option 2).

4. Cost model based service placement optimization

4.1. Service placement optimization algorithm

Finding the optimal service placement requires the calculation of the total cost for each possible service placement option on a private cloud and a number of federated public clouds. To achieve this, our cost model based service placement optimization algorithm, which is called COMBSPO, initializes the main input variables (Section 2.4) and, then, executes a loop over all possible service placements. Within the loop, the calculation of the total cost (Eq. (6)) is performed based on our cost model (Section 3.2.5). At the end of the loop, the cost of the newly calculated service placement options is compared with the currently lowest cost one. If the new one is lower, it will become the new lowest cost service placement option. The algorithm terminates, after the set of all relevant service placement options have been checked. Note, the set of relevant service placement options could comprise all possible service placements, which is a brute-force approach and not scalable, or could be selected based on a heuristic (line 9 of the service placement optimization algorithm (Fig. 3)). However, for demonstrating the workings of our COMBSPO algorithm, the selection of the service placement option is not of importance. The detailed service placement optimization algorithm is shown in Fig. 3.

It should be noted though that a corporate's policy (e.g., security policy) might invalid a few service placement options. Consequently, the number of eligible service placement options will be reduced.

4.2. Scenario and data set

To demonstrate the workings of the cost model and the service placement algorithm, we define a small scenario. The scenario comprises a cloud customer with demand for 1–3 VMs (i.e., 1–3 services). As the cloud customer has always a minimum demand for one VM (e.g., to run mission-critical applications), the company runs one VM (service 1 of Fig. 4) on its private cloud. All other

Table 3
The minimum and maximum cost factor values used for the scenario analysis are based on Amazon AWS in December 2013 [51].

Cost type	Cost factor	Basis of cost estimation	Cost (€ / Month)	
			Minimum	Maximum
	(f2) Cloud server use	Historical data and market price	51 €/Month	1728 €/Month
(f) Cloud service	(f3) Data transfer into cloud	Historical data and market price	0 €/GB	0 €/GB
	(f4) Data transfer from cloud	Historical data and market price	0.07 €/GB	0.12 €/GB
	(f5) Cloud storage use	Historical data and market price	0.065 €/GB/Month	0.065 €/GB/Month
	(f7) Data transfer between clouds	Historical data and market price	0.02 €/GB to 0.10 € /GB	0.12 €/GB
(g) Deployment	(g1) Number of deployments	Historical data and market data	0.06 €/Deployment (i.e., 1 h server use)	14.40 €/Deployment (i.e., 1 h server use and 100 GB data transfer)

services need to run on federated clouds, if the services are demanded. For simplification of the scenario, we assume that all services require the same amount of computational resources, i.e., one VM of the same type. Furthermore, two clouds are assumed to exist within the cloud federation at two geographically different locations (Fig. 4). Furthermore, it is assumed that, initially, service 2 runs on the public cloud 1.

When the demand for service 3 emerges (Fig. 4), the company has to decide on a new service placement. For our scenario, we assume further that an extension of the private cloud is not considered for strategic business reasons of the cloud customer. Without this assumption, however, an extension of the private cloud would be a valid option that needs to be analyzed.

Therefore, in our scenario, there are only three service placement options to be considered: Service 2 and service 3 run on public cloud 1 (option 1); Service 2 runs on public cloud 1 and service 3 runs on public cloud 2 (option 2); Service 2 and service 3 run on public cloud 2 (option 3). Service 1 is kept on the private cloud due to security reasons.

To decide on the cost-minimizing service placement, the cloud customer has to calculate the total cost for each of the three options. As the company of our scenario owns its data center, the values for the fixed cost factors (Eq. (4)) can easily be obtained from the accounting department [49]. The values for the variable cost factors (Eq. (5)) need to be estimated using the past consumption and the price developments in the market. It is to be noted that, as there are many different pricing schemes (e.g., spot market prices, reservation prices, flat rate pricing, tiered pricing), finding the best fitting pricing scheme is time-consuming [50]. It requires comparing the potential overall cost of all pricing schemes for a given cloud server use. For our scenario calculation, we assume that this comparison has been conducted.

