QoS- and Security-aware Composition of Cloud Collaborations

Olga Wenge, Ulrich Lampe and Ralf Steinmetz
Multimedia Communications Lab (KOM), TU Darmstadt, Rundeturmstr. 10, 64283 Darmstadt, Germany

Keywords: Cloud Computing, Collaboration, Quality of Service, Security.

Abstract: While cloud computing promises virtually unlimited resource supplies, smaller providers may not be able to offer sufficient physical IT capacity to serve large customers. A solution is cloud collaborations, in which multiple providers unite forces in order to conjointly offer capacities in the market. Unfortunately, both the QoS and security properties of such collaborations will be determined by the "weakest link in the chain", hence resulting in a trade-off between the cumulative capacity and the non-functional characteristics of a cloud collaboration. In this position paper, we examine how cloud collaborations can be optimally composed in a QoS- and security-aware fashion within a market scenario involving multiple cloud providers and users. We propose a Mixed Integer Programming-based exact optimization approach named CCCP-EXA.KOM. Based on a quantitative evaluation, we find that the practical applicability of CCCP-EXA.KOM is limited to small-scale problem instances and conclude that the development of tailored heuristic approaches is required.

1 INTRODUCTION

Cloud computing promises to supply virtually unlimited IT capacities in a scalable, pay-as-you-go fashion (Buyya et al., 2009). Yet, specifically smaller providers may not be able to satisfy the resource demands of large customers on their own due to insufficient data center capacity. A solution lies in cloud collaborations, i. e., the cooperation of multiple providers to conjointly satisfy user demands. Unfortunately, such cloud collaborations have both Quality of Service (QoS) and security implications: since a user may potentially be served by any provider within a collaboration, the non-functional service attributes – e. g., availability, latency, or data center location – will be determined by the "weakest link in the chain", i. e., the provider with the lowest guarantees.

Take the example of two providers, one of which uses encryption for data storage and one which does not. Once these providers join forces within a collaboration and act as one common provider, data may be stored at either one of them. Hence, the collaboration *cannot* be assumed to offer data encryption, even though the data may in fact physically reside with the first provider, i. e., the one which applies encryption. A similar problem occurs if two providers reside in different jurisdictions, such as the European Union (EU) and the United States, where data privacy laws substantially differ (Wenge et al., 2012).

Yet, given the wide range of legal and regulatory requirements that apply in many industries, a cloud user faces certain requirements in choosing his/her cloud provider, and these requirements may not be fulfilled once two or more providers join forces within a collaboration. Based on this scenario, we examine the *Cloud Collaboration Composition Problem* (CCCP) in the work at hand. Our focus is on a broker within the cloud market, who aims to maximize his/her profit through the composition of cloud collaborations from a set of providers and assignment of users to these collaborations. In that process, QoS and security requirements should also be satisfied.

This work introduces the CCCP as a new research problem in the context of cloud computing. The paper also presents a formal optimization model, which permits the computation of exact, i. e., profit-optimal, solutions for specific problem instances.

The remainder of this paper is structured as follows: In Section 2, we describe the problem in detail and introduce formal notations. Based on this, the subsequent Section 3 introduces an exact optimization approach, called CCCP-EXA.KOM, which is quantitatively evaluated in Section 4. Section 5 gives an overview of related work, and Section 6 concludes the paper with a summary and outlook.

Wenge O., Lampe U. and Steinmetz R..
QoS- and Security-aware Composition of Cloud Collaborations.
DOI: 10.5220/0004940405780583
In *Proceedings of the 4th International Conference on Cloud Computing and Services Science* (CLOSER-2014), pages 578-583
ISBN: 978-989-758-019-2

2 FORMAL PROBLEM STATEMENT

In our work, we take the perspective of a cloud broker, who is acting within a cloud market. This cloud market consists of a set of cloud providers and a set of users, formally denoted as $P = \{1, 2, \ldots, P^{\#}\}$ and $U = \{1, 2, \ldots, U^{\#}\}$, respectively.

Each user $u \in U$ exhibits a certain resource demand of $RD_u \in \mathbb{R}^+$ units, for which he/she is willing to pay a total of $M_u^+ \in \mathbb{R}^+$ monetary units. Furthermore, each cloud provider $p \in P$ is able to provide a resource supply of $RS_p \in \mathbb{R}^+$ units at a total cost of $M_p^- \in \mathbb{R}^+$. Please note that resource demands and supplies could also be expressed in a multi-dimensional fashion, i. e., with respect to different resource types. However, for the sake of simplicity, we assume one-dimensional resource constraints at this point, a notion that is also followed in related publications (Hans et al., 2013).

