An Agency Perspective to Cloud Computing

Frank Pallas[1,2]

[1] Karlsruhe Institute of Technology, Center for Applied Legal Studies,
76131 Karlsruhe, Germany
frank.pallas@kit.edu
[2] FZI Forschungszentrum Informatik,
10117 Berlin, Germany

Abstract. The field of cloud computing is strongly affected by conflicts of interest between providers and users of resources. A comprehensive and integrative model for representing and analyzing these conflicts on a theoretically well-founded basis is, however, still lacking. Therefore, this paper establishes such a model based on economic agency theory. Employing two realistic example scenarios, we identify representative challenges faced by cloud users and generalize them as typical problems present in agency relations. Based on this conception, we correlate existing practices and strategies from cloud computing with corresponding abstract instruments from agency theory. Finally, we identify approaches that are – even if suggested by economic theory – not practically employed in the cloud domain and discuss the potential to utilize them in future technical and non-technical developments.

Keywords: Cloud Computing, Agency, Principal-Agent, Adverse Selection, Moral Hazard, Hold-Up.

1 Introduction

Given its paradigmatic differences from traditional IT usage, cloud computing introduces completely novel challenges of technical and non-technical nature. In particular, the relation between the provider and the user of a certain cloud resource deserves specific attention: While the fulfillment of certain requirements can be of crucial importance for the user's decision to employ cloud computing at all, it does in many cases not lie in the interest of the provider. A certain level of physical datacenter security might, for example, be an essential precondition for a potential user to actually consider cloud storage as a viable option. For the provider, however, establishing the required level of physical security raises significant costs which he will usually try to limit to the extent that actually pays off for him. Under certain conditions (which will be laid out in more detail throughout this paper) and without proper countermeasures being used, this would lead a rational provider to not establish the security level aspired by the would-be customer. A potential customer, in turn, can then be assumed to be aware of this fact and therefore to abstain from the use of cloud storage at all.

Comparable problems arise with regard to a multitude of further aspects of cloud computing. Whether in matters of security, legal compliance, long-term availability or many other characteristics: We are always confronted with fundamental conflicts of interest between the user and the provider which, on the large scale, may hinder the broad application of cloud computing in general. There are thus good reasons to search for appropriate strategies and countermeasures for adressing them properly.

Ongoing and ever-increasing research in this field notwithstanding, we do, however, still lack well-established abstract models for representing, understanding and counteracting cloud-specific conflicts of interests between users and providers of resources. It is the aim of this paper to establish such an abstract model which covers the most constitutive factors shaping typical settings of cloud computing. Furthermore, the model shall be applicable across different conflict domains and allow for the development of novel and theoretically well-founded approaches for counteracting the identified conflicts in practice.

For developing our abstract model, we employ economic theory to categorize and understand relevant factors and conflicts arising in the field of cloud computing, thus following an approach of positive economics [8]. In particular, we mainly build upon the well-known agency model as established by [14] and extensively analyzed by [20].[3] Even if the relevance of agency-theory has occasionally been recognized for selective aspects of cloud [9, 12], grid [7] and service computing [31] as well as for the closely related domain of IT-outsourcing [11, 6, 3] before, these considerations have not yet been condensed into an integrative abstract model as established herein. As we will see, however, such a model allows us 1) to better understand and structure the conflicts and challenges arising in the field of cloud computing on an abstract level, 2) to categorize and assess existing concepts for counteracting these conflicts and challenges and 3) to identify possible starting points for necessary extensions to the status quo in matters of technical and non-technical instruments for heightening the broad applicability of cloud computing.

To illustrate our rather conceptional considerations and to demonstrate the broad practical applicability of our abstract model, we use two simple scenarios exemplifying some common materializations of the generic conflicts of interests arising in the context of cloud computing. Of the broadly recognized cloud service models [26], these scenarios mainly comprise infrastructure (IaaS) and platform (PaaS) services, but our considerations and findings are similarly applicable to Software (SaaS) services. Of the various deployment models [26], in turn, we concentrate on settings following a "public-cloud" model and explicitly ignore those where cloud technologies are merely used in-house without involving external providers ("private cloud"). To a certain extent, our findings will also be applicable to intermediate models like "hybrid" and "community clouds", but our primary focus is on "public" offerings.

[3] For an overview of agency theory also discussing its different understandings and lines of research, see also [5].

Finally, we explicitly confine our considerations to two-party relations with just one cloud provider and one cloud user involved. Being well aware that practial applications of cloud computing often involve multiple parties with sometimes highly complex interdependencies [27, 21, 15], we consciously do so herein because a sound and scientifically well-founded understanding of the fundamental two-party relation provides an indispensable basis for corresponding future deliberations on rather complex settings.

