Cost Efficiency of Hybrid Cloud Storage: Shortening Acquisition Cycle to Mitigate Volume Variation

Completed Research Paper

Gabriella Laatikainen
Department of Computer Science and Information Systems
University of Jyväskylä
Jyväskylä, Finland
gabriella.laatikainen@jyu.fi

Oleksiy Mazhelis
Department of Computer Science and Information Systems
University of Jyväskylä
Jyväskylä, Finland
mazhelis@jyu.fi

Pasi Tyrväinen
Agora Center
University of Jyväskylä
Jyväskylä, Finland
pasi.tyrvainen@jyu.fi

Abstract

Hybrid cloud storage infrastructure, which combines cost-effective but inflexible private resources and flexible but premium-priced public cloud storage, allows organizations to operate cost-efficiently under demand volume uncertainty. The extant literature, however, offers a limited analytical insight into the effect that the variation of demand has on the cost-efficient mix of internal and external resources. This paper considers the storage capacity acquisition cycle, i.e. the interval at which the organization re-assesses and acquires additional resources, as a parameter shaping the optimal mix of resources. It introduces a model capturing the compound effect of the acquisition cycle and volume variation on the cost-efficiency of hybrid cloud storage. The model is analytically investigated to demonstrate its inherent regularities, and empirically evaluated in numerical experiments. The analysis indicates that shortening the acquisition cycle reduces the volume variability thus reducing the costs. The costs decrease further if shortening the cycle reduces the demand volume uncertainty.

Keywords: hybrid cloud storage; volume variation; reassessment interval; concurrent sourcing

Introduction

The multi-faced phenomenon of cloud computing brings together the technological advances in hardware virtualization, networking, and multi-tenancy, among other technological innovations, and blends them into the highly flexible shared computing resources that are accessible by multiple customers over the Internet (Babcock 2010, Armbrust et al. 2009). The emergence of cloud computing has changed the way the organizations purchase information technology (IT), as well as the role the IT function has in the organization, especially with respect to enabling innovativeness and creating new networked business models (Weinhardt et al. 2009, Schlagwein et al. 2014). In the very core of this multitude of impacts of cloud computing lies the utmost flexibility of shared computing capacity and related decrease in capital expenditures that are enabled, among other factors, by the increased availability and the decreased cost of communicating with external cloud computing and storage systems (Mazhelis and Tyrväinen 2012, Chen and Wu 2013). Without this flexibility, the transformation of the IT function and the emergence of innovative networked models would unlikely succeed (Venters and Whitley 2012, Schlagwein et al. 2014).

The hybrid cloud infrastructure, i.e. a combination of concurrently used private and public cloud infrastructure resources (Armbrust et al. 2009), offers further flexibility as well as cost savings (Mazhelis and Tyrväinen 2012). In this context, the public cloud refers to the computing, storage, and other infrastructure resources provided publicly by an infrastructure service provider to any organization willing to use these resources, on demand, over the Internet (Mell and Grance 2011). These infrastructure service providers often charge for the use of their resources based on the volume of usage. Whereas the pricing for small scale use is competitive, especially for small enterprises lacking IT competences, the high profit margins of the providers (Gauger 2013) make their services expensive for larger enterprises.

Hybrid cloud infrastructure can be seen as an instantiation of the so-called concurrent sourcing phenomenon that represents a simultaneous use of market contracting and vertical integration, i.e. producing and buying the same good or service (Parmigiani 2007, Parmigiani and Mitchell 2009, Mols 2010, Heide et al. 2013). Meanwhile, the hybrid cloud infrastructure as a concurrent sourcing phenomenon has attracted little attention in the IS research community. Indeed, whereas the concurrent sourcing has been widely studied outside of IS in the context of industries varying from automotive (Gulati et al. 2005) and metal forming (Parmigiani 2007) to fashion garments industries (Jacobides and Billinger 2006), to the best knowledge of the authors, the paper by Mazhelis and Tyrväinen (2012) is the only work where the hybrid cloud infrastructure is discussed as an instantiation of the concurrent sourcing. Therefore, the research enquiry on cloud-enabled flexibility, and in particular on the hybrid cloud and its impact on future cloud services is defined as one of the directions for further research by Venters and Whitley (2012).

A widely cited justification for the use of concurrent sourcing derives from transactional cost theories and neoclassical economics: namely, it is claimed to reduce production costs when firms face the so-called volume uncertainty, i.e., the difficulty in accurately predicting demand volumes (Adelman 1949, Parmigiani 2003, 2007). When the demand is fluctuating and it is difficult to forecast it accurately, the risk of diseconomies of scale due to unutilized excess capacity may be mitigated by serving the high probability component of demand with in-house resources and by using external suppliers for the peak demand only (Heide 2003, Puranam et al. 2013). Thus, the degree of uncertainty has an impact on how much to produce internally vs. how much to procure from external sources, and it determines the volume of cost savings that are attainable by sourcing concurrently. However, the empirical results on whether the use of concurrent sourcing is motivated by the presence of volume uncertainty are contradictory (Parmigiani 2003, Krzeminska et al. 2013).

Observe that the volume uncertainty reflects the difficulty in accurately predicting demand volumes and thus can be defined as the degree of (in)precision with which the volume is predicted (Parmigiani 2003, 2007). However, besides this prediction inaccuracy, also the natural variation in the volume of the demand referred to as *variability* – e.g., seasonal fluctuations – can be the reason of the diseconomies of scale in case the firm decides to invest in production for the peak demand (Puranam et al. 2013); note that, in principle, this natural variation may be perfectly predictable. Together, the volume uncertainty and variability comprise the *variation* in the volume of the demand (van Belle 2008). To the best knowledge of the authors, the variability aspect of the variation has not been explicitly considered in the concurrent sourcing literature.

A key question both in recent cloud computing and concurrent sourcing literature is the optimal mix of internal and external sourcing. Indeed, the cost-optimal mix of the private and public cloud resources has been one of the crucial themes in cloud computing literature, predominantly focusing on the dynamic allocation of available resources (Trummer et al. 2010, Shifrin et al. 2013, Wang et al. 2013, Altmann and Kashef 2014), and to a less extent on proactive resource provisioning (Weinman 2012, Mazhelis and Tyrväinen 2012). Likewise, in concurrent sourcing literature, multiple factors were found to affect the optimal mix including, among others, resource co-specialization, supplier selection, as well as the cost and benefits of producing in-house resources and buying from external parties (Puranam et al. 2013).

One of the parameters shaping the optimal mix of sourcing is the *acquisition cycle*, also referred to as the *reassessment interval* in the paper, which can be defined as the time period between the successive time points when the organization re-assesses its sourcing needs and acquires additional resources for in-house use (Laatikainen et al. 2014). For instance, if the company acquires additional private resources once a year, then the length of the reassessment interval is one year. Demand reassessment interval affects the degree of volume variation, since both the expected change of the demand and the difficulty of estimating it increase with the length of the interval. Therefore, it can be hypothesized (cf. Figure 1) that the demand reassessment interval, through its affect on the volume variation, impacts on how much to produce internally vs. how much to procure from external sources, and determines the volume of cost savings that are attainable by hybrid cloud storage.

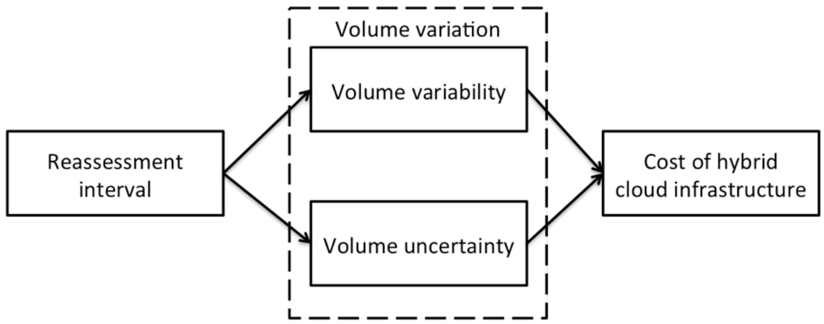

Figure 1. The impact of demand reassessment interval on the hybrid cloud costs

It shall be noticed that, while many research works in cloud computing literature focus on optimal dynamic resource allocation and scheduling in line with various requirements and constraints (e.g. Cerviño et al. 2013, Strebel and Stage 2010, Bjorkqvist et al. 2012, Zhang et al. 2014, Bittencourt and Madeira 2011, Wang et al. 2013, Vecchiola et al. 2012, Shifrin et al. 2013, Calheiros et al. 2011), the studied mechanisms provide little transparency on the effects that individual variables have on the optimal outcome, as well as on the inter-dependencies between different variables. Besides, to the best of the authors' knowledge, the effect of the demand reassessment interval and the volume variation on the costs of hybrid cloud storage has not been considered yet in the literature. Therefore, there is a need for analytical inquiry to increase our understanding of the economic effect that the reassessment interval and volume variation have on the cost of hybrid cloud infrastructure.

