This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Laatikainen, Gabriella; Mazhelis, Oleksiy; Tyrväinen, Pasi

Title: Cost benefits of flexible hybrid cloud storage : Mitigating volume variation with shorter acquisition cycle

Year: 2016

Version: Submitted version (Preprint)

Cost Benefits of Flexible Hybrid Cloud Storage: Mitigating Volume Variation with Shorter Acquisition Cycle

Gabriella Laatikainen[a,*], Oleksiy Mazhelis[a], Pasi Tyrvainen[b]

[a]*University of Jyvaskyla, Department of Computer Science and Information Systems, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*
[b]*University of Jyvaskyla, Agora Center, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*

Abstract

Hybrid cloud storage combines cost-effective but inflexible private storage along with flexible but premium-priced public cloud storage. As a form of concurrent sourcing, it offers flexibility and cost benefits to organizations by allowing them to operate at a cost-optimal scale and scope under demand volume uncertainty. However, the extant literature offers limited analytical insight into the effect that the non-stationarity (i.e., variability) and non-determinism (i.e., uncertainty) of the demand volume – in other words, the demand variation – have on the cost-efficient mix of internal and external sourcing. In this paper, we focus on the reassessment interval – that is, the interval at which the organization re-assesses its storage needs and acquires additional resources –, as well as on the impacts it has on the optimal mix of sourcing. We introduce an analytical cost model that captures the compound effect of the reassessment interval and volume variation on the cost-efficiency of hybrid cloud storage. The model is analytically investigated and empirically evaluated in simulation studies reflecting real-life scenarios. The results confirm that shortening the reassessment interval allows volume variability to be reduced, yielding a reduction of the overall costs. The overall costs are further reduced if, by shortening the interval, the demand uncertainty is also reduced.

*Corresponding author
   Email addresses: gabriella.laatikainen@jyu.fi (Gabriella Laatikainen), mazhelis@jyu.fi (Oleksiy Mazhelis), pasi.tyrvainen@jyu.fi (Pasi Tyrvainen)

1. Introduction

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

Hybrid cloud infrastructure, where there is a combination of concurrently used private and public cloud infrastructure resources (Armbrust et al., 2010), 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 real 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 infrastructure providers (Gauger, 2013) may make their services overly expensive for larger enterprises.

Cloud computing, as a form of on-demand computing, represents a special form of outsourcing (Willcocks and Lacity, 2012; Venters and Whitley, 2012; Chen and Wu, 2013; Son et al., 2014), whereby the property or decision rights regarding the IT infrastructure are transferred to an external organization.

2

Furthermore, the hybrid cloud infrastructure can be seen as an instantiation of concurrent sourcing, which is a simultaneous use of market contracting and vertical integration, that is, a situation in which the same good or service is produced as well as bought(Parmigiani, 2007; Parmigiani and Mitchell, 2009; Mols, 2010; Heide et al., 2013).

Outsourcing and make-or-buy decisions have been the subject of extensive study in the field of information systems (IS) (Gregory et al., 2013; Lacity et al., 2011; Kotlarsky et al., 2014), as well as in strategic management and operations management research (Freytag and Kirk, 2003; van de Water and van Peet, 2006; Weigelt and Sarkar, 2012). Along with the need to focus on core capabilities, cost-savings represent the most frequently cited reasons behind the decisions to outsource in general (Lacity et al., 2009), and the decision to use public cloud infrastructure in particular (Venters and Whitley, 2012).

Meanwhile, hybrid cloud infrastructure as a concurrent sourcing phenomenon has attracted little attention from the IS research community. Whereas concurrent sourcing has been widely studied outside of IS in the automotive (Gulati et al., 2005), metal forming (Parmigiani, 2007) and 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 concurrent sourcing. Therefore, research inquiry into cloud-enabled flexibility, and in particular into the hybrid cloud and its impact on future cloud services, has been indentified as one of the directions for further research (Venters and Whitley, 2012).

Concurrent sourcing has been studied from the viewpoint of theories such as transaction cost economics, agency theory, resource-based theory, neoclassical economics, life cycle theory, resource and capability view, theories of multi-profit center firms, marketing channels, options theory, and knowledge-based theory (Mols, 2010; Mols et al., 2012). A widely cited justification for the use of concurrent sourcing derives from transactional cost theories and neoclassical economics. Specifically, it is claimed that this form of governance reduces production costs when firms face so-called volume uncertainty (Adelman, 1949; Parmigiani, 2003; Mols, 2010), that is, difficulty in accurately predicting demand volumes (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

3

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 versus 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).

It has been observed that volume uncertainty reflects the difficulty in accurately predicting demand volumes and can be defined as the degree of (in)precision with which volume is predicted (Parmigiani, 2003, 2007). However, besides this prediction inaccuracy, the natural variation in the volume of the demand referred to as *variability* (e.g., seasonal fluctuations) can be the reason for 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 fully deterministic and perfectly predictable. Together, the volume uncertainty and volume variability comprise the *variation* in the volume of the demand (van Belle, 2008). To the best knowledge of the authors, the variability aspect of variation has not been explicitly considered in the concurrent sourcing literature.