This being said, the remainder of this paper is structured as follows: In section 2, we sketch our example scenarios that will be used throughout the paper. In section 3, we briefly outline the fundamentals of agency theory and some related concepts from new institutional economics, which together form the theoretical basis of our model. Section 4 then maps these concepts onto the previously sketched example scenarios, establishes our agency perspective to cloud computing and deductively identifies measures that are suggested by theory but not yet broadly applied in practice. Finally, section 5 sums up our findings and points to auspicious strands for future research.

2 Example Scenarios

As a first representative example of cloud usage and the conflicts of interests arising in this context, we assume a European medium enterprise wanting to ensure fallback availability of its most important internal systems for the case of large-scale local infrastructure failures. Instead of erecting and maintaining a complete remote site on their own, the enterprise decides to use a model of "cloud standby" [23] for this purpose. In this model, the fallback-infrastructure is held available in the form of virtual machines which are updated to the current state of software and data on a regular basis (e.g. once a day) but which are apart from these update cycles constantly inactive until a disaster actually happens. As opposed to the operation of a complete remote fallback site, this model would provide significant benefits to the company [22] as long as all requirements are met.

Besides fundamental functional properties (e.g. regarding operational capability, performance, etc.), these requirements also include several nonfunctional ones. For example, we can assume that the infrastructure to be replicated also comprises databases holding personal data. As these data must be accessible within the virtualized fallback infrastructure, they cannot be stored in anonymized form, which, in turn, raises several legal restrictions. In our case, these obligate the enterprise to ensure that personal data does not leave the European Union[4], that data are not used (e.g. by the provider) for other pur-

[4] In fact, the actual legal regulations are far more sophisticated. For instance, the transfer of personal data beyond the EU can be allowed if the destination country is explicitly recognized as providing an "adequate level of protection" or in case the transfer is covered by instruments like the "safe harbor agreement" or the so-called "standard contractual clauses". This would, however, under certain circumstances (like, e.g., under German legislation) invalidate the legal construct of "processing

poses than originally collected for, and that certain security procedures are in place to prevent unwanted disclosure. In ensuring compliance with these legal restrictions, the enterprise is obviously dependent on the cloud provider's conduct and will therefore insist upon respective contractual agreements. As soon as the adherence to such agreements can hardly be verified, however, the cloud provider may still have incentives to act against his customer's interest because data replication beyond Europe may be the cheapest way to meet certain availability guarantees, because exploiting the data might provide additional business value or because maintaining certain strong security procedures would raise significant costs. The question is, then, how the medium enterprise can ensure that the cloud provider does not opportunistically serve his own goals.

The same question also arises within our second example. Here, we assume a small startup firm that pursues the strategy of erecting their whole IT infrastructure "in the cloud" from the very beginning in order to ensure strong adaptability to changing requirements (e.g. in matters of sudden growth or highly variable usage patterns) without having to invest substantially into physical infrastructure. Due to their specific use case, the startup firm needs cloud-based resources 1) for running virtual machines with a self-deployed distributed web application, 2) for storing and analyzing large amounts of social media data retrieved from third parties, 3) for storing and handling own datasets (including user data) in a consistent way, and 4) for payments processing. These functionalities are to be realized on the basis of advanced, ready-to-use cloud-based services which are, for reasons of communication performance, to be sourced from one single provider.

Again, this scenario raises basic, functional requirements as well as further, non-functional ones. In particular, we assume the startup firm to have a vital interest in stability and availability of the overall system, in updates, patches, etc. being promtly installed to the virtualization layer by the provider, and in the establishment and proper maintenance of efficient mechanisms for anomaly and intrusion detection. Furthermore, the startup has an interest in a high quality of the fraud detection mechanisms employed in the payment service and, finally, in its core assets – the large amounts of social media data – not being used for own purposes by other parties including the provider himself. Like in the above example, acting in the best interest of his customer is not in the interest of the cloud provider because this would raise costs and efforts for heightening availability, for constant system updates, for the maintenance and enhancement of anomaly, intrusion and fraud detection mechanisms, etc. Again, this implies the question how the startup firm can ensure its requirements to be appropriately met even if this does – a priori – not lie in the interest of the provider.

As we see, the conflicts of interests between the user and the provider of cloud resources can be manifold and originate from various directions including security, legal compliance, quality of results, etc. Instead of – as it is often done

---

on behalf of the controller" and thus result in further complications. Without going more into detail and for the sake of clarity, we therefore assume – like it is usually done in practice – a strict "EU-internal" requirement.

– addressing each of these domains directly, we do herein strive for a rather generic, abstract model for representing and analyzing these conflicts on a theoretical basis. This model shall later serve as starting point for the identification of foreseeable obstacles to a broader adoption of cloud computing and for the development of respectice countermeasures. Economic agency theory matches this aim and the basic setting with user and provider facing a goal divergence quite well and will, supplemented by further concepts from new institutional economics, therefore serve as the theoretical basis of our model. Its foundations shall thus be outlined in brief before applying it to our exemplary scenarios in section 4.