The paper studies the case of the hybrid cloud storage as a subset of hybrid cloud infrastructure whose popularity has increased dramatically in recent years, and is predicted to increase further (McClure 2014). The practical issue addressed in this paper is that of determining how much storage to provision from in-house resources and how much to procure on-demand from the public cloud resources. Whereas numerous factors including the need to deliver the required level of service and comply with applicable legislation have an affect on the cloud sourcing decisions (Fadel and Fayoumi 2013, Andrikopoulos et al. 2013), this paper focuses on the cost-efficiency of the resulting mix of resources which is a key factor affecting these decisions (Agarwala et al. 2011) and, thus, is a crucial issue faced by the cloud infrastructure practitioners (Weinman 2012, Altmann and Kashef 2014).

In earlier works on hybrid cloud computing, it has been shown that the cost-optimal time of using the public cloud computing resources is the inverse of the premium charged by the public cloud provider (Weinman 2012, Mazhelis and Tyrväinen 2011, 2012). Once the future workload is known or estimated, the cost-optimal time of using the public cloud can be found, and the cost-optimal portion of the workload to serve in-house can be estimated. For this, the fluctuating demand curve is re-arranged to be a monotonically non-decreasing function, and the maximum workload at the time when the in-house resources only are used indicates the volume of resources to be provisioned in-house (Mazhelis and Tyrväinen 2012). In case of storage, however, the fluctuations are rare; instead, the demand for storage usually is a monotonically non-decreasing function (Laatikainen et al. 2014). Nevertheless, during the time period between subsequent sourcing decisions, the same logic of determining the cost-optimal mix of in-house and external storage resources can be used, thus suggesting that the use of the hybrid approach can yield cost benefits also to the context of cloud storage resources.

The research question addressed in this paper can be formulated as follows:

How does the demand reassessment interval, through its affect on the volume variation experienced by the organization, impact the cost-efficient mix of internal and external sourcing in hybrid cloud storage?

The paper subscribes to the design science research (DSR) paradigm (Hevner et al. 2004, Peffers et al. 2008, Gregor and Hevner 2013) wherein an innovative artifact – in a form of an analytical model – is constructed and evaluated, in order to increase our understanding of the concurrent sourcing. In terms of the DSR knowledge contribution framework (Gregor and Hevner 2013), this paper reports a single design and evaluation cycle that aims at contributing to the creation of a nascent theory of concurrent IT sourcing. According to (Hevner et al. 2004), this cycle includes the build process, wherein an artifact in a form of conceptual-analytical model is constructed, and the evaluate process, wherein the built model is evaluated, i.e. (i) analytically investigated to demonstrate the inherent regularities of the model, and then (ii) empirically evaluated in simulation studies reflecting real-life scenarios.

The contribution of this paper is both theoretical and practical. From the theoretical perspective, the paper contributes to the cloud economics literature by analyzing the impact of the reassessment interval on the cost savings attainable through the use of hybrid cloud storage. The paper also fills the gap in concurrent sourcing literature by studying the hybrid cloud storage as an instantiation of the concurrent sourcing. From the practical perspective, the paper is expected to call practitioners' attention on the possibility of reducing hybrid cloud storage costs by reassessing their storage needs more often. In real life, companies may deal with both relatively easily predictable storage demand (e.g. periodic electricity metering data) and varying data volume that is difficult to predict (e.g. data originated from the customers). Therefore, both scenarios have been analyzed in the paper and thus, practitioners may use the findings in both cases.

The paper is organized as follows. Section 2 introduces the analytical model of hybrid cloud storage costs in cases of known and forecasted demand, and investigates the regularities inherited in the model. This model is further empirically evaluated in section 3 using simulation studies reflecting real-life scenarios. Finally, section 4 concludes the paper with the discussion of the theoretical and practical implications of the constructed model, and the outline of the directions for further research.

Cost Model of Hybrid Cloud Storage

Let us consider the cost of a hybrid cloud storage system, where a private and a public cloud infrastructure together serve the organization's storage demand. The system can be decomposed into two subsystems: the private subsystem provided by the in-house resources, and the public subsystem provided by the public cloud.

Assumptions

Before introducing the analytical model for hybrid storage costs, several assumptions have to be made. The core assumptions are listed below, while the others are introduced as appropriate later in the text.

1. First, we shall assume that the storage demand is a non-decreasing function in time. Indeed, as opposite to the demand for computing resources that often exhibits seasonal and other periodic fluctuations, the demand for storage tends to accumulate over time, due to the fact that newly created digital content only partially replaces the content already stored (Laatikainen et al. 2014). As a result, the digital universe as a whole grows 40% a year, according to a recent study by IDC (2014).

2. Second, we shall assume that the organization aims to achieve cost savings by allocating the cost-optimal amount of resources to the private subsystem. For this, the organization is assumed to periodically reassess its future storage needs and proactively acquire additional storage capacity. It shall be noted that, to minimize the storage-related costs, the organization may intentionally decide to acquire the storage resources to fulfill less than 100% of its future storage needs.

3. Third, for the sake of simplifying the analysis, we shall assume that each unit of data is atomic in the sense that (i) it bears the same level of criticality (i.e., the confidentiality, reliability, availability, and other considerations), and (ii) it is stored either on public or private portion of the system. In other words, it is assumed that no unit of data is distributed between private and public subsystems and neither is replicated in another cloud infrastructure. As a result, the interaction between the private and public subsystems can be assumed negligible.

4. Finally, the organization is assumed to allocate the storage on the private-first-public-second basis. Specifically, whenever a need to allocate storage emerges, the required storage space is allocated from the pool of the organization's in-house resources, provided that unused storage is still available in-house. However, when the demand for storage exceeds the capacities available in-house, the storage space to accommodate the excess demand is allocated from the public cloud.

Using these assumptions, let us consider the cost components comprising the hybrid storage cost model. Different cost components are relevant for the private and the public subsystems. For the private subsystem, the relevant cost constituents include the cost of hardware and software acquisition, integration, configuration, upgrade costs, as well as the recurring costs of renting floor space, power, bandwidth, and the cost of administration and maintenance. The overall cost of the private storage subsystem is thus a function of the demand, as well as its growth pattern and its predictability, the time interval between storage acquisitions, and the pricing of the needed equipment, software, and personnel, among other costs. Meanwhile, for the public cloud storage, the cost components include the usage-dependent costs, such as, in case of Amazon S3, the costs of storage capacity, data transfer, and input/output requests. Depending on the charging policy of the storage service provider, the cost of the storage may be determined by the maximum or average volume of storage occupied during the charging period. For instance, Amazon Web Services (AWS) offerings charge its customers based on the maximum storage capacity used in 12-hour intervals[1].

In case of a hybrid cloud storage, the process of acquiring, provisioning and paying for the necessary storage resources differs for the private and the public subsystems. On the one hand, the *private* subsystem's resources cannot be added instantly when the need comes, for there is a definite amount of time for the resources to be provisioned upon request — this time period is referred to as the provisioning interval (Weinman 2011c, 2012). Therefore, the organization has to manage the private subsystem proactively, i.e. it has to periodically estimate its future demand and in advance acquire and deploy the additional resources for the in-house storage infrastructure. The interval between the subsequent resource acquisitions based on the future demand estimates will be referred to as the *reassessment interval*. The cost of the private storage subsystem is incurred in the beginning of each reassessment interval and it, thus, depends on the maximum storage capacity and the estimation accuracy rather than on the actual use of storage resources.

On the other hand, the *public* subsystem's resources can be provisioned with a negligible delay. Thus, when the demand for storage exceeds the available private cloud capacity, the organization can acquire additional resources from the public cloud provider and then deploy the excess data to the public subsystem. As opposite to the private subsystem, the organization pays for the public subsystem's resources only when they are used, and only for the volume of actually used storage.