4.3. Scenario analysis using the COMBSPO service placement algorithm

In order to populate the trafficCostRateMatrix (*TCR* in Section 3.2.1) and the deploymentCostVector (*DCV* in Section 3.2.2) of the COMBSPO service placement algorithm (Fig. 3) with realistic values, we draw uniformly values from the ranges of cost factor values that are specified in Table 3. The range of values is given through the minimum and maximum cost factors values. The minimum and maximum values are obtained from Amazon AWS in 2013 [51]. Although the data comes from a single provider only, the data is sufficient for determining a realistic order of magnitude of values.

With respect to populating the number Of Deployment Vector (Q^d in Section 3.2.2), we assume that the emergence of the demand for service 3 is a onetime event. Only option 3 needs to consider the cost for migrating service 2 from public cloud 1 to public cloud 2.

In detail, the total cost of option 1 is based on a data transfer price of 0.00 Euro/GB between service 2 and service 3, 0.09 Euro/GB for data transfer to and from service 1, and on the cost of two VMs of 172 Euro/month (public cloud 1).

The total cost of option 2 is based on a data transfer price of 0.09 Euro/GB between service 2 and service 1, a data transfer price of 0.11 Euro/GB between service 2 and service 3, a data transfer price of 0.13 Euro/GB between service 3 and service 1, the cost of a VM of 86 Euro/month (public cloud 1), and the cost of a VM of 83 Euro/month (public cloud 2).

The total cost of option 3 is based on a data transfer price of 0.00 Euro/GB between service 2 and service 3, 0.13 Euro/GB for data transfer to and from service 1, and the cost of two VMs of 83 Euro/month (public cloud 2). In addition to this, option 3 has to take into consideration the deployment cost for migrating service 2 from public cloud 1 to public cloud 2. The deployment cost, which considers data transfer cost and unusable time of VM is assumed to be 3.5 Euro per deployment.

Furthermore, it is assumed that the amount of data transfer is the same not only for each option but also between the three services (service 1, service 2, and service 3). However, the exact amount of data transfer has not been set by the cloud customer yet. The amount of data transfer depends on the size of the data set processed by a service. Therefore, a sensitivity analysis is conducted to determine the impact of the size of the data set on the overall cost of using federated hybrid clouds.

For this sensitivity analysis, the dataTrafficMatrix (*DTM* in Section 3.2.1) is populated with different data transfer values, ranging from 0 GB to 150 GB in total. For each of the populated dataTrafficMatrix, the COMBSPO service placement algorithm is executed. The results of all COMBSPO calculations are shown in Fig. 5.

The results of the sensitivity analysis show that option 2 is better than option 1 and option 3 (Fig. 5), if the amount of data transfer is less than 25 GB (in total). Option 3 is better than option 1 and option 2, if the amount of data transfer is between 25 GB and 95 GB. If the amount of data transfer is above 95 GB, option 1 will be better than the other two options.

These results demonstrate two issues that need to be considered closely in federated clouds. First, the deployment cost has a strong impact on the optimal service placement. As can be seen from Fig. 5, option 2 is the preferred option for a data transfer amount of less than 30 GB due to the 3.5 Euro deployment cost that is incurred in option 3. Without this cost factor, option 3 would have always been less costly than option 2. Second, the data transfer cost is a significant cost factor in federated clouds. Fig. 5 shows that the low price of servers in option 3 becomes irrelevant, if the amount of data transfer is larger than 95 GB. At this point, the data