3 The General Agency Model

Basically, the general agency model consists of two parties interacting with each other – the principal and the agent. Both are assumed to be opportunistic utility maximizers and thus to primarily serve their own individual goals. Under these givens, the "principal engages the agent to perform some service on his behalf, and to facilitate the achievement of the activity, he delegates some decision-making authority to the agent" [10, p. 162]. Furthermore, the model assumes that information is "asymmetric in the sense that (1) the agent's action is not directly observable by the principal [...] or (2) the agent has made some observation that the principal has not made [where] in the first case, one speaks of hidden action, in the second of hidden information" [10, p. 162]. These givens lead to the first problem present in principal-agent relationships: Moral hazard.

3.1 Moral Hazard

Basically, the above-mentioned information asymmetries suggest an opportunistic agent to not always act in the best interest of the principal but to primarily serve his own goals instead. Being aware of the principal's hidden action problem, for example, an agent can put low effort at carrying out the task delegated to him while the fact of hidden information allows him to attribute poor outcomes of his efforts to adverse situational conditions instead of, say, to the fact of having been negligent. In order to counteract this so-called moral hazard problem, different countermeasures can be employed:

- *Monitoring* refers to the principal's activities for reducing the information asymmetries between him and the agent during service provision. In particular, monitoring can refer to agent behavior (adressing hidden action) as well as to situational givens (adressing hidden information). While more comprehensive monitoring reduces the expectable divergence between actual agent behavior and the principal's interest, it also raises costs by itself. These must be weighed against the accomplished reduction of moral hazard.
- *Bonding*, in turn, is a strategy where the agent makes efforts or expenditures in order to guarantee that he will not act opportunistically against the

principal's interest. A typical example for this strategy are bonds deposited with the principal which are "forfeited if the agent is caught cheating" [25, p. 195]. Like monitoring, bonding mechanisms raise costs by themselves and their extent must be weighed against the achieved reduction of moral hazard.

Due to the costs induced by monitoring and bonding mechanisms, it is usually rational for the principal to accept a certain level of opportunistic agent behavior, which is referred to as "residual loss". Together, monitoring costs, bonding costs and residual loss make up the overall "agency costs" a principal has to accept in the course of delegating a task to an agent instead of performing it himself.

Besides the moral hazard problem (and the respective costs) occuring *during* the provision of a delegated service, agency relations are characterized by further challenges. Even if not being part of agency theory's core, the problems of "adverse selection" and "hold-up" apply to many agency relationships and shall therefore also be included in the following analysis.

3.2 Adverse Selection

Adverse selection refers to the principal's initial selection of an agent: While the agent is very well aware of his characteristics and, thus, of the expectable quality of a certain service delivered by him, the principal cannot perfectly assess these characteristics beforehand. Without additional countermeasures being taken, this would lead to a so-called "lemons-market" [1] where high-quality agents cannot achieve appropriate payments and therefore leave the market while low-quality agents stay in. The principal can then only choose between different low-quality agents providing inferior service quality. Again, different strategies can be employed to prevent such unwanted outcomes:

- *Screening* is conducted by the principal and encompasses all activities of examination aimed at directly increasing the principal's knowledge about the agent's actual quality. A typical example of screening are assessment centers in the employment market.
- *Signaling* is basically conducted by the agent and refers to certificates, references and other evidence that shall demonstrate the agent's quality. In order to be meaningful, it must be ensured that such signals are less expensive to emit for high-quality agents than for low-quality ones [28] – like it is, for example, the case for university degrees.
- *Self-Selection*, in turn, refers to models with different contracts being offered to agents, featuring characteristics that make high-quality agents choose other options than low-quality ones. Typical examples of such mechanisms include insurance contracts with different deductibles or employment contracts for salespersons with different commission rates. For this scheme to work properly, it must be ensured that the facts determining the agents outcome (e.g., the sales volume) can easily and doubtlessly be ascertained.