For the organization that aims to achieve cost savings by allocating the cost-optimal amount of resources to the private subsystem, this cost-optimal allocation depends on the forecasted or known storage demand, on the utility premium of the cloud provider, and on the length of the reassessment interval, that are the main subject of the analysis in this section. It shall be noted that the total cost of a hybrid cloud storage is also affected by many additional factors, such as the cost of taking a hybrid infrastructure into use, data transfer costs, pricing trends, volume discounts, cost savings achievable by storing only the provenance data and regenerating the rest when needed, or cost savings due to data transformations, among other factors. Combined, these factors are likely to have a complex, non-linear effect on the overall costs making them difficult to analyze (Mazhelis and Tyrväinen 2012). In order to simplify the analysis, in this paper it is assumed that these additional factors either have a minor effect or affect similarly the costs of both the private and the public subsystems, and hence are left outside of the scope of the analysis.

In the next subsections, first we introduce a general storage cost model where the storage needs are known or easy to estimate. Then we discuss the case when the demand is difficult to forecast, and introduce a cost model where demand forecasting errors are taken into account as well. The advantage of this model is its greater flexibility that makes it possible to account for the forecasting errors, at the expense of greater model complexity, which makes it less tractable analytically.

---

[1] See http://aws-portal.amazon.com/gp/aws/developer/common/amz-storage-usage-type-help.html.

General Hybrid Storage Cost Model

Let us first assume that the storage needs are known or easily predictable. Let us define the demand function $s(t) \mapsto \mathbf{R}$ that maps from time to quantity of needed resources. As stated above, due to the increasingly growing nature of storage needs, this function is assumed to be positive and increasing. Note that the form of the demand function $s(t)$ reflects the former aspect of the volume variation – *variability*, i.e. the non-stationary nature of the demand.

Let us consider the total cost of a hybrid storage during the reassessment interval of length w. Since, in a hybrid solution, the private and public subsystems are used in combination, the total hybrid cloud storage costs C_{H1} are the sum of the private costs C_o and public costs C_p.

First, let us evaluate the costs of owing the private storage subsystem C_o during the reassessment interval of length w. As described above, in the beginning of each reassessment interval, the company estimates the amount of resources needed during the following reassessment interval and acquires the necessary storage accordingly. Thus, having denoted the total cost of owning a unit of private storage capacity v over time t as $p_o(v)$, the cost of owning in-house capacity C_o can be estimated as:

$$C_o = v_o \ p_o(v_o) \ w, \tag{1}$$

where v_o is the maximum private storage capacity to be used within the next reassessment interval. In case the actual demand $s(t)$ exceeds v_o, the difference $s(t) - v_o$ will be served by using public cloud resources.

The cost of the public cloud storage subsystem C_p can be evaluated by calculating the costs of public storage over the period when the public cloud is used. Let $p_p(s(t))$ denote the price of a unit of storage per unit of time set by the public storage provider. We will assume for simplicity that the demand for the public resources is served immediately. Thus, the cost of public storage C_p accumulated over the reassessment interval of length w is

$$C_p = \int_{t_0}^{w} p_p(s(t)) \ s(t) \ dt - p_p(v_0) \ v_0 \ (w - t_0), \tag{2}$$

where t_0 is the time point when $s(t_0) = v_0$ and, therefore, $w - t_0$ is the length of the time interval, during which the public subsystem is used.

We shall assume that the price of a unit of public storage capacity is greater as compared with the cost of a unit of private storage. This is justified by the fact that the public storage provider charges a premium for the organization's flexibility in rapidly provisioning and de-provisioning the resources (Weinman 2011a); as a result, some organizations found it significantly less expensive to host own storage facilities than to use the storage capacity of Amazon, with the difference up to the factor of 26 (Nufire 2011). Thus, it can be written that $p_p(s(t)) = u(s(t)) \ p_o(s(t))$, where $u(s(t)) > 1$ is the utility premium ratio, or in short, utility premium of the public storage vendor. To simplify the further analysis, the utility premium shall incorporate (i) the cost of transferring the excess data to/from the public subsystem and (ii) the cost of transferring the cumulated public storage capacity to the private subsystem once the private capacity gets increased.

In order to make the analysis tractable, we shall further assume that the prices are not subject to volume discounts. Therefore, for brevity, we shall refer to $p_p(s(t))$ and $p_o(s(t))$ as p_p and p_o, respectively, and thus, Eq. (2) can be rewritten as follows:

$$C_p = u \, p_o \int_{t_0}^{w} s(t) \ dt - u \, p_o \, v_0 \ (w - t_0). \tag{3}$$

Thus, the total hybrid cloud storage costs C_{H1} are:

$$C_{H1} = p_o v_o w + u p_o \left(\int_{t_0}^{w} s(t)\ dt - v_0 \ (w - t_0) \right). \tag{4}$$

Let us now consider the cost-impact of shortening the reassessment interval. Specifically, let us consider the case when the reassessment interval is refined, i.e., it is divided into two adjacent reassessment intervals P_1 and P_2 of the lengths z and $w - z$, respectively (cf. Figure 2(a) and 2(b)). Let us mark the maximum private storage over the period P_1 with v_1, while over the period P_2 with v_2.

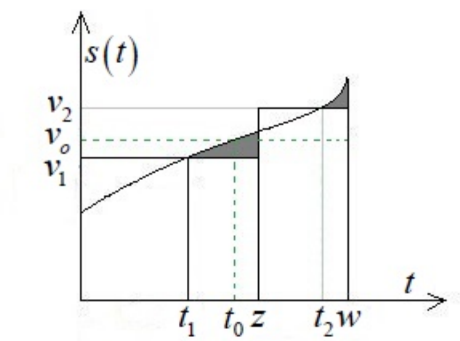

Figure 2(a). Demand for storage with a refinement of the reassessment interval exemplified for $z > t_0$; a cost- optimal allocation of resources to the private cloud is assumed

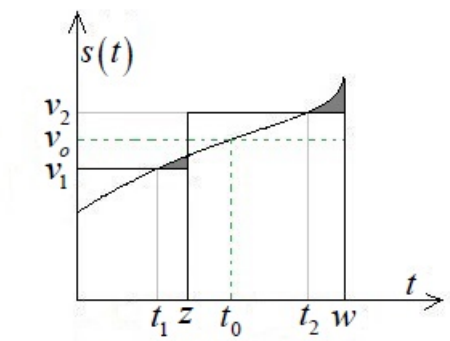

Figure 2(b). Demand for storage with a refinement of the reassessment interval exemplified for $z < t_0$; a cost- optimal allocation of resources to the private cloud is assumed

Using the same notations as introduced above, the total hybrid cloud storage costs in the period P_1 can be expressed as

$$C_{HP1} = p_o v_1 z + u p_o \left(\int_{t_1}^{z} s(t)\ dt - v_1 (z - t_1) \right), \tag{5}$$

where $s(t_1) = v_1$ and $z - t_1$ is the length of the time period where the public cloud is used.

Similarly, the total hybrid cloud storage costs for the period P_2 are

$$C_{HP2} = p_o v_1 (w - z) + u p_o \left(\int_{t_2}^{w} s(t)\ dt - v_2 (w - t_2) \right), \tag{6}$$

where $s(t_2) = v_2$ and $w - t_2$ is the length of the time period when the public cloud is used.

We will mark the total hybrid costs with C_{H2} for the case when the reassessment interval is refined. C_{H2} can then be calculated as the sum of the costs for reassessment intervals P_1 and P_2:

$$C_{H2} = C_{HP1} + C_{HP2}. \tag{7}$$

Let us define the cost difference function $f = C_{H1} - C_{H2}$. If $f > 0$, then refining the reassessment interval is worth cost-wise. Otherwise, if $f < 0$, the hybrid costs are increasing when the reassessment interval is divided into two smaller intervals.

The cost difference function can be written as:

$$\begin{aligned} f &= p_o v_0 w - u\, p_o v_0 \left(w - t_0\right) - p_o\, v_1 z + u\, p_o v_1 \left(z - t_1\right) - p_o v_2 \left(w - z\right) + u\, p_o v_2 \left(w - t_2\right) \\ &+ u\, p_o \left(\int_{t_0}^{w} s(t)\; dt - \int_{t_1}^{z} s(t)\; dt - \int_{t_2}^{w} s(t)\; dt \right). \end{aligned} \quad (8)$$

It can be seen that the sign of f depends on the utility premium charged by the public cloud provider, on the length of time period when the public cloud is used, on the demand function, and on the percentage of the actual demand that is allocated to the private cloud.