Consumption and provision of services is subject to certain QoS and security constraints, which we refer to by the common term of *non-functional* constraints. Specifically, we assume two sets, $A = \{1, 2, \ldots, A^{\#}\}$ and $\hat{A} = \{1, 2, \ldots, \hat{A}^{\#}\}$, of quantitative and qualitative non-functional attributes. Quantitative attributes represent numerical properties, e. g., availability or latency. In contrast, qualitative attributes correspond to nominal properties, e. g., data center location in the European Union or adherence to a certain security policy.

The cloud providers make certain guarantees with respect to the non-functional attributes. For each quantitative attribute $a \in A$, the value guaranteed by provider $p \in P$ is denoted as $AG_{a,p} \in \mathbb{R}$. For each qualitative attribute $\hat{a} \in \hat{A}$, the corresponding information is given by $\hat{AG}_{\hat{a},p} \in \{0, 1\}$.

Inversely, the cloud users specify certain requirements concerning the non-functional attributes. With respect to each quantitative attribute $a \in A$, the value required by user $u \in U$ is denoted as $AR_{a,u} \in \mathbb{R}$. Likewise, $\hat{AR}_{\hat{a},u} \in \{0, 1\}$ denotes the requirement for each qualitative attribute $\hat{a} \in \hat{A}$, i. e., indicates whether this attribute is mandatory or not. Without loss of generality, we assume that the users specify lower bounds (e. g., minimum availability) on their quantitative non-functional requirements. Upper bounds (e. g., maximum latency) can be easily incorporated into the model by negation of the respective values.

As it has been briefly explained in the previous section, the challenge for the cloud broker consists in composing cloud collaborations, consisting of multiple cloud providers, and subsequently assigning users to them. In that process, the objective for the broker is to maximize his/her profit, i. e., the difference between the revenue from the served cloud users and the spending on the incorporated cloud providers. As most important constraint, each collaboration should obviously offer sufficient resource supplies to serve the assigned users. The process is further subject to the constraint that the QoS and security requirements of each user are fulfilled by the cloud collaboration to which he/she has been assigned.

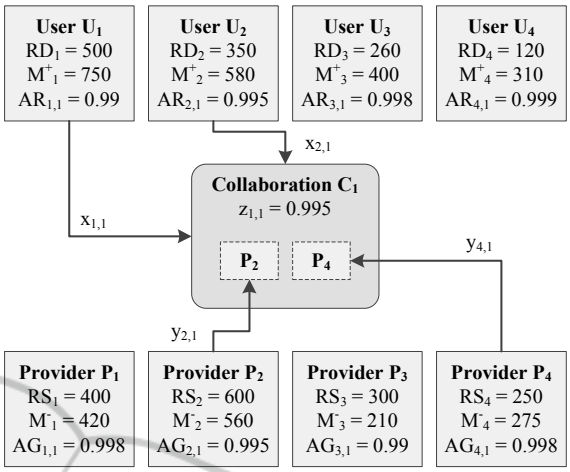

Figure 1: Tangible example of a CCCP instance with four users and providers.

A tangible, simplified example for a CCCP instance is provided in Figure 1. The instance exhibits four users and providers with different resource demands/supplies and non-functional requirements/guarantees, respectively. In the example, providers P_2 and P_4 form a collaboration, which enables them to conjointly serve users U_1 and U_2 under the given constraints. Both providers substantially profit from the collaboration, since their combined resource supply permits to serve larger customers and allows to achieve a higher degree of resource utilization.

3 OPTIMIZATION APPROACH CCCP-EXA.KOM

Based on the notations that were introduced in the previous section, the CCCP can be transformed into an optimization model. The result is given in Model 1 and will be explained in the following.

To start with, $x_{u,c}$ and $y_{p,c}$ are the *main* decision variables in the model (cf. Equation 11). They are defined as binary and indicate whether user u or provider p, respectively, has been assigned to collaboration c or not. As additional axillary decision vari-

ables, we introduce $y'_{p,c}$, which are also binary and serve as complement to $y_{p,c}$, hence indicating the non-assignment of a provider p to a collaboration c. Furthermore, $z_{a,c}$ and $\hat{z}_{\hat{a},c}$ are specified (cf. Equation 12). They are defined as real and binary, respectively, and represent the cumulative value of the non-functional property a or $\hat{a}$, respectively, for collaboration c. The variables x and y are referred to as main decision variables, since they have a direct impact on the objective function. In contrast, y', z, and $\hat{z}$ only have an indirect influence.