3.3 Hold-Up

Hold-up situations, in turn, emerge from specific, non-recoverable investments being made in preparation of or within an agency relation by one party, thereby providing a certain possibility of opportunistic exploitation – e.g. through demanding price reductions or other detrimental condition changes – to the other [16, pp. 310ff]. In agency relations, hold-up problems share some characteristics of moral hazard but can apply to both, principal and agent, depending on the concrete setting and the specific investments having to be done. Whoever makes a significant investment specific to his counterpart will be at disadvantage and face the risk of his "locked in" position being exploited. Beyond those already mentioned above for counteracting moral hazard in general (esp. bonding), economic theory suggests several countermeasures to address such hold-up situations:

- *Long-term contracts* can, if they are sufficiently complete and easily enforceable, mitigate the hold-up problem through eliminating the risk of ex-post condition changes demanded by one party. On the other hand, long term contracts may also introduce new, "inverted" hold-up situations because of the formerly strong party now being bound to a single contractor, for example [16, pp. 308f].
- *Non-contractual long-term relations* based on mutual trust, reputation and anticipated future rewards are another way for dealing with the risk of hold-up [13, pp. 80ff].
- *Multiple, substitutable counterparties* ensure that there are always alternatives which can be employed in case of one partner trying to exploit a hold-up situation. Thereby the risk of such opportunistic behavior can be limited [13, pp. 80ff]. On the other hand, this also implies that a highly specific investment must be done multiple times – either by one party having to adopt to the specifics of multiple partners or by the different partners each adopting to the same specifics of their common counterparty.
- *Vertical Integration*, finally, also is a possible strategy for dealing with the need to make specific investments. Instead of trying to manage the hold-up problem, the involved parties are simply merged into one entity. Especially in case of "extensive hold-up problems", vertical integration can prove advantageous over remaining with a two-party agency relation [13, p. 74].

After having laid out the fundamentals of agency theory and strongly related problems from new institutional economics in general, we will now show how these abstract concepts can be mapped to our exemplary scenarios and what we can learn from this mapping with regard to well-known concrete challenges of cloud computing.

4 An Agency Model of Cloud Computing

Given the above-mentioned fundamental characteristics of agency relations, the application of agency-theory to cloud-based settings seems promising. In the

following, we will therefore demonstrate how abstract concepts from agency theory on the one and concrete problems and approaches from the field of cloud computing on the other hand can be mapped onto each other by interpreting the cloud user (in our example scenarios, this is the medium enterprise or the startup firm, respectively) as the principal and the cloud provider as the agent. For each of the main challenges – adverse selection, moral hazard and hold-up – we will furthermore demonstrate in brief how agency theory can be employed to deduce starting points for the development of novel or enhanced approaches for diminishing uncertainties and risks currently hampering a broader adoption of cloud computing in general.

4.1 Adverse Selection in Cloud Computing

Like any agency relation, every use of cloud computing begins with the selection of an appropriate provider. In the example of cloud standby, for instance, this means that the medium enterprise has to find a provider that offers the required functionalities (operational capability, performance, etc.) and at the same time provides sufficient certainty that the enterprise can still comply with its legal obligations regarding the handling of personal data replicated to the cloud.

While the fulfillment of basic functional requirements can easily be estimated in advance through test installations, benchmarks, or provider-independent comparisons based on long-term measurements[5], this is not the case for non-functional ones. In particular, it is basically unknown to the potential cloud user what security precautions a certain provider has in place and of what quality the provider's security competences are in general. While the ex-ante information asymmetries seem manageable with regard to functional requirements, they are thus significant in matters of characteristics like security [9].

Faced with this uncertainty (and in the absence of any measures for counteracting it), the enterprise having to select a provider would for any potential choice have to assume a medium level of security and thus a medium probability of being able to meet its legal obligations. This, in turn, would imply a medium willingness to pay, drive those providers out of the market that invest considerable efforts in the provision of high security and, in the end, lead to a "lemons market" for security [2] where a rational customer would have to expect none of the remaining providers to offer sufficient security for complying with his legal obligations. In this case, the enterprise would certainly abstain from employing a cloud-standby model at all. The same effect could also be expected for the second example of a startup firm pursuing an all-encompassing cloud strategy from the very beginning. Beyond the above security-related considerations, the adverse selection effect could here also be expected for the quality of the fraud and misuse detection mechanisms employed in the payment service, for example.

In order to counteract this adverse selection effect, agency theory suggests screening, signaling and self-selection as potential strategies. These can be mapped

[5] Like, e.g., CloudHarmony – https://www.cloudharmony.com/

to different measures commonly used in or at least proposed for the practice of cloud computing.