In order to simplify the analysis, we will take into account the assumption that, at the beginning of each reassessment interval, the organization acquires storage capacity to the private cloud so as to minimize the overall hybrid storage costs. According to Mazhelis and Tyrväinen (2012) and Weinman (2012), the cost-optimal portion of time to use public cloud is the inverse of the premium charged by the cloud provider (see Corollary 1 in (Mazhelis and Tyrväinen 2012)). It follows from this assumption that

$$t_0 = \frac{u-1}{u} w, \quad (9)$$

$$t_1 = \frac{u-1}{u} z, \quad (10)$$

$$t_2 = z + \frac{u-1}{u}\left(w - z\right). \quad (11)$$

As a result, it can be shown that the cost difference function simplifies to:

$$f = u\, p_o \left(\int_{\frac{u-1}{u} w}^{w} s(t)\; dt - \int_{\frac{u-1}{u} z}^{z} s(t)\; dt - \int_{z + \frac{u-1}{u}(w-z)}^{w} s(t)\; dt \right). \quad (12)$$

Proposition 1. *Assuming the allocation of cost-optimal amount of storage capacity to private cloud, re-evaluating the storage needs more often is always beneficial cost-wise, i.e.* $f > 0$.

Proof. Let $F(t)$ be an anti-derivative of $s(t)$. In this case, Eq. (12) can be rewritten as follows:

$$\begin{aligned} f &= u\, p_o \left[F(w) - F\left(\frac{u-1}{u} w\right) - F(z) + F\left(\frac{u-1}{u} z\right) - F(w) + F\left(z + \frac{u-1}{u}\left(w - z\right)\right) \right] \\ &= u\, p_o \left(\int_{\frac{u-1}{u} w}^{z + \frac{u-1}{u}(w-z)} s(t)\; dt - \int_{\frac{u-1}{u} z}^{z} s(t)\; dt \right). \end{aligned} \quad (13)$$

Having introduced an auxiliary function $g(t) = s\left(t + \frac{u-1}{u}\left(w - z\right)\right)$, Eq. (13) can be rewritten in a form:

$$f = u\, p_o \left(\int_{\frac{u-1}{u} z}^{z} g(t)\; dt - \int_{\frac{u-1}{u} z}^{z} s(t)\; dt \right). \quad (14)$$

Since $w > z$, it follows that $g(t) > s(t)$. Using the property of integral monotonicity, it further follows that $f > 0$, and, thus, re-evaluating the storage needs more often reduces the overall hybrid cloud storage cost. Q.E.D.

Cost Model Taking into Account Demand Forecasting Errors

Let us now turn to the case when the demand function $s(t)$ is not known, but needs to be estimated instead. The estimated demand $\hat{s}(t)$ is likely to diverge from its real value:

$$\hat{s}(t) = s(t)\,(1+\varepsilon), \tag{15}$$

where the estimation error $\varepsilon = \varepsilon(t_y - t_x)$ is a function of the length of the forecasting horizon $t_y - t_x$, i.e. the interval between the current time at which the prediction is made t_x and the time for which the prediction is made t_y. Note that this estimation error manifests the latter aspect of the volume variation — volume *uncertainty*, i.e. the inaccuracy with which the demand volumes are predicted.

A couple of assumptions need to be made about the estimation error function. The estimation error may be additive or multiplicative depending on the application; in this work, we assume the estimation error to grow as the amount of estimable storage increases, and hence we use ε to denote a multiplicative error that grows with the storage demand. The demand function is also assumed to have no or negligible bias.

Thus, the estimation error contaminates the estimates of required storage capacity v :

$$\hat{v}_0 = s(t_0)\,(1+\varepsilon_0), \text{ where } \varepsilon_0 = \varepsilon(t_0); \tag{16}$$

$$\hat{v}_1 = s(t_1)\,(1+\varepsilon_1), \text{ where } \varepsilon_1 = \varepsilon(t_1); \tag{17}$$

$$\hat{v}_2 = s(t_2)\,(1+\varepsilon_2), \text{ where } \varepsilon_2 = \varepsilon(t_2 - z). \tag{18}$$

Importantly, the error in the estimates of v_i spreads also into the "effective" value of t_i, denoted as $\hat{t}_i$, where $i = \{0,1,2\}$. For instance, if v is overestimated ($\varepsilon > 0$), it effectively means that the public cloud storage will start to be used later than originally envisioned, i.e., $\hat{t} > t$. Having denoted the error function impacting t_i as ξ_i, we can express the "effective" values of t_i as follows:

$$\hat{t}_0 = t_0(1+\xi_0), \text{ where } \xi_0 = \xi(t_0); \tag{19}$$

$$\hat{t}_1 = t_1(1+\xi_1), \text{ where } \xi_1 = \xi(t_1); \tag{20}$$

$$\hat{t}_2 = z + (t_2 - z)(1+\xi_2) = z + \frac{u-1}{u}(w-z)(1+\xi_2), \text{ where } \xi_2 = \xi(t_2 - z). \tag{21}$$

Several notes shall be made. First, ξ is assumed to act as a multiplicative error, too, in line with ε. Second, the errors ξ and ε are covarying, i.e., if $\varepsilon > 0$, then $\xi > 0$, and vice versa. Finally, it is important to observe that $\hat{s}(t) = s(\hat{t})$.

Taking into account the estimation errors introduced above, for the cost-optimal storage allocation as specified in Eqs. (9-11), the cost difference function f can be rewritten as:

$$\begin{aligned} f \; &= p_o s(t_0)(1+\varepsilon_0)w + up_o\left[\int_{\hat{t}_0}^{w} s(t)\,dt - s(t_0)(1+\varepsilon_0)(w-\hat{t}_0)\right] \\ &- p_o s(t_1)(1+\varepsilon_1)z + up_o\left[\int_{\hat{t}_1}^{z} s(t)\,dt - s(t_1)(1+\varepsilon_1)(z-\hat{t}_1)\right] \\ &- p_o s(t_2)(1+\varepsilon_2)(w-z) + up_o\left[\int_{\hat{t}_2}^{w} s(t)\,dt - s(t_2)(1+\varepsilon_2)(w-\hat{t}_2)\right]. \end{aligned} \tag{22}$$

Having opened $\hat{t}$, it can be rewritten in the form

$$\begin{aligned} f = f^* \quad &+\left(p_o s(t_0)(1+\varepsilon_0) w\xi_0 (u-1) - up_o \int_{t_0}^{\hat{t}_0} s(t)\,dt\right) \\ &-\left(p_o s(t_1)(1+\varepsilon_1) z\xi_1 (u-1) - up_o \int_{t_1}^{\hat{t}_1} s(t)\,dt\right) \\ &-\left(p_o s(t_2)(1+\varepsilon_2)(w-z)\xi_2 (u-1) - up_o \int_{t_2}^{\hat{t}_2} s(t)\,dt\right), \end{aligned} \qquad (23)$$

where $f^* = up_o\left[\int_{t_0}^{w} s(t)\,dt - \int_{t_1}^{z} s(t)\,dt - \int_{t_2}^{w} s(t)\,dt\right]$ is the value of f in case the estimation of demand is free of estimation error, as specified in Eq. (12). Based on Eq. (23), the difference $f - f^*$ can be expressed as

$$\begin{aligned} \Delta = f - f^* \quad &= \left[p_o s(t_0)(1+\varepsilon_0) w\xi_0 (u-1) - up_o \int_{t_0}^{\hat{t}_0} s(t)\,dt\right] \\ &-\left[p_o s(t_1)(1+\varepsilon_1) z\xi_1 (u-1) - up_o \int_{t_1}^{\hat{t}_1} s(t)\,dt\right] \\ &-\left[p_o s(t_2)(1+\varepsilon_2)(w-z)\xi_2 (u-1) - up_o \int_{t_2}^{\hat{t}_2} s(t)\,dt\right] = \alpha(\varepsilon_0,\xi_0) - \alpha(\varepsilon_1,\xi_1) - \alpha(\varepsilon_2,\xi_2), \end{aligned} \qquad (24)$$

where $\alpha(\varepsilon_0,\xi_0)$, $\alpha(\varepsilon_1,\xi_1)$, and $\alpha(\varepsilon_2,\xi_2)$ represent the terms in square brackets. It can be shown that $\alpha(\varepsilon_0,\xi_0)$, $\alpha(\varepsilon_1,\xi_1)$, and $\alpha(\varepsilon_2,\xi_2)$ are positive terms. Therefore, the sign of Δ depends on the interplay between them. Among other factors, the absolute values of the estimation errors determine the relative magnitude of these terms and therefore affect the sign of Δ.