As outlined before, the objective consists in profit maximization (cf. Equation 1). That is, the difference between the revenue from the served cloud users and the spending on the used cloud providers should be maximized, depending on the values of the decision variables. Please note that other objectives, such as maximizing the number of served users or the overall resource utilization, may easily be incorporated into the optimization model as well. However, given the general idea of a competitive cloud market, our initial focus is placed on monetary objectives.

Equations 2 and 3 make sure that each user and provider is assigned to not more than one collaboration. Thus, the broker may opt to *not* satisfy certain users' demands, but also to not exploit cloud providers as part of a collaboration. Equation 4 determines the inverse variable $y'_{p,c}$ for each decision variable $y_{p,c}$. This definition is used in the following two Equations 6 and 7. They determine the cumulative non-functional values for quantitative and qualitative attributes, respectively. Both equations are formulated such that quantitative properties are given by the "worst" value among all providers in a certain collaboration, i. e., the "weakest link in the chain". Equations 8 and 9 make sure that users can only be assigned to such collaborations that make sufficient non-functional guarantees, given the users' specific non-functional requirements. Lastly, Equation 10 defines a set of potential cloud collaborations. The underlying notion for the given definition is that no user or provider will be assigned to more than one collaboration (recall Equations 2 and 3). Hence, the maximum number of collaborations is given by the number of users or providers, whichever is lower.

Based on the given model, it can easily be seen that the CCCP constitutes a Mixed Integer Program (MIP), i. e., a special form of Linear Program (LP) that features both integer (in this case, binary) and natural decision variables. Thus, the problem can be solved using off-the-shelf optimization algorithms, such as branch-and-bound (Hillier and Lieberman, 2005), in order to obtain an exact (i. e., profit maximal) solution.

While branch-and-bound is known to perform very well on many Integer Program (IP) and MIP problems, it is ultimately still based on the principle of enumeration (Hillier and Lieberman, 2005). Thus, in the worst case, the computational complexity of obtaining an exact solution grows with the size of the solution space. In the specific case of the CCCP, this translates into an exponential growth with the problem size, i. e., the number of considered providers and users.

4 EVALUATION

To assess the practical applicability of our proposed approach CCCP-EXA.KOM, we have prototypically implemented it in Java 7. In order to transfer Model 1 into a programmatic representation, we use the free JavaILP framework[1]. While this potentially permits for the application of different backend solver framework, we have selected the commercial IBM ILOG CPLEX framework[2] as default due to its favorable performance (Meindl and Templ, 2012) and its popularity in related research, e. g., (Hans et al., 2013; Mashayekhy and Grosu, 2012).

4.1 Evaluation Setup and Procedure

The main objective of our evaluation is to assess the required computation time of CCCP-EXA.KOM for different problem sizes. This allows us to judge the applicability of the proposed approach under practical conditions, where time constraints in the decision process play an important role. Thus, formally, we regard *computation time* as the *dependent* variable of our evaluation.

As *independent* variables, we include the number of considered users and providers, i. e., $U^{\#}$ and $P^{\#}$. In contrast, the number of quantitative and qualitative non-functional attributes were fixed ($A^{\#} = 1$ and $\hat{A}^{\#} = 1$); hence, they constitute *controlled* variables. This is justified by two aspects: First, these variables are likely also predefined in practice. Second, they do not have an impact on the number of decision variables and hence, the size of the solution space. Each specific combination of $U^{\#}$ and $P^{\#}$ results in a *test case*. For each test case, we created 100 specific CCCP instances with the according dimensions.

The parameter values or distributions that were used in the problem generation process are summa-

[1] http://javailp.sourceforge.net/.

[2] http://www.ibm.com/software/integration/optimization/cplex-optimizer/.