- *Screening:* Conducting in-depth security audits, for example, would from the agency perspective represent a screening activity and could in fact reduce ex-ante uncertainty about the provider's security-related capabilities. Unfortunately, such audits would raise significant screening costs as compared to the expectable overall volume of most cloud contracts, rendering it an unfeasible option for most cases.[6] Technologies providing trustworthy information about provider-internal givens to external parties [17, 4, 30] might, however, also be harnessed for assessing the capabilities of different providers and therefore be employed as screening tools for diminishing the risk of adverse selection.
- *Signaling:* Audit certificates issued by third parties are an instrument often referred to with regard to cloud computing [29]. Even if chronically misunderstood as proving certain assured facts about actual givens and conduct currently present on the provider's side, such certificates rather have to be considered as credible signals about the provider's capabilities in the respective domain. For instance, a security certificate issued a year ago only confirms that a provider demonstrated *his ability* to establish a certain level of security to the auditing third party that one year ago. It does, however, say nothing about whether the provider still makes the same effort of maintaining a certain level of security today or about whether he opportunistically exploits his customer's data for own purposes. Nonetheless, and perfectly in line with agency theory, certificates can still be an effective measure for counteracting the adverse selection problem faced by cloud users as they help to distinguish between providers with strong and those with weak capabilities in the respective domain.[7]
- *Self-Selection:* Contract schemes employing self-selection to make the provider reveal information about his capabilities, in turn, are currently not broadly established on the cloud market. They might, however, provide an interesting option for future arrangements. Strictly following the abstract concept of self-selection laid out above, this would require a menu of contracts being offered by the cloud user which contains options more interesting to high-quality providers on the one and other options preferred by low-quality providers on the other hand, while the facts that determine the provider's ultimate outcome must be easily ascertainable. For our example cases, one could, for instance, think of different contract options obligating the provider to pay different penalties in case a breach of data becomes publicly known. A provider with poor security capabilities would then tend to choose the option with the lowest penalty even if this implies a lower base price while a

[6] High-volume contracts concluded by government agencies or large, multinational corporations might be the exception here.

[7] In order to actually provide this functionality, it is a necessary (but not sufficient) condition that receiving the certificate is more expensive for less capable providers than for strong ones. This can, however, be assumed for typical certification schemes.

high-quality provider would rather choose an option with a higher possible compensation in exchange for a higher base price. Of course, more complex contract menus could also be thought of but it should even from this simple example become clear that self-selection schemes offer interesting options for counteracting ex-ante information asymmetries and the resulting problem of adverse selction in cloud computing scenarios. This would, however, require that cloud users actually have the necessary bargaining power and are not, as it is mostly the case today, simply facing "take it or leave it" offers from providers.

4.2 Moral Hazard in Cloud Computing

Moral hazard problems are probably the most decisive obstacles for a further development of the cloud market today. As delineated above, they emanate from information asymmetries between the principal and the agent regarding agent behavior (hidden action) and the circumstances of this behavior (hidden information).

Mapped to our cloud standby example, this moral hazard problem would particularly emerge with regard to the cloud provider's data handling, as the cloud user cannot know whether the provider actually transfers personal data beyond the European Union, whether he exploits the data for own purposes, or what efforts he actually spends in order to maintain security procedures preventing unintended disclosure of the data. Being aware of the user's nescience, the provider then has a clear incentive to act opportunistically: Whenever secretly transferring data beyond the EU, exploiting the data, or cutting down on security-related efforts would provide a benefit, he could – as long as no countermeasures are in place – be assumed to do so, even if he is contractually bound to abstain from such conduct. A well-informed user would be aware of this fact and thus know that he cannot ensure his legal obligations to be met. In the end, he would not employ the model of cloud standby at all.

For our startup example, the situation would be quite similar. The startup firm also cannot know whether security patches are promtply installed, whether intrusion and fraud detection mechanisms are constantly developed further, or whether the provider exploits the massive social media data for his own purposes to generate additional income. Regarding the requirements for maximum stability and availability, in turn, one could think that the provider's efforts are very well observable for the user through external monitoring of the efforts' outcome [12, p. 448]. This does, however, raise the problem of hidden information: For instance, the provider could explain that observed downtimes or crashes are not caused by his inappropriate conduct but rather by unforeseeable adverse exogenous conditions like a massive DDoS-attack or self-propagating large-scale glitches that lay beyond his influence. Without knowledge about such circumstances, the user could not assess the appropriateness of the provider's conduct, which, again, results in possibilities for the provider to engage in opportunistic shirking instead of acting in the interest of his customer. In order to prevent such outcomes, agency theory suggests monitoring and bonding.