In particular, if the error terms are declining with the length of the forecasting horizon (i.e., $|\varepsilon_0| > |\varepsilon_1|$, $|\varepsilon_0| > |\varepsilon_2|$, $|\xi_0| > |\xi_1|$, $|\xi_0| > |\xi_2|$, then it is likely that $\alpha(\varepsilon_0,\xi_0) >> \alpha(\varepsilon_1,\xi_1)$ and $\alpha(\varepsilon_0,\xi_0) >> \alpha(\varepsilon_2,\xi_2)$, and hence $\Delta = f - f^* > 0$. However, if the errors fail to decline with the length of the forecasting horizon, then $\alpha(\varepsilon_0,\xi_0) < \alpha(\varepsilon_1,\xi_1)$ and/or $\alpha(\varepsilon_0,\xi_0) < \alpha(\varepsilon_2,\xi_2)$, and hence $\Delta = f - f^* < 0$. In other words, if the refinement of the reassessment interval allows the volume uncertainty to be reduced, as reflected in the declining values of the estimation errors, then the economic benefit of the refinement is greater when the volume uncertainty is present. On the other hand, if the interval refinement fails to reduce the volume uncertainty, then the economic surplus due to the refinement gets smaller.

Numerical Experiments

In the previous section, our hybrid cloud storage cost model has been analytically investigated, with the aim to reveal its inherent properties. This section expands our effort at evaluating this model, by means of the numerical simulations taking into account the context of a real world organization.

Design of Numerical Experiments

Numerical simulation is a simulation that relies on numerical methods to quantitatively represent the evolution of a physical system (Colombo and Rizzo 2009). By analogy with laboratory experiments, these calculations with numerical models are referred to as numerical experiments (Bowman et al. 1993, Bacour et al. 2002, Winsberg 2003). Each numerical experiment studies how a particular combination of input parameters affects the output parameter of interest, and the set of the experiments are designed so as to maximize the amount of relevant information from a limited number of simulation runs (Hunter et al. 1978). In order to resemble reality, our simulation needs to rely on the real demand for storage experienced by a real-world organization, as well as on the real pricing for the private and public storage resources, as described below.

Demand for storage. The real demand for storage as experienced by the archival system of the National Center for Atmospheric Research/University Corporation for Atmospheric Research (NCAR/UCAR) is utilized in the experiments. This organization has been chosen for the study for three reasons. First, NCAR is an example of real-world organization that maintains and develops a large-scale storage solution whose storage demand and its growth can be considered as representative. Second, a long-time trace of storage massives in use at NCAR allows the historic development of storage needs to be observed. Finally, as opposite to commercial organizations that keep their infrastructure details in secret, the traces of storage growth at NCAR were publicly available for this study.

The historical development of the Archival System at NCAR is documented at the organization's website[2]. Over the years, a number of developments at NCAR were made to accommodate the growing needs for the storage, either by expanding the available storage massives or by replacing them with more efficient solutions. Due to a constant need to evolve while providing service continuity, multiple storage technologies have co-existed within the NCAR's archival systems.[3] At present, the archival system represents a combination of the new tape libraries of High Performance Storage System (HPSS)[4], and the legacy tape libraries maintained by a subcontractor[5]. This tape-based archival storage is used in concert with the GLobally Accessible Data Environment (GLADE), the centralized disk-based storage service using high-performance GPFS shared file system technology.[6]

For the purposes of this study, we use the storage metrics with monthly granularity that was kindly provided by the NCAR. In Figure 6, the growth profile of the NCAR's archival system during the period 1.9.1986-1.4.2014 is shown. As evidenced by the figure, the demand for data storage exhibited an exponential growth during these years, rising from 2TB in 1986 to over 30PB in 2014.

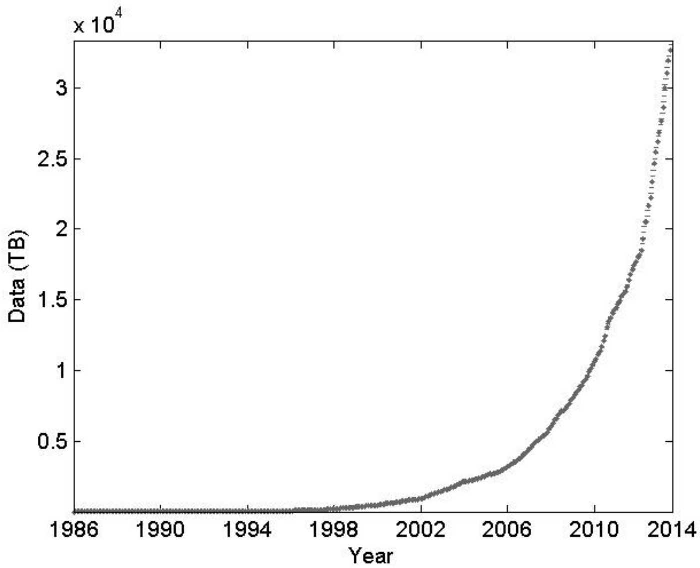

Figure 3. Growth of the NCAR archiving storage during 1986-2014

[2] See the annual reports of the Computational & Information Systems Laboratory that manages the archival system; these are available at http://nar.ucar.edu/.

[3] See the storage technologies used at NCAR by 2006 at http://www.cisl.ucar.edu/nar/2006/links/2.3.mss.lg.jsp.

[4] See https://www2.cisl.ucar.edu/docs/hpss.

[5] See http://www.nar.ucar.edu/2009/CISL/1comp/1.3.6.amstar.php.

[6] See https://www2.cisl.ucar.edu/resources/glade.

Public storage. The unit price of the public storage can be estimated for example by consulting the price list of Amazon Web Services (AWS) by Amazon, which is one of leading providers of public cloud infrastructure services (Leong et al. 2014).

Assuming the Reduced Redundancy Storage (RRS) is used as a public storage equivalent[7], it costs $0.024, $0.0236, and $0.0232 per GB per month to store the first TB, the next 49TB, and the next 450 TB of data, respectively. Further, transferring the data out of the cloud costs $0.12, $0.09, and $0.07 per GB for the first 10TB, next 40TB, and the next 100TB, respectively. It shall be noted that the request pricing is not taken into account for simplicity, for the contribution of the request-based charges to the overall cost is rather modest in case of the archival solutions.

Instead of RRS, Amazon Glacier could have been used as an inexpensive public tape storage equivalent that only charges $0.01 to store 1GB for a month. However, significant costs are incurred for transferring the data out of the service since, in addition to the data transfer fee above, the transfer may incur a significant retrieval fee that depends on the desired retrieval rate; furthermore, deleting the files stored for less than three months incurs fees, too. All this makes the use of Glacier economically inefficient in case the data is stored for short periods of time, as considered in the paper.

Private storage. The unit price of the private storage for newly designed storage solutions can be approximated using the costs incurred by Backblaze (Nufire 2011). Specifically, in order to provision a PB of storage, back in 2011 Backblaze reportedly was spending $94563 over three years for hardware, space, power, bandwidth, and maintenance, which corresponds to $2.57 per TB per month. By 2014, the cost of storage hardware declined from $0.055 per GB in 2011 to $0.0517 per GB in 2014, owing to more efficient design and declining component prices (Klein 2014); however, we will assume the total cost per TB unchanged due to a likely increase in other costs, such as rents and labor costs.

It should be noted that, along with the storage hardware, also the software solutions for managing the storage (such as IBM Tivoli Storage Manager) along with related services are likely to be needed, thus increasing the cost of the storage solution further. However, we will assume that these software and service costs are minor when compared to the other storage-related costs, and hence may be neglected for the sake of simplicity.

Utility premium. The value of the utility premium u varies depending on the type and the volume of storage to be provisioned, as well as on the pricing set by the public cloud storage provider and the cost-efficiency of the private solution. For instance, storing 100 TB of data on disk over six-months period cost: (i) $1539 if the data is stored in-house using Backblaze's type of storage, and (ii) $22878 if the data is stored in Amazon Reduced Redundancy Storage and transferred therefrom at the end of the storage period; this results in the utility premium value of $22878/$1539 = 14.9.

Storing the same volume of data on tape will cost: (i) $2550 if the in-house tape storage is used as described in (Reine and Kahn 2013), and (ii) $16786 if Amazon Glacier is used instead[8], thus resulting in the utility premium of $16786/$2550 = 6.6.

It shall be noted that the costs of the private storage solutions may be underestimated. First, additional labor costs are incurred to design, implement, and maintain growing in-house storage facilities. Second, additional costs will be required if higher redundancy level is needed, especially if geographically distributed facilities are deployed. Finally, for lower-scale data storage solutions, the absence of volume discounts is likely to increase the prices for the components. Due to these and possibly other factors, the value of the utility premium may be lower, but still notably greater than one[9].

---

[7] The details of RRS pricing are available at http://aws.amazon.com/s3/; the prices used in the research are for US Standard region and are valid on 2.8.2014.