Model 1: Cloud Collaboration Composition Problem

$$\text{Max. } Pr(x,y,y',z,\hat{z}) = \sum_{u\in U, c\in C} x_{u,c} \times M_u^+ - \sum_{p\in P, c\in C} y_{p,c} \times M_p^- \quad (1)$$

such that

$$\sum_{c\in C} x_{u,c} \leq 1 \quad \forall u \in U \quad (2)$$

$$\sum_{c\in C} y_{p,c} \leq 1 \quad \forall p \in P \quad (3)$$

$$y_{p,c} + y'_{p,c} = 1 \quad \forall p \in P, \forall c \in C \quad (4)$$

$$\sum_{u\in U} x_{u,c} \times RD_u \leq \sum_{p\in P} y_{p,c} \times RS_p \quad \forall c \in C \quad (5)$$

$$z_{a,c} \leq y_{p,c} \times AG_{p,a} + y'_{p,c} \times \max_{p\in P}(AG_{p,a}) \quad \forall p \in P, \forall c \in C, \forall a \in A \quad (6)$$

$$\hat{z}_{\hat{a},c} \leq y_{p,c} \times \hat{AG}_{p,\hat{a}} + y'_{p,c} \quad \forall p \in P, \forall c \in C, \forall \hat{a} \in \hat{A} \quad (7)$$

$$z_{a,c} \geq x_{u,c} \times AR_{u,a} \quad \forall u \in U, \forall c \in C, \forall a \in A \quad (8)$$

$$\hat{z}_{\hat{a},c} \geq x_{u,c} \times \hat{AR}_{u,\hat{a}} \quad \forall u \in U, \forall c \in C, \forall \hat{a} \in \hat{A} \quad (9)$$

$$C = \{1, 2, \ldots, \min(P^{\#}, U^{\#})\} \quad (10)$$

$$x_{u,c} \in \{0,1\} \quad \forall u \in U, \forall c \in C \quad (11)$$
$$y_{p,c} \in \{0,1\} \quad \forall p \in P, \forall c \in C$$

$$y'_{p,c} \in \{0,1\} \quad \forall p \in P, \forall c \in C \quad (12)$$
$$z_{a,c} \in \mathbb{R} \quad \forall a \in A, \forall c \in C$$
$$\hat{z}_{\hat{a},c} \in \{0,1\} \quad \forall \hat{a} \in \hat{A}, \forall c \in C$$

rized in Table 1. The specifications of the non-functional parameters are based on the notion that the sole quantitative and qualitative attribute represent availability (a QoS aspect) and data center location in the European Union (a security aspect), respectively. Furthermore, monetary parameters were set such that higher availability results in quickly increasing values, based on the observation that each additional "nine" in the availability figure results in doubled cost (Durkee, 2010). In contrast, a EU data center location only leads to a moderate increase of 10%, which closely corresponds to the price difference observed for Eastern U.S. and Ireland-located Amazon EC2 VM instances (Amazon Web Services, Inc., 2013).

Table 1: Parameter values and distributions used in the problem instance generation. Abbreviations: Uni – Uniform distribution; Ber – Bernoulli distribution.

Param.	Value/Distribution
$AR_{1,u}$	Uni$(0.99, 0.9995)$
$\hat{AR}_{1,u}$	Ber(0.5)
$AG_{1,p}$	Uni$(0.995, 0.9995)$
$\hat{AG}_{1,p}$	Ber(0.5)
RD_u	Uni$(1000, 5000)$
RS_p	Uni$(1000, 5000)$
M_u^+	$\alpha_u \times RD_u \times log_{10}(1 - AR_{1,u})^2 \times (1.1^{\hat{AR}_{1,u}})$
M_p^-	$\beta_p \times RS_p \times log_{10}(1 - AG_{1,p})^2 \times (1.1^{\hat{AG}_{1,p}})$
α_u	Uni$(1.5, 1.75)$
β_p	Uni$(1.0, 1.25)$

Following the generation, we computed a solution to each problem instance using our prototypical implementation of CCCP-EXA.KOM. In that process, we imposed a timeout of 300 seconds (i. e., five minutes) per problem instance. Based on the resulting sample of computation times for the successfully solved problems, we computed the mean computation time, as well as the 95% confidence interval. The evaluation was conducted on a desktop computer, equipped with an Intel Core 2 Duo E7500 processor and 4 GB of memory, operating under the 64-bit edition of Microsoft Windows 7.