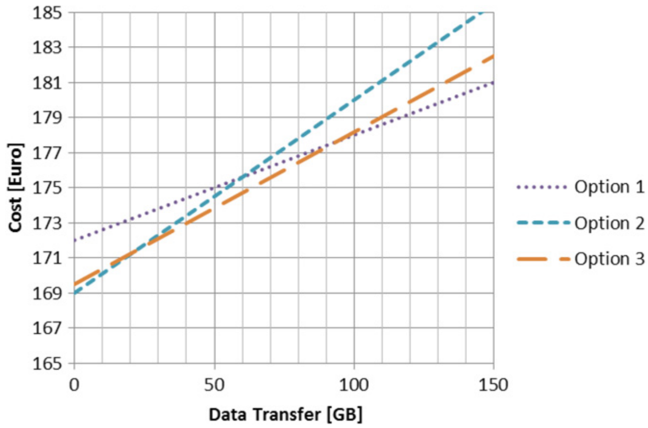

Fig. 5. Sensitivity analysis of the three service placement options for the scenario shown in Fig. 4. For the comparison, the COMBSPO algorithm is executed with different data traffic matrices. The total traffic is varied from 0 to 150 GB.

transfer price, which is higher than the data transfer price of option 1, becomes dominant.

This analysis does not only demonstrate the workings of the COMBSPO service placement algorithm but also illustrates that the COMBSPO service placement algorithm helps understanding the dependencies between different cost factors of federated hybrid clouds. It has shown that the deployment cost can offset data transfer cost. Consequently, it can be stated that an increase in the number of service migrations can have a substantial negative impact on the overall benefit of cost-minimizing service placements. In this case, a cost-minimizing service placement option would not be implemented.

5. Conclusion and future work

In this paper, we have comprehensively analyzed the cost factors that need to be considered for estimating cloud computing cost of federated hybrid clouds, i.e., the most general cloud architecture. As part of this work, the cost factors have been evaluated and categorized. This analysis has also been the basis for our cloud cost model, which can be used by any cloud customer to estimate the total cost of clouds (i.e., private clouds, public clouds, hybrid clouds, distributed clouds, federated clouds, or federated hybrid clouds). The cost model defines the cost through cost formulas (i.e., a fixed cost formula, a variable cost formula, and a total cost formula).

Based on this cost model, we also presented a cost model based service placement optimization algorithm, called COMBSPO. It compares the total cost of all possible service placement options and identifies the cost-minimizing service placement option. Compared to existing service placement algorithms, the major difference of this service placement algorithm is that it considers the data transfer cost between public clouds and the deployment cost of services, which are incurred through VM migrations.

Furthermore, our analysis of a federated hybrid clouds scenario demonstrates the workings and the benefits of our service placement algorithm. It shows that COMBSPO not only helps companies in finding the cost-minimizing service placement option in federated hybrid clouds but also helps understanding the dependencies between different cost factors of federated hybrid clouds. It has shown that the deployment cost can offset data transfer cost. Consequently, we can state that an increase in the number of service migrations can have a substantial negative impact on the overall benefit of cost-minimizing service placements. In this case, a cost-minimizing service placement option would not be implemented.

In our future work, we will address the limitations of our current work. Besides investigating an extension of the service placement algorithm with respect to performance factors, we will also look into the application of the proposed service placement algorithm to a more realistic scenario with respect to the number of clouds and services considered. Finally, we will investigate the differences in cost calculated with COMBSPO and other cost models.

Acknowledgment

This work has been funded by the Korea Institute for Advancement of Technology (grant no. ITEA2 10014) as part of the ITEA2 project EASI-CLOUDS.