[8] We further assume that the data is transferred from Amazon Glacier to the in-house storage solution at the end of the storage period, reserving two weeks for the retrieval.

[9] Otherwise, the in-house storage solutions would not be economically justifiable, as was shown analytically by Weinman (2011a).

It shall be also noted that the prices for storage components and storage services tend to decline with time (Walker et al. 2010, Reine and Kahn 2013, Jackson 2014). However, we will assume that approximately the same decline rate applies to both the private and the public portions. Likewise, it shall be noted that the time value of money is not taken into account in the cost estimates, since, within a single reassessment interval, the present value of money changes insignificantly, and thus has a limited effect on the overall costs.

In order to evaluate the total cost of hybrid storage and the cost difference between the cases when reassessment happens once or twice during the given time period, the following steps are taken:

1. First, the storage demand function is determined and its parameters are estimated.
2. The unit price of public and private storage is estimated. If needed, the utility premium can then be estimated by dividing the unit price of public storage by the unit price of private one.
3. The maximum private storage capacity to be used within the reassessment interval is estimated.
4. In case the reassessment happens only once at the end of the given time interval, the total cost of hybrid storage within the time interval is evaluated based on eq. (4).
5. If the demand is reassessed two times during the given time interval, the total cost of hybrid storage within the time interval is evaluated based on equations (5), (6) and (7).
6. Finally, the cost difference function is calculated based on eq. (8).

In the simulation below, we compare the costs incurred by the organization facing the growing demand for storage as experienced by NCAR (i) in case the organization is re-estimating the storage needs and acquiring additional in-house resources on a yearly basis, and (ii) in case the organization is doing the reassessment twice a year.

Results

Let us consider the compound effect of the reassessment interval and the volume variation – reflected in changing demand function and its forecasting inaccuracy – on the total cost of the hybrid cloud storage.

Known demand for storage. Let us first estimate the cost of hybrid cloud storage under the assumption that the future changes of the demand for storage are known in advance. In this case, as explained in the preceding section, both the time of using public cloud resources and the volume of the private storage to be acquired can be set to minimize the overall cost.

The cost estimate includes both the cost of storage C_{H1} (calculated with the parameters described above based on eq. (4)) as well as the data transfer cost. The data transfer cost is estimated based on the pricing of Amazon EC2, assuming that 5% of the stored data is requested and transferred monthly, and that the whole volume of the data in the public subsystem is transferred to the private subsystem. Furthermore, the effective value of the utility premium is estimated based on the total monthly volumes of storage. This estimate varies between 2.27 and 10.29; the median value of u = 2.88 is therefore used in the cost calculations, unless explicitly specified otherwise.

In Figure 4, the total yearly cost of a hybrid cloud storage is shown for the reassessment interval of six and twelve months. In order to make the figure more readable, only the costs over the last six years (2008–2013) are shown. As evidenced by the figure, the total hybrid storage costs are lower if the organization reassesses its storage needs more often, i.e., once every six months instead of once a year. This is in line with Proposition 1 that claims more frequent re-evaluation of storage needs to be cost-beneficial.

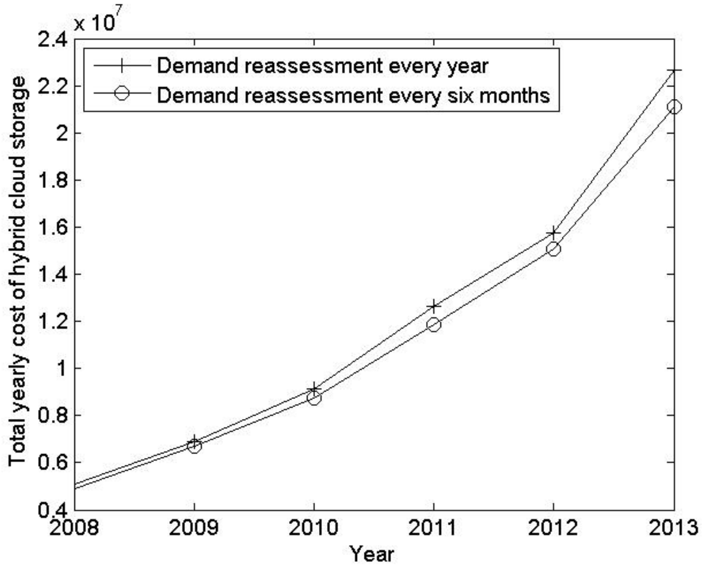

Figure 4. The total yearly cost of a hybrid cloud storage for the reassessment interval of six and twelve months

Forecasted demand for storage. Let us now turn to the case when the future demand for storage is not known and is therefore forecasted based on the traces of demand observed in the past. As in the previous experiment, here we consider the compound effect of the reassessment interval and the volume variation on the total cost of the hybrid cloud storage. However, whereas in the preceding experiment the volume variation was limited to the changing demand function, in this numerical experiment, the more realistic settings are studied by considering the volume variation as reflected both in changing demand function and its forecasting inaccuracy.

Specifically, this experiment relies on forecasting the future demand at the beginning of every reassessment period based on the historical data. The forecasting is performed by using the non-linear least square fitting to estimate the parameters of an exponential growth function and the estimation errors are calculated based on the forecasted and real demand. The forecasted demand largely follows the original demand, although there are periods when the demand is under- or over-estimated. In the results presented next, we exclude the data for the first year, since there was no historical data to base the forecast on. We will also exclude the data for the last year (2014), since the available data for this year is incomplete. It is important to notice that the estimates of the storage costs occurred in earlier years do not take into account the prices relevant at that time; therefore, these estimates shall not be treated as absolute values, but the (sign of the) difference between the costs without and with refining the reassessment interval should be considered instead.

As was analytically shown in the previous section, the effect of the refinement of reassessment interval on the cost savings depends on whether the forecasting inaccuracy decreases with the refinement. Indeed, as Table 1 manifests, the change in the cost savings Δ (eq. (24)) greatly correlates with the change in the estimation errors: for 21 years out of 26, the sign of Δ matches the sign of $\varepsilon_0 - \varepsilon_2$. Furthermore, in the cases when $\varepsilon_0 \approx \varepsilon_2$ (i.e., more formally, when $|\varepsilon_0 - \varepsilon_2| < 0.01$), the sign of Δ depends on the change of the time estimation error ξ: when ξ declines or remains the same after refinement ($\xi_0 - \xi_2 \leq 0$), the cost difference increases (years 2001 and 2002), whereas for the years when the error increases ($\xi_0 - \xi_2 > 0$), the cost difference declines (years 2011 and 2012). Thus, in line with the analytical considerations above, it can be observed that, while the refinement of the reassessment interval does cut the cost of the hybrid storage, the magnitude of the cost cut further depends on the inaccuracy of the demand forecasting; in particular, when the refinement allows the estimation errors to be reduced, the cost reduction increases, and decreases otherwise.

Year	ε_0	ε_1	ε_2	ξ_0	ξ_1	ξ_2	Δ
1988	0,620	0,325	0,192	0,50	0,50	0,50	2062,88
1989	0,096	0,097	0,044	0,50	0,25	0,50	186,97
1990	0,158	0,067	0,126	0,50	0,50	0,50	501,39
1991	0,136	0,107	0,056	0,50	0,50	0,50	259,59
1992	0,036	0,010	0,023	0,13	0,00	0,00	63,12
1993	0,019	0,007	0,050	0,13	0,00	0,50	-144,51
1994	0,104	0,074	0,067	0,50	0,50	0,50	183,90
1995	0,091	0,069	0,049	0,38	0,50	0,50	123,04
1996	0,045	0,036	0,032	0,25	0,25	0,25	90,20
1997	-0,013	0,036	-0,027	-0,13	0,25	-0,50	-624,79
1998	-0,031	-0,007	-0,273	-0,25	-0,25	-0,50	-20252,56
1999	-0,292	-0,253	-0,143	0,50	-1,00	-1,00	15570,07
2000	-0,026	-0,058	0,023	-0,25	-0,50	0,00	-4537,85
2001	0,066	0,047	0,064	0,25	0,25	0,25	2478,82
2002	0,112	0,055	0,117	0,50	0,25	0,50	5247,23
2003	-0,005	0,040	-0,032	-0,13	0,00	-0,25	-8441,47
2004	0,013	-0,002	0,002	0,00	-0,25	0,00	-580,63
2005	0,159	0,085	0,138	0,50	0,50	0,50	55074,42
2006	0,226	0,167	0,150	0,50	0,50	0,50	115820,24
2007	0,093	0,095	0,032	0,25	0,50	0,25	-10105,75
2008	0,005	-0,015	0,018	0,00	-0,25	0,00	-16279,51
2009	-0,005	-0,023	0,009	-0,25	-0,50	0,00	-27617,11
2010	0,038	0,016	0,019	0,13	0,00	0,00	24527,86
2011	-0,023	0,016	-0,030	-0,13	0,00	-0,50	-60162,71
2012	0,050	0,024	0,043	0,25	0,50	0,50	-41240,11
2013	0,000	0,048	-0,073	-0,13	0,00	-0,50	-527163,54

Table 1. Estimation errors and the change in the cost savings

Discussion and concluding remarks

The core benefit of cloud computing can be attributed to the business flexibility achievable by converting the capital IT expenditures to on-demand operational expenditures. As compared with traditional in-house IT infrastructure, this provides both a low-cost option to scale a business and the ability to make frequent and rapid changes in business models.