– *Monitoring:* As laid out above, monitoring refers to all activities of the cloud user aimed at reducing the information asymmetries about the provider's actual conduct as well as about the circumstances the conduct took place in. Random auditing, for instance, is often suggested to cloud users in order to ensure appropriate provider behavior [12, p. 448]. This approach does, however, seem hardly valuable in practice, as it would repeatedly induce considerable monitoring costs for the user which would hardly be justified by the expectable reduction in opportunistic shirking. Technologies for providing trustworthy event logging [17] or "digital evidence" [4, 30], in contrast, also help the cloud user to assess the provider's actual conduct albeit featuring a much better ratio between monitoring costs and reduction of opportunistic provider behavior. For our first scenario, we can, for instance, expect any technology providing the cloud user with trustworthy information about storage location, data access or on the provider's security-related effort to be highly welcome by the medium enterprise because this would significantly lower the risk of his data being handled illegitimately, thereby making cloud standby a considerable option at all. For the startup example, this also extends to contextual information. With trustworthy information revealing whether, for instance, a DDoS-attack actually took place, the cloud provider could be prevented from attributing good outcomes to his own efforts while blaming external conditions for bad ones.
– *Bonding:* Like the strongly related concept of self-selection for addressing adverse selection, bonding mechanisms suggested by agency theory for counteracting moral hazard are not yet established in the domain of cloud computing. Their implementation would, however, be quite straight forward: Through depositing a certain (e.g. monetary) bond with the cloud user or a commonly trusted third party, the provider could convince his customer that he will refrain from opportunistically exploiting the cloud user's data, from cutting down on security-related efforts, etc. As soon as the detection of shirking behavior is sufficiently probable (where monitoring plays an important role again), such bonding schemes could thus represent a valuable measure for cloud providers to prove their willingness not to act opportunistically against their customers.

Together, such bonding schemes and technologies reducing monitoring costs for cloud users could thus significantly limit the omnipresent moral hazard problem in cloud computing and thereby play a crucial role for the further establishment of the cloud market.

4.3 Hold-Ups in Cloud Computing

Vendor lock-in is the most important kind of hold-up situation that can be expected in the context of cloud computing. Basically, such a vendor lock-in emerges whenever specific and non-recoverable investments are made by one party of the agency relation. As switching to another partner would be prohibitively expensive for the investing party once the investment is done, it then

faces the hold-up risk of his counterparty demanding detrimental conditions. Consistently with the theoretical concept outlined above, research has shown that potential customers are very well aware of this risk [12, p. 456], constituting a significant obstacle for them to adopt cloud computing at all.

Within our cloud standby example, such specific investments could, for instance, refer to the establishment of the fallback infrastructure and the respective update procedures on the basis of the virtualization environment and the management services (orchestration etc.) made available by the cloud provider. As soon as a once-established infrastructure and the respective automated processes cannot be easily migrated from one cloud provider to another, the medium enterprise faces switching costs which result in a vendor lock-in [19, p. 33]. Finally, this leads to a hold-up situation allowing the provider to, for example, stay with the once-established rates and performance points while the rest of the market follows the usual IT-cycles of price reductions or performance increases.

In the startup example, the hold-up risks are even more significant as the cloud user heavily relies on advanced, ready-to-use services. The specific investments in this case refer to all development activities utilizing the advanced services' programming interfaces and following their predetermined programming paradigms. Especially for advanced services like the massive data storage and analysis service or the payment service assumed in our example, these interfaces and paradigms usually are highly provider-specific and the implementations based upon them are thus hardly transferable. The more our assumed startup firm thus ties itself to the advanced services of a given provider, the higher are the obstacles for switching to another provider and the higher is the risk of this hold-up situation being exploited by the provider through, e.g., inappropriate prices or lowered availability commitments.

Again, agency theory suggests several strategies and instruments for counteracting hold-up risks in the context of cloud computing:

- *Long-Term Contracts:* Of the four strategies suggested by agency theory for counteracting hold-ups, long-term contracts could reduce such risks through ex-ante agreements on periodic price reductions or performance increases. Due to the requirement of such long-term-contracts to be complete and easily enforceable, their application seems, however, at least questionable with regard to the foreseeability of every potential source of hold-up risks as well as in matters of covering security-related and other aspects which are hard to ascertain and which are therefore also subject to the moral hazard problem (see above).
- *Availability of Alternatives:* Ensuring the availability of alternatives which could easily be switched to whenever a provider tries to exploit a hold-up possibility would therefore seem highly worthwile for both of our assumed example users. This would, however, require them to always adopt their solutions to the specific givens of at least two different providers, thus causing significant additional costs. In this regard, technologies abstracting from provider-specific programming interfaces and thus allowing to realize the re-

spective implementations in a provider-independent way[8] play an important role as they significantly reduce the costs of maintaining different options in parallel. Such technologies are, however, currently only available for rather foundational, generic resources like storage or the management and execution of complete virtual machines. More specific, higher-level functionalities like those provided by the data analysis and the payment service assumed in the startup-scenario are, at least currently, not covered by established abstraction frameworks and it is questionable to what extent such abstractions are realistic and reasonable at all for higher-level services implementing rather specific functionalities [19, pp. 34f].[9] For those cases not covered by abstraction technologies, maintaining multiple provider alternatives would thus require to constantly adopt implementations to the specifics of the respective alternatives, thus causing significant costs.