References

[1] B.C. Tak, B. Urgaonkar, A. Sivasubramaniam, To move or not to move: The economics of cloud computing. HOTCLOUD 2011, Third USENIX Workshop on Hot Topics in Cloud Computing. 2011.
[2] M. Klems, J. Nimis, S. Tai, Do clouds compute? A framework for estimating the value of cloud computing. Designing E-Business Systems. Markets, Services, and Networks, Lecture Notes in Business Information Processing, vol. 22, pp. 110-123. 2009.
[3] A. Bany Mohammed, J. Altmann, J. Hwang, Cloud computing value chains: understanding businesses and value creation in the cloud, Econ. Models Algorithms Distrib. Syst. (2010) 187–208.
[4] D. Kondo, B. Javadi, P. Malecot, F. Cappello, D.P. Anderson, Cost-benefit analysis of cloud computing versus desktop grids. IPDPS 2009, IEEE International Symposium on Parallel & Distributed Processing. May 2009.
[5] P. Mell, T. Grance, The NIST definition of cloud computing, Natl. Inst. Stand. Technol. 53 (6) (2009) 50.
[6] C.D. Patel, A.J. Shah, Cost Model for Planning, Development and Operation of a Data Center. HPL-2005-107(R.1). June 2005.
[7] A. Cole, The future belongs to hybrid clouds. IT Business Edge. September 2011, http://www.itbusinessedge.com/cm/blogs/cole/the-future-belongs-to-hybrid-clouds/?cs=30482, 2013 (accessed on December 2013).
[8] G. Orenstein, Forecast for 2010: The rise of hybrid clouds. GIGAOM. January 2010. http://gigaom.com/2010/01/01/on-the-rise-of-hybrid-clouds, 2013 (accessed on December 2013).
[9] A. Cole, Hybrid Clouds Allow Organizations to 'Own the Base and Rent the Peak'. IT Business Edge. July 2011. http://www.itbusinessedge.com/cm/community/features/interviews/blog/hybrid-clouds-allow-organizations-to-own-the-base-and-rent-the-peak/?cs=47971, 2013 (accessed on December 2013).
[10] B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I.M. Llorente , R. Montero, Y. Wolfsthal, E. Elmroth, J. Caceres, M. Ben-Yehuda, W. Emmerich, F. Galan, The reservoir model and architecture for open federated cloud computing, IBM J. Res. Dev. 53 (4) (2009) 4:1–4:11.
[11] E. Elmroth, L. Larsson, Interfaces for placement, migration, and monitoring of virtual machines in federated clouds, in: Eighth International Conference on Grid and Cooperative Computing, GCC 2009, IEEE, 2009, pp. 253–260.
[12] E. Deelman, G. Singh, M. Livny, B. Berriman, J. Good, The cost of doing science on the cloud: The montage example. SC 2008, International Conference for High Performance Computing, Networking, Storage and Analysis. pp.1-12. November 2008.
[13] D. D'Agostino, A. Galizia, A. Clematis, M. Mangini, I. Porro, A. Quarati, A QoS-aware broker for hybrid clouds, J. Comput. 95 (1) (2013) 89–109. Springer.
[14] S. Pandey, A. Barker, K.K. Gupta, R. Buyya, Minimizing execution costs when using globally distributed cloud services. AIDA 2010, 24th IEEE International Conference on Advanced Information Networking and Applications (AINA). pp. 222–229. April 2010.
[15] L. Bittencourt, E. Madeira, HCOC: A cost optimization algorithm for workflow scheduling in hybrid clouds, J. Internet Serv. Appl. 2 (3) (2011) 207–227.
[16] D.M. West, Saving money through cloud computing, Governance Studies, Brookings (2010).
[17] M. Bjorkqvist, L.Y. Chen, W. Binder, Cost-driven service provisioning in hybrid clouds. SOCA 2012, 5th International Conference on Service-Oriented Computing and Applications, 2012.
[18] R. Van den Bossche, K. Vanmechelen, J. Broeckhove, Online cost-efficient scheduling of deadline-constrained workloads on hybrid clouds, Future Gener. Comput. Syst. 29 (4) (2013) 973–985.
[19] Y. Kessaci, N. Melab, E.-G. Talbi, A Pareto-based metaheuristic for scheduling HPC applications on a geographically distributed cloud federation, J. Cluster Comput. 16 (3) (2013) 451–468.
[20] M. Kashef, J. Altmann, A cost model for hybrid clouds, in: 9th International Workshop on Economics of Grids, Clouds, Systems, and Services, GECON 2012, Springer, 2012, pp. 46–60.
[21] J. Metzler, S. Taylor, Cloud Computing: Reality vs. Fiction, Network World (2008) http://www.networkworld.com/article/2216572/lan-wan/cloud-computing–reality-vs–fiction.html, 2014 (accessed on August 2014).