In order to deliver such flexibility, public cloud providers have to be capable of providing scalability for services whose demand grows by factor of 100 or 1000 in a few months; in response, these providers may request utility premium as high as 2 to 20 times the in-house costs. Such a high premium is well tolerable for small firms with no in-house IT capability; for larger enterprises with in-house IT capabilities, however, the use of the resources available in-house may prove less expensive in the longer term.

The hybrid cloud solutions combining fixed in-house cloud resources and flexible public cloud resources provide cost optimal solutions when the volume variation is high. In such cases, the cost can be minimized by serving the high probability component of demand with in-house resources and by using public cloud for the peak demand only. Importantly, the need to communicate between in-house cloud resources and public cloud resources reduces the benefit of using public cloud resources. This implies that the cloud storage associated with cloud computing capacity may be a critical factor limiting the benefits of cloud adoption.

This paper contributes to the cloud economic literature by analyzing the hybrid cloud storage that combines in-house and public cloud storage. In particular, the paper has analyzed the impact of refining the reassessment interval on the cost savings attainable by using hybrid cloud storage. It has been shown that the magnitude of the cost savings depends on two distinct dimensions of the volume variation, i.e. both on non-stationary (demand variability) and on nondeterministic nature (demand volume uncertainty) of the demand volume. This analysis proves that shortening the reassessment interval and the acquisition of public cloud storage capacity allows the volume variability to be reduced, yielding a reduction of the overall costs. Specifically, the findings are that:

- The refinement of the reassessment interval does indeed reduce the cost of hybrid cloud storage;
- If the refinement allows the forecasting inaccuracy to be reduced, then the economic benefit of the refinement further increases, and decreases otherwise.

This paper also sheds some light on the economical viability of organizational transformation towards cloud adoption, by re-engineering or replacing the organization's information systems to become cloud-enabled. Indeed, the economical viability of cloud transformation can be questioned, since the information system renewal incurs costs, since using cloud enabled software has some performance penalty (5-15%), and especially due to high premium costs of public cloud offerings (2-20 times in-house costs). When we assume no costs for such a cloud transformation, this paper's analytical model explains how the optimal cost of cloud storage can be achieved by using concurrent sourcing, i.e. a combination of in-house private cloud and limited volume of public cloud. In the numerical example from a conventional storage organization, cutting annual resource acquisition cycle to 6 months would provide 15% savings of cloud storage costs assuming no costs for speeding up the storage acquisition cycle and executing it twice a year. This 15% saving is thus the incentive for the enterprise to acquire the capability needed for concurrent sourcing in cloud environments, that is, adopting cloud platform internally to be able to gain the cost benefit of the hybrid cloud solution through concurrent sourcing. In short, the cost benefit of flexibility in concurrent sourcing motivates the firm to carry out a cloud transformation.

This case considered in this paper is further connecting the concurrent sourcing to the literature of strategic flexibility (Sanchez 1995), especially in resource flexibility (Sanchez 2004) and real options (Brydon 2006), as well as back to the transaction cost economics (Williamson 1985). As long as we assume no extra cost from repeating the capacity estimation and acquisition cycle more often, the faster cycle provides an option to minimize sum of volume diseconomies and utility premium of resource vendors, as well as revise the acquisition plan to mitigate estimation errors. The additional costs related to extra acquisition cycles can in this case be compared with the savings representing 15% or $100K for halving the cycle for the case organization. In an organization of this size, the benefits exceed the costs, while for a small organization; the 15% savings could easily be smaller than the resource acquisition costs thus recommending the use of an annual capacity acquisition cycle.

While this paper showed that the cost benefit of flexibility in concurrent sourcing motivates an enterprise like the case organization to adopt hybrid cloud approach, which requires in-house cloud capability and thus cloud transformation of incumbents, the case could be somewhat different in case of new ventures with extreme volume variation. They need not carry the legacy IT with them and can build cloud-enabled IT infrastructure from the beginning, thus dropping the cost of the transformation. In case of small firms the overhead of establishing in-house information systems, maintaining in-house servers etc. may also be a capital intensive cost factor, which can or even must be avoided. Thus they may have the tendency to use public cloud only until the storage demand has increased substantially and the cost benefit of hybrid cloud approach overrules the capital spending and inflexibility of the in-house storage. Future research, therefore, should address the economical view to flexibility and premium costs of public cloud only approach compared to the cost optimal hybrid cloud solution in case of fast growing small enterprises with high volume variation.

Future work shall also focus on enhancing and further elaborating the cost model to closer resemble the reality. Specifically, storage request arrival can be treated as stochastic process, and this stochasticity can be taken into account in further work, e.g., by defining the estimation errors as random variables with certain probability distributions. Besides, in order to address the needs of practitioners, the model can be re-formulated as the optimization problem with the acquisition cycle as the decision variable.

Acknowledgements

This work was partly supported by TEKES as part of the Need for Speed (N4S) Program of DIGILE (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT and digital business).

References

Adelman, M. A. 1949. "The large firm and its suppliers.", *The Review of Economics and Statistics* (31:2), pp. 113–118.

Agarwala, S., Jadav, D., Bathen, L.A. 2011. "iCostale: Adaptive cost optimization for storage clouds," in *Proceedings of 2011 IEEE International Conference on Cloud Computing,* IEEE, pp. 436–443.

Altmann, J., and Kashef, M.M. 2014. "Cost model based service placement in federated hybrid clouds," *Future Generation Computer Systems (*41), pp. 79–90.

Andrikopoulos, V., Song, Z., and Leymann, F. 2013. "Supporting the migration of applications to the cloud through a decision support system," in *Proceedings of the 6th IEEE International Conference on Cloud Computing*. IEEE, pp. 565–572.

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. 2010. "A view of cloud computing," *Communications of ACM (*53), pp. 50–58.

Babcock, C. 2010. "Management Strategies for the Cloud Revolution: How Cloud Computing Is Transforming Business and Why You Can't Afford to Be Left Behind", *McGraw-Hill Education*.

Bacour, C., Jacquemoud, S., Tourbier, Y., Dechambre, M., and Frangi, J.-P. 2002. "Design and analysis of numerical experiments to compare four canopy reflectance models," *Remote Sensing of Environment (*79:1), pp. 72–83.

Bittencourt, L.F. and Madeira, E. R. M. 2011. "HCOC: a cost optimization algorithm for workflow scheduling in hybrid clouds", *Journal of Internet Services and Applications* (2:3), pp. 207–227.

Bjorkqvist, M., Y Chen, L. and Binder, W. 2012. "Cost-driven service provisioning in hybrid clouds", in *Proceedings of the 5th IEEE International Conference on Service-Oriented Computing and Applications (SOCA)*. IEEE, pp. 1–8.

Bowman, K. P., Sacks, J., and Chang, Y.-F. 1993. "Design and analysis of numerical experiments," *Journal of the atmospheric sciences (*50:9), pp. 1267–1278.

Brydon, M. 2006. "Evaluating strategic options using decision-theoretic planning," *Information Technology and Management (*7), pp. 35–49.

Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. AF. and Buyya, R. 2011. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", *Software: Practice and Experience* (41:1), pp. 23–50.

Cerviño, J., Rodríguez, P., Trajkovska, I., Escribano, F., and Salvachúa, J. 2013. "A cost-effective methodology applied to video conference services over hybrid clouds." *Mobile Networks and Applications* (18:1), pp. 103–109.

Chen, P., and Wu, S. 2013. "The impact and implications of on-demand services on market structure," *Information Systems Research (*24:3), pp. 750–767.

Colombo, Simone P., Christian L. Rizzo. 2009. *Numerical Simulation Research Progress*. Nova Science Publishers.