– *Non-Contractual Long-Term Relationships:* Especially for our second example massively employing such higher-level services, the third option of countervailing hold-up risks through long-term relationships based on reputation and trust is thus to be considered. Given the limited feasibility and establishedness of the two strategies discussed above, it could very well be argued that reputation and trust are actually the most relevant mechanisms currently keeping the cloud market running at all. Their general role for the cloud market is, however, not yet well-understood [18] and requires further examination. In particular, this also includes the question to what extent the intrinsically informal processes around reputation and trust can actually be formalized and proceduralized [24] for the domain of cloud computing.

In any case, countermeasures for diminishing hold-up risks will be of crucial relevance for the further development of the cloud market in general because only with effectual mechanisms available, it will in the long run be rational for a potential cloud user to not fall back on the fourth option suggested by agency theory: *vertical integration*. For the context of cloud computing, this would basically imply to *not* make use of external providers at all but rather stay with the traditional model of providing the respective resources in-house. The availability of effective mechanisms for countering hold-up risks is therefore clearly in the very interest of cloud providers, too.

5 Conclusion and Research Implications

As we have demonstrated, agency theory is highly suitable for representing, understanding and analyzing the conflicts of intersts arising between providers and users of cloud resources on a theoretically well-founded basis. While isolated aspects of agency theory have been applied to cloud computing before, the

[8] Like, e.g., the Apache toolkits jclouds (https://jclouds.apache.org/) and libcloud (https://libcloud.apache.org/)

[9] The same arguments do, of course, also apply to initiatives for cloud-related standardization which are basically also aimed at reducing the hold-up risk by establishing provider substitutability.

integrative perspective developed herein provides a comprehensive mapping of agency theory to cloud computing. This, in turn, allows us to evaluate concrete existing instruments like auditing or certifications in the light of corresponding abstract concepts and, in particular, to identify strategies suggested by agency theory which have no or only dysfunctional practical equivalents yet.

In particular, the fact of audit certificates being chronically misunderstood as proving facts while rather representing credible *signals* on his capabilities sent out by the provider has only briefly been addressed above and should in the future be investigated in more detail. Furthermore, agency-specific contract schemes are – except, maybe, for high-volume customers like governments or multinational corporations – not broadly established in the cloud market. Nonetheless, we showed that such advanced contract schemes are suggested by agency theory for countering *adverse selection*, *moral hazard* and *hold-ups* and that the respective abstract concepts could very well be transferred to the concrete domain of cloud computing. Further research on such advanced contract schemes for cloud computing thus seems highly valuable. As such contract schemes require the customer to be in a sufficiently strong bargaining position, this research should comprise policy aspects, too.

Regarding the further development of technology, we demonstrated that mechanisms providing credible information on the provider's capabilities and actual conduct could significantly reduce *screening* and *monitoring* costs and thus lead to more efficient outcomes or make cloud computing a viable option at all. In order to prevent a cloud provider from incorrectly attributing bad outcomes to detrimental external conditions, agency theory suggests these mechanisms to cover the externally given conditions of provider conduct, too. As such instruments would address the core problem of information asymmetries directly and at comparatively low operating costs, we see much potential for counteracting agency-related inefficiencies here. The whole field of providing credible digital evidence to the cloud user is thus a domain that we will explicitly focus our future research on – not only from the technical perspective but also regarding its legal and regulatory dimension.

As intended, however, the main result of this paper is the establishment of an agency perspective to cloud computing itself. Due to its theoretical well-foundedness and its broad practical applicability, this perspective opens up the whole field of agency-related research for application to the cloud computing domain. Certainly, this is not limited to the development of further novel approaches for counteracting well-recognized practical problems as exemplarily done herein. Instead, the abstract perspective established throughout this paper will also serve the prospective analysis of ongoing technological developments and thereby help to better foresee upcoming but yet unrecognized conflicts and challenges in the cloud market. This, in turn, will prove highly valuable from the perspective of policymaking, too.

Acknowledgement: The research presented in this paper has been partially funded by the European Commission in the context of the Research Framework Program Seven (FP7) project SECCRIT (Grant Agreement No. 312758, https://seccrit.eu).