[22] M. Maurer, V.C. Emeakaroha, M. Risch, I. Brandic, J. Altmann, Cost and benefit of the SLA mapping approach for defining standardized goods in cloud computing markets, in: UCC 2010, International Conference on Utility and Cloud Computing (2010).
[23] J. Altmann, C. Courcoubetis, M. Risch, A marketplace and its market mechanism for trading commoditized computing resources, Ann. Telecommun. 65 (11–12) (2010) 653–667.
[24] J.F. Rayport, A. Heyward, Envisioning the cloud: the next computing paradigm and its implication for technology policy. White Paper. Marketspace LLC. March 2009, http://marketspacenext.files.wordpress.com/2011/01/envisioning_the_cloud_presentation-deck.pdf, 2013 (accessed on December 2013).
[25] J. Weinman, Mathematical proof of the inevitability of cloud computing, White Paper, January 2011, http://www.joeweinman.com/Resources/Joe_Weinman_Inevitability_Of_Cloud.pdf, 2013 (accessed on December 2013).
[26] H.L. Truong, S. Dustdar, Composable cost estimation and monitoring for computational applications in cloud computing environments. ICCS 2010, International Conference on Computational Science. Procedia Computer Science. vol. 1, no. 1, pp. 2175–2184.
[27] T. Alford, G. Morton, The economics of cloud computing: Addressing the benefits of infrastructure in the cloud. White Paper. Booz Allen and Hamilton, October 2009.
http://www.boozallen.com/media/file/Economics-of-Cloud-Computing.pdf, 2013 (accessed on December 2013).
[28] A. Khajeh-Hosseini, I. Sommerville, I. Sriram, Research challenges for enterprise cloud computing. arXiv:1001.3257, January 2010.
[29] J. Koomey, K. Brill, P. Turner, J. Stanley, B. Taylor, A simple model for determining true total cost of ownership for data centers. White Paper. Uptime Institute. 2007.
http://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/%28TUI3011B%29SimpleModelDetermingTrueTCO.pdf, 2013 (accessed on December 2013).
[30] A. Khajeh-Hosseini, D. Greenwood, J.W. Smith, I. Sommerville, The cloud adoption toolkit: supporting cloud adoption decisions in the enterprise, J. Softw. Pract. Exp. 42 (4) (2012) 447–465.
[31] A. Khajeh-Hosseini, D. Greenwood, I. Sommerville, Cloud migration: A case study of migrating an enterprise IT system to IaaS, in: Computing Research Repository (CoRR), ACM, 2010.
[32] K. Buell, J. Collofello, Transaction level economics of cloud applications. SERVICES 2011, IEEE World Congress on Services, 2011.
[33] R.-H. Hwang, C.-N. Lee, Y.-R. Chen, D.-J. Zhang-Jian, Cost optimization of elasticity cloud resource subscription policy, IEEE Trans. Serv. Comput. (2013) http://doi.ieeecomputersociety.org/10.1109/TSC.2013.35.
[34] C. Chen, B. He, A framework for analyzing monetary cost of database systems in the cloud, in: Web-Age Information Management, LNCS 7923, Springer, 2013, pp. 118–129.
[35] M.E. Aldana Diaz, E.N. Huh, An Efficient IPTV Cost-Based Policy Management of Hosting Service for 3rd Party Content. CUTE 2010, 5th International Conference on Ubiquitous Information Technologies and Applications, 2010.
[36] M.E. Aldana Diaz, E.N. Huh, Cost analysis on IPTV hosting service for 3rd party providers. ICUIMC 2011, 5th International Conference on Ubiquitous Information Management and Communication, 2011.
[37] M. Risch, J. Altmann, Cost analysis of current grids and its implications for future grid markets. GECON 2008, 5th International Workshop on Grid Economics and Business Models. pp. 13-27. 2008.