Fadel, A.S., and Fayoumi, A.G. 2013. "Cloud resource provisioning and bursting approaches," in *Proceedings of the 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD),* IEEE, pp. 59–64.

Gauger, A. 2013. "Amazon's 'mountain of margin' in cloud services: over 80% profit," VentureBeat, available online at http://venturebeat.com/2013/09/05/amazons-mountain-of-margin-in-cloud-services-over-80-profit/ (retrieved March 10, 2015).

Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *Management Information Systems Quarterly* (37 :2), June, pp. 337–355.

Gulati, R., Lawrence, P.R., and Puranam, P. 2005. "Adaptation in vertical relationships: beyond incentive conflict," *Strategic Management Journal (*26:5), pp. 415–440.

Heide, J.B. 2003. "Plural governance in industrial purchasing," *Journal of Marketing (*67:4), pp. 18–29.

Heide, J.B, Kumar, A., and Wathne, K.H. 2013. "Concurrent sourcing, governance mechanisms, and performance outcomes in industrial value chains," *Strategic Management Journal* (35:8), pp. 1164–1185.

Hevner, A., March, S., Park, J., and Ram, S. 2004, "Design Science Research in Information Systems," *Management Information Systems Quarterly* (28 :1), March, pp. 75–105.

Hunter, William Gordon, J Stuart Hunter, EP George. 1978. "Statistics for experimenters: an introduction to design, data analysis, and model building". *Wiley New York*.

IDC 2014. "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", *EMC Digital Universe with Research & Analysis by IDC*, Executive Summary, available from http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm (retrieved March 10, 2015).

Jackson, J. 2014. "Price war! Amazon cuts cloud costs to counter Google," *IDG News Service, Computerworld*, available from http://www.computerworld.com/article/2489105/cloud-computing/price-war--amazon-cuts-cloud-costs-to-counter-google.html (retrieved on 30.9.2014).

Jacobides, M.G., Billinger, S. 2006. "Designing the boundaries of the firm: From "make, buy, or ally" to the dynamic benefits of vertical architecture," *Organization Science (*17:2), pp. 249–261.

Klein, A. 2014. "Storage pod 4.0: Direct wire drives – faster, simpler and less expensive," *Backblaze blog*, available from https://www.backblaze.com/blog/backblaze-storage-pod-4/ (retrieved on 29.9.2014).

Krzeminska, A., Hoetker, G., and Mellewigt, T. 2013. "Reconceptualizing plural sourcing," *Strategic Management Journal (*34:13), pp. 1614–1627.

Laatikainen, G., Mazhelis, O., and Tyrväinen, P. 2014. "Role of acquisition intervals in private and public cloud storage costs," *Decision Support Systems (*57), pp. 320–330.

Leong, L., Toombs, D., Gill, B., Petri, G., and Haynes, T. 2014. "Magic Quadrant for Cloud Infrastructure as a Service," *Gartner*, 28 May 2014.

Mazhelis, O., and Tyrväinen, P. 2011. "Role of data communications in hybrid cloud costs," in *Proceedings of the 37th EUROMICRO Conference on Software Engineering and Advanced Applications,* S. Biffl, M. Koivuluoma, P. Abrahamsson, and M. Oivo (Eds.), IEEE, pp. 138–145.

Mazhelis, O., and Tyrväinen, P. 2012. "Economic aspects of hybrid cloud infrastructure: User organization perspective," *Information Systems Frontiers (*14:4), pp. 845–869.

McClure, T. 2014. "Hybrid model offers more secure file sync and share," *Storage Magazine (*13:1).

Mell, P., and Grance, T. 2011. "The NIST definition of cloud computing," *Special publication 800-145, National Institute of Standards and Technology*, available from http://www.csrc.nist.gov/groups/SNS/cloud-computing/.

Mols, N.P. 2010. "Economic explanations for concurrent sourcing," *Journal of Purchasing and Supply Management (*16:1), pp. 61–69.

Nufire, T. 2011. "Petabytes on a budget v2.0: Revealing more secrets," *Backblaze blog*, available from http: //blog.backblaze.com/2011/07/20/petabytes-on-a-budget-v2-0revealing-more-secrets/, (retrieved on 29.9.2014).

Parmigiani, A.E. 2003. "Concurrent sourcing: When do firms both make and buy?" Ph.D. thesis, Duke University.

Parmigiani, A.E. 2007. "Why do firms both make and buy? An investigation of concurrent sourcing," *Strategic Management Journal (*28:3), pp. 285–311.

Parmigiani, A.E., and Mitchell, W. 2009. "Complementarity, capabilities, and the boundaries of the firm: the impact of within-firm and interfirm expertise on concurrent sourcing of complementary components," *Strategic Management Journal (*30:10), pp. 1065–1091.

Peffers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. 2008, "A Design Science Research Methodology for Information Systems Research," *Journal of MIS* (24:3), pp. 45–77.

Puranam, P., Gulati, R., and Bhattacharya, S. 2013. "How much to make and how much to buy? An analysis of optimal plural sourcing strategies," *Strategic Management Journal (*34:10), pp. 1145–1161.

Reine, D., and Kahn, M. 2013. "Revisiting the search for long-term storage – a TCO analysis of tape and disk," *The Clipper Group Calulator bulletin*, available from http://www.clipper.com/Clipper_Tutorials_Index.htm (retrieved on 29.9.2014).

Sanchez, R. 1995. "Strategic flexibility in product competition," *Strategic Management Journal (*16:S1), pp. 135–159.

Sanchez, R. 2004. "Understanding competence-based management: Identifying and managing five modes of competence," *Journal of Business research (*57:5), pp. 518–532.

Schlagwein, D., Thorogood, A., and Willcocks, L.P. 2014. "How commonwealth bank of Australia gained benefits using a standards-based, multi-provider cloud model," *MIS Quarterly Executive (*13:4), pp. 209–222.

Shifrin, M., Atar, R., and Cidon, I. 2013. "Optimal scheduling in the hybrid-cloud," in *Proceedings of IFIP/IEEE International Symposium on Integrated Network Management,* IEEE, pp. 51–59.

Strebel, J., and Stage, A. 2010. "An economic decision model for business software application deployment on hybrid cloud environments", in *Multikonferenz Wirtschaftsinformatik 2010*, M. Schumann, L.M. Kolbe, M.H. Breitner, and A. Frerichs (eds), pp. 195–206.

Trummer, I., Leymann, F., Mietzner, R. and Binder, W. 2010. "Cost-optimal outsourcing of applications into the clouds," in *Proceedings of IEEE Second International Conference on Cloud Computing Technology and Science, pp*. 135 –142.

van Belle, G. 2008. *Statistical Rules of Thumb*. Wiley & Sons.

Vecchiola, C, Calheiros, R. N., Karunamoorthy, D. and Buyya, R. 2012. "Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka", in *Future Generation Computer Systems (*28:1), pp. 58–65.

Venters, W., and Whitley, E.A. 2012. "A critical review of cloud computing: researching desires and realities," *JIT (*27:3), pp. 179–197.

Walker, E., Brisken, W., and Romney, J. 2010. "To lease or not to lease from storage clouds," *Computer (*43), pp. 44–50.

Wang, W.-J., Chang, Y.-S., Lo, W.-T., and Lee, Y.-K. 2013. "Adaptive scheduling for parallel tasks with QoS satisfaction for hybrid cloud environments," *The Journal of Supercomputing (*66:2), pp. 783–811.

Weinhardt, C., Anandasivam, A., Blau, B., Borissov, N., Meinl, T., Michalk, W., and Stößer, J. 2009. "Cloud computing – a classification, business models, and research directions," *Business & Information Systems Engineering (*1:5), pp. 391–399.

Weinman, J. 2011a. "Cloudonomics: A rigorous approach to cloud benefit quantification," *The Journal of Software Technology (*14), pp. 10–18.

Weinman, J. 2011c. "*Time is money: The value of "on-demand"*," Working paper, available from http://www.joeweinman.com (retrieved on February 28, 2012).

Weinman, J. 2012. *Cloudonomics: The Business Value of Cloud Computing*. John Wiley & Sons.

Williamson, O.E. 1985. *The Economic Institutions of Capitalism*. The Free Press, New York.

Winsberg, E. 2003. "Simulated experiments: Methodology for a virtual world," *Philosophy of science (*70:1), pp. 105–125.

Zhang, H., Li, P., Zhou, Z., Wu, J. and Yu, X. 2014. "A privacy-aware virtual machine migration framework on hybrid clouds". *Journal of Networks* (9:5), pp. 1086–1095.