4.2 Evaluation Results and Discussion

The results of our evaluation, i. e., the observed mean computation times per test case, are graphically illustrated in Figure 2. As can be clearly seen, the computation times quickly increase with the problem size, i. e., the considered number of users and providers. The effect is less pronounced for the smallest two problem classes (with $U^{\#} \leq 6$ and $P^{\#} \leq 9$); in fact, for these two test cases, there is no statistically significant difference in mean computation time observable at the 95% confidence level. In absolute terms, we already find absolute computation times in the order of magnitude of one-hundred seconds and one second respectively for the medium-sized test cases with $U^{\#} \leq 8$. For these test cases, increasing the number of

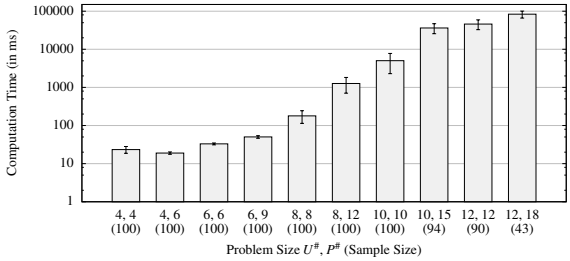

Figure 2: Evaluation results, i. e., observed mean computation times (with 95% confidence intervals) for CCCP-EXA.KOM by test case. Please note the logarithmic scaling of the ordinate.

providers increases the computation time by a factor of approximately ten already.

For the four largest test cases (with $U^{\#} \geq 10$ and $P^{\#} \geq 10$), the absolute computation times reach the order of magnitude of seconds and ten seconds. All observed increases are statistically significant at the 95% confidence level. In addition, the ratio of solved problem instances sharply drops with growing problem size. This effect is most notable for the largest problem class that involves 12 users and 18 providers, where only 43% of the 100 problem instances could be solved within the timeout period of five minutes.

Given that the considered problem dimensions are still relatively small in the context of a large cloud market, it can be concluded that the practical applicability of the proposed optimization approach CCCP-EXA.KOM is rather limited. As it has already been explained before, a broker will likely have to decide on the composition of collaborations under rigid time constraints, since users likely require resources at short notice.

Hence, an important future challenge consists in the development of appropriate heuristics, which permit to trade reductions in computation time against small degradations in broker profit, and are consequently applicable to practically relevant, large-scale problem instances. In that context – apart from its potential application to small-scale problem instances – CCCP-EXA.KOM can serve as a valuable performance benchmark.

5 RELATED WORK

Niyato et al. (Niyato et al., 2011) study the cooperative behavior of multiple cloud providers in order to cooperate and support the establishment of resource pools to offer services to public cloud users. The authors present a stochastic LP game model which takes the random internal demand of cloud providers and a transferable utility into account to define and commit the optimal offer of cooperated cloud providers. In contrast to our work, Niyato et al. do not consider non-functional constraints, i. e., QoS and security requirements.

In a more recent work, Niyato et al. (Niyato et al., 2012) examine building coalitions between cloud providers as a novel approach to optimize the capacity expansion and maximize the mobile cloud providers' monetary benefits. The authors consider cooperative game theory and the Nash equilibrium principles in their approach and propose admission control and revenue sharing strategies for building cloud provider coalitions and a resource pool for mobile applications. The provided results illustrate improvements in cloud providers' capacity and profit maximization by entering such cloud coalitions. Similar to their previous work, the authors do not consider non-functional constraints, which are an important aspect of our work.

Gohad et al. (Gohad et al., 2013) propose a dynamic algorithm for forming self-adaptive cloud collaborations based on the identifying most appropriate healthy set of cloud provider resources (cloud provider capabilities and functional abilities at the SaaS layer), cost modeling and tenancy requirements. The approach is highlighted with a realistic example. In contrast to us, Gohad et al. focus on ad-hoc resource provisioning, rather than the long-term formation of cloud collaborations, and do not consider security aspects. This specifically includes the cumulative security properties of cloud collaborations that were a focal point of our work.

Song et al. (Song et al., 2010) examine the problem of task selection and allocation to physical machines in the context of dynamic cloud collaborations. Their objective consists in the balancing of resource demands under consideration of different resource types, such as CPU and memory. For that purpose, the authors propose three heuristic optimization approaches, and demonstrate that a cooperative heuristic has benefits with respect to the objective of balanced resource utilization. In contrast to us, Song et al. focus on individual cloud providers and do not regard security requirements.

Mashayekhy and Grosu (Mashayekhy and Grosu, 2012) model a cloud federation formation problem based on the game theory and formulate a corresponding IP-based optimization approach. In their model, the authors consider the cooperative provisioning of VM instances and storage by federated cloud providers. Their objective consists of profit maximization combined with the formation of stable coalitions, i. e., coalitions in which cloud providers do not have a monetary incentive to switch to differ-

ent coalitions. In contrast to our work, the authors only consider resource constraints, but do not regard non-functional requirements. Their work also aims at low-level VM provisioning, rather than strategic composition of collaboration.