References

1. Akerlof, G.A.: The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics 84(3), 488-500 (1970)
2. Anderson, R., Moore, T.: The Economics of Information Security. Science, 314(5799), 610–613 (2006)
3. Aubert, B.A., Patry, M., Rivard, S.: Assessing the Risk of IT Outsourcing. Proc. 31st HICSS, Vol. 6, 685–692 (1998)
4. Bless, R., Flittner, M.: Towards Corporate Confidentiality Preserving Auditing Mechanisms for Clouds. Proc. 3rd International Conference on Cloud Networking (CloudNet), to appear (2014)
5. Eisenhardt, K.M.: Agency theory: An Assessment and Review. Academy of Management Review 14(1), 57–74 (1989)
6. Elitzur, R., Gavious, A., Wensley, A.K.: Information Systems Outsourcing Projects as a Double Moral Hazard Problem. Omega, 40(3), 379–389 (2012)
7. Eymann, T., Knig, S., Matros, R.: A Framework for Trust and Reputation in Grid Environments. Journal of Grid Computing, 6(3), 225–237 (2008)
8. Friedman, M.: The Methodology of Positive Economics. In: D. M. Hausman (ed.), The Philosophy of Economics – An Anthology (second ed.)., 180–213. Cambridge, New York: Cambridge University Press (1994 [1953])
9. Friedman, A.A., West, D.M.: Privacy and Security in Cloud Computing. Center for Technology Innovation at Brookings (2010)
10. Furubotn, E.G., Richter, R.: Institutions & Economic Theory: The Contribution of the New Institutional Economics. 2nd ed., The University of Michigan Press (2005)
11. Hamlen, K.W., Thuraisingham, B.: Data Security Services, Solutions and Standards for Outsourcing. Computer Standards & Interfaces, 35(1), 1–5 (2013)
12. Hauff, S., Huntgeburth, J., Veit, D.: Exploring Uncertainties in a Marketplace for Cloud Computing: A Revelatory Case Study. Journal of Business Economics, 84(3), 441-468 (2014)
13. Holmström, B., Roberts, J.: The Boundaries of the Firm Revisited. Journal of Economic Perspectives 12(4), 73–94 (1998)
14. Jensen, M.C., Meckling, W.H.: Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. Journal of Financial Economics 3(4), 305–360 (1976)
15. Kim, K., Altmann, J.: Evolution of the Software-as-a-Service Innovation System Through Collective Intelligence. International Journal of Cooperative Information Systems, 22(3), 1340006 (2013)
16. Klein, B., Crawford, R.G., Alchian, A.A.: Vertical Integration, Appropriable Rents, and the Competitive Contracting Process. Journal of Law & Economics 21, 297–326 (1978)
17. Ko, R.K., Lee, B.S., Pearson, S.: Towards Achieving Accountability, Auditability and Trust in Cloud Computing. In: Abraham, A., Mauri, J.L., Buford, J.F., Suzuki, J., Thampi, S.M. (eds.) Advances in Computing and Communications, CCIS 193, 432–444 (2011)
18. Kshetri, N.: Privacy and Security Issues in Cloud Computing: The Role of Institutions and Institutional Evolution. Telecommunications Policy, 37(4), 372–386 (2013)
19. Kurze, T., Klems, M., Bermbach, D., Lenk, A., Tai, S., Kunze, M.: Cloud federation. Proc. Second International Conference on Cloud Computing, GRIDs, and Virtualization, 32–38 (2011)

20. Laffont, J.J., Martimort, D.: The Theory of Incentives - The Principal-Agent Model. Princeton University Press (2002)
21. Leimeister, S., Böhm, M., Riedl, C., Krcmar, H.: The Business Perspective of Cloud Computing: Actors, Roles and Value Networks. Proc. ECIS 2010, Paper 56 (2010)
22. Lenk, A., Pallas, F.: Modeling Quality Attributes of Cloud-Standby-Systems. In: Lau, K.K., Lamersdorf, W., Pimentel E. (eds.) Proc. ESOCC 2013, LNCS 8135, 49–63 (2013)
23. Lenk, A., Tai, S.: Cloud Standby: Disaster Recovery of Distributed Systems in the Cloud. In: Villari, M., Zimmermann, W., Lau, K.K. (eds.) Proc. ESOCC 2014, LNCS 8745, 32–46 (2014)
24. Macias, M., Guitart, J.: Cheat-Proof Trust Model for Cloud Computing Markets. In: Vanmechelen, K. , Altmann, J., Rana, O.F. (eds.) Proc. GECON 2012, LNCS 7714, 154-168 (2012)
25. Milgrom, P., Roberts, J.: Economics, Organization & Management. Prentice Hall (1992)
26. National Institute of Standards and Technology (NIST): Special Publication 800-145 – The NIST Definition of Cloud Computing. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf (2011)
27. National Institute of Standards and Technology (NIST): Special Publication 500-292 – NIST Cloud Computing Reference Architecture. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909505 (2011)
28. Spence, M.: Job Market Signaling. The Quarterly Journal of Economics 87 (3), 355–374 (1973)
29. Sunyaev, A., Schneider, S.: Cloud Services Certification. Communications of the ACM, 56(2), 33–36 (2013)
30. Taylor, M., Haggerty, J., Gresty, D., Hegarty, R.: Digital Evidence in Cloud Computing Systems. Computer Law & Security Review, 26(3), 304–308 (2010)
31. Vykoukal, J., Wolf, M., Beck, R.: Services Grids in Industry – On-Demand Provisioning and Allocation of Grid-Based Business Services. Business & Information Systems Engineering, 1(2), 177–184 (2009)