[38] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, M. Zaharia, Above the clouds: A Berkeley view of cloud computing, University of California, Berkeley, 2009.
[39] M. Hajjat, X. Sun, Y.-W.E. Sung, D.A. Maltz, S. Rao, K. Sripanidkulchai, M. Tawarmalani, Cloudward Bound: Planning for Beneficial Migration of Enterprise Applications to the Cloud, ACM, 2010.
[40] A. Opitz, H. König, S. Szamlewska, What does grid computing cost? J. Grid Comput. 6 (4) (2008) 385–397.
[41] J. Altmann, J. Rohitratana, Software Resource Management Considering the Interrelation between Explicit Cost, Energy Consumption, and Implicit Cost. MKWI 2010, Multikonferenz Wirtschaftsinformatik, pp. 327–339. 2010.
[42] Amazon AWS., Simple monthly calculator. 2013
http://calculator.s3.amazonaws.com/calc5.html, 2013 (accessed on December 2013).
[43] J. Altmann, J. Rohitratana, Software resource management considering the interrelation between explicit cost, energy consumption, and implicit cost, Future Generation Computing Systems (2010) 327–339.
[44] K.-T. Tran, N. Agoulmine, Adaptive and cost-effective service placement, in: Global Telecommunications Conference, GLOBECOM 2011, IEEE, 2011.
[45] F. Koch, M. Assunção, M.S. Netto, A cost analysis of cloud computing for education, in: 9th International Conference on Economics of Grids, Clouds, Systems, and Services, GECON 2012, Springer, 2012, pp. 182–196.
[46] M.Q. Dang, J. Altmann, Resource allocation algorithm for light communication grid-based workflows within an SLA context, International Journal of Parallel, Emergent and Distributed Systems 24 (1) (2009) 31–48.
[47] J. Altmann, L. Rhodes, Dynamic Netvalue analyzer —a pricing plan modeling tool for ISPs using actual network usage data, in: International Workshop on Advance Issues of E-Commerce and Web-Based Information Systems, IEEE WECWIS 2002, IEEE, 2002.
[48] J. Altmann, K. Chu, A Proposal for a Flexible Service Plan that is Attractive to Users and Internet Service Providers. IEEE InfoCom2001, Conference on Computer Communications, Anchorage, Alaska, USA, April 2001.
[49] S. Hernandez, J. Fabra, P. Alvarez, J. Ezpeleta, Cost Evaluation of Migrating a Computation Intensive Problem from Clusters to Cloud. in: Altmann, J. Vanmechelen, K., Rana, O., GECON 2013, Intl. Conference on Economics of Grids, Clouds, Systems, and Services, Zaragoza, Spain, September 2013.
[50] R. Riolo, J.H. Moore, M. Kotanchek, Genetic programming theory and practice XI, Series: Genetic and Evolutionary Computation (2014).
[51] Amazon AWS., http://calculator.s3.amazonaws.com/calc5.html, 2013 (accessed on December 2013).

Jörn Altmann is Professor for Technology Management, Economics, and Policy at the College of Engineering, Seoul National University. Prior to this, he taught computer networks at UC Berkeley, worked as a Senior Scientist at Hewlett-Packard Labs, and has been a postdoc at EECS and ICSI of UC Berkeley. Dr. Altmann's research centers on Internet economics with a focus on economic analysis of Internet services and on integrating economic models into Internet infrastructures.

Mohammad Mahdi Kashef is a Ph.D. student at the Technology Management, Economics, and Policy Program of the College of Engineering, Seoul National University. His research interest is service placement in clouds.