Lastly, Hans et al. (Hans et al., 2013) have examined the cost-efficient selection of cloud data centers for the delivery of multimedia services. In that context, the authors propose an exact optimization approach based on IP. While their work is similar with respect to the consideration of resource and QoS constraints, it focuses on a single cloud provider and does neither regard the composition of collaborations nor qualitative non-functional aspects.

In conclusion, to the best of our knowledge, we are the first to examine the profit-maximal, strategic composition of cloud collaborations under consideration of cumulative non-functional properties that result from the very formation of these collaborations, i. e., are determined by the "weakest link in the chain". Apart from the identification of that specific problem, our main contribution consists in the proposal of an exact optimization approach, which can serve as benchmark for future heuristic approaches.

6 SUMMARY AND OUTLOOK

While cloud computing promises access to virtually unlimited IT resources, the physical infrastructure of cloud providers is actually limited. Hence, smaller providers may not be able to serve the demands of larger customers. A possible solution is cloud collaborations, where multiple providers join forces to conjointly serve customers. Unfortunately, in such scenario, non-functional QoS and security properties are determined by the "weakest link in the chain", rendering the process of composing collaborations cumbersome.

In this work, we introduced the corresponding *Cloud Collaboration Composition Problem*. We proposed an initial solution approach named CCCP-EXA.KOM based on Mixed Integer Programming, and evaluated it with respect to its computation time requirements. Our results indicate that exact optimization approaches are only applicable to small-scale problem instances, thus indicating the need for the development of custom-tailored heuristic approaches.

Accordingly, the development of such heuristics will constitute the primary focus of our future work. In addition, we plan to extend the proposed model to cater for more complex non-functional constraints, such as conditional requirements (e. g., strong data encryption is only required if data is placed outside the European Union).

ACKNOWLEDGEMENTS

This work has partly been sponsored by the E-Finance Lab e. V., Frankfurt a.M., Germany (www.efinancelab.de.).

REFERENCES

Amazon Web Services, Inc. (2013). Amazon EC2 Pricing, Pay as you go for Cloud Computing Service. http://aws.amazon.com/en/ec/pricing/.

Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., and Brandic, I. (2009). Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Generation Computer Systems*, 25(6):599–616.

Durkee, D. (2010). Why Cloud Computing Will Never Be Free. *Queue*, 8(4):20–29.

Gohad, A., Ponnalagu, K., Narendra, N. C., and Rao, P. S. (2013). Towards Self-Adaptive Cloud Collaborations. In *2013 Int. Conf. on Cloud Engineering*.

Hans, R., Lampe, U., and Steinmetz, R. (2013). QoS-Aware, Cost-Efficient Selection of Cloud Data Centers. In *6th Int. Conf. on Cloud Computing*.

Hillier, F. and Lieberman, G. (2005). *Introduction to Operations Research*. McGraw-Hill, 8th edition.

Mashayekhy, L. and Grosu, D. (2012). A Coalitional Game-Based Mechanism for Forming Cloud Federations. In *5th Int. Conf. on Utility and Cloud Computing*.

Meindl, B. and Templ, M. (2012). Analysis of Commercial and Free and Open Source Solvers for Linear Optimization Problems. Technical report, Technische Universität Wien.

Niyato, D., Vasilakos, A. V., and Kun, Z. (2011). Resource and Revenue Sharing with Coalition Formation of Cloud Providers: Game Theoretic Approach. In *11th Int. Symp. on Cluster, Cloud and Grid Computing*.

Niyato, D., Wang, P., Hossain, E., Saad, W., and Han, Z. (2012). Game Theoretic Modeling of Cooperation Among Service Providers in Mobile Cloud Computing Environments. In *2012 Wireless Communications and Networking Conf.*

Song, B., Hassan, M. M., and Huh, E.-N. (2010). A Novel Heuristic-Based Task Selection and Allocation Framework in Dynamic Collaborative Cloud Service Platform. In *2nd Int. Conf. on Cloud Computing Technology and Science*.

Wenge, O., Siebenhaar, M., Lampe, U., Schuller, D., and Steinmetz, R. (2012). Much Ado About Security Appeal: Cloud Provider Collaborations and Their Risks. In *1st Europ. Conf. on Service-Oriented and Cloud Computing*.