A key question in the recent literature on cloud computing as well as on concurrent sourcing is the optimal mix of internal and external sourcing. Indeed, the cost-optimal mix of 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 lesser extent on proactive resource provisioning (Weinman, 2012; Mazhelis and Tyrväinen, 2012). Likewise, in the literature on concurrent sourcing, multiple factors have been found to affect the optimal mix, including resource co-specialization, supplier selection as well as the cost and benefits of producing in-house resources and buying from external parties (Sako et al., 2013; Puranam et al., 2013), with volume uncertainty found among the factors that warrant additional studies (Sako et al., 2013).

One of the parameters shaping the optimal mix of sourcing is the *reassessment interval* (also referred to as acquisition cycle time), which can be defined as the time period between successive time points when the organization reassesses 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. The demand reassessment interval affects the degree of volume variation, because 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 that the demand reassessment interval, through its effect on volume variation, impacts on how much to produce internally versus how much to procure from external sources, and determines the volume of cost savings that are attainable by hybrid cloud storage.

The objective of this paper is to increase our understanding of the economic effect that the reassessment interval and volume variation have on the cost of hybrid cloud infrastructure. In particular, the paper studies hybrid cloud storage as a subset of hybrid cloud infrastructure, the popularity of which has increased dramatically in recent years and which is predicted to increase even further (TwinStrata, 2013; 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 effect 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 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 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 the case of storage, fluctuations are rare; instead, the demand for storage is usually a monotonically non-decreasing function (Laatikainen et al., 2014). Nevertheless, within a single 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 yields cost

5

benefits in the context of cloud storage resources as well.

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

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

The following two hypotheses are formulated:

(i) Shortening the reassessment interval leads to smaller unutilized excess capacity, thereby reducing the inefficiencies of scale.

(ii) Shortening the reassessment interval may reduce the demand estimation error, thereby further reducing inefficiencies of scale by minimizing the departure from cost-optimal sourcing.

The paper subscribes to the design science research (DSR) paradigm (Hevner et al., 2004; Peffers et al., 2007; Gregor and Hevner, 2013) wherein an innovative artifact – in the form of a conceptual-analytical model – is constructed and evaluated, in order to increase our understanding of concurrent sourcing in the IS domain and to address the practical issue of determining the cost-optimal mix of internal and external storage resources. This model is systematically evaluated in the paper, being (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 two-fold. First, by studying the concurrent sourcing phenomenon in the context of cloud infrastructure, the specific aspects of the latter, such as non-decreasing demand for storage and relatively high utility premiums, can be taken into account, thus deepening our understanding of concurrent sourcing in the IS domain. Second, this paper contributes to the theoretical foundations of IS management by allowing managers to compare the savings attainable through shortening the resource acquisition cycle with the cost of acquiring the organizational capability needed for shortening the cycle. In other words, the paper helps in applying agile principles to IS management by offering a tool for comparing the cost savings gained through the flexibility of concurrent sourcing with the costs of cloud transformation for the purpose of enabling such sourcing.

6

Furthermore, the results of the study can be added to the general body of knowledge of concurrent sourcing, thereby helping to resolve the contradictions present in the contemporary concurrent sourcing literature. Studying the role of volume uncertainty analytically, as a central concept of concurrent sourcing literature, provides previously unknown insights into the role of volume uncertainty in an organization's choice of different sourcing forms as well as the optimal allocation between buy and make, and helps in achieving one of the goals of concurrent sourcing – maximizing the volume of cost saving, which is particularly crucial during economic downturns.

The paper is organized as follows. In the next section, related research in the field of hybrid cloud storage is surveyed. Section 3 introduces the analytical model of hybrid cloud storage costs and investigates the regularities inherited in the model. This model is further empirically evaluated in Section 4 using simulation studies reflecting real-life scenarios. In Section 5, the theoretical and practical implications of the constructed model are discussed, and the directions for further research are outlined. Finally, conclusions are presented in Section 6.

2. Related Work

This section provides an overview of related work in the field of cloud storage in general (subsection 2.1), and hybrid cloud storage in particular (subsection 2.2). In subsection 2.3, hybrid cloud storage is discussed as an instantiation of the concurrent sourcing phenomenon. The section ends with a summary of the applicable cost factors, and it indicates the gaps in research where further study is needed (subsection 2.4).

2.1. Cloud Storage

The popularity of using cloud storage services has increased dramatically in recent years, and it is predicted to increase further, partially due to the fact that the growth of storage capacity demand outpaces the capacity growth attainable in-house. For example, 84% of survey respondents attending the Cloud Computing Expo in New York in June 2013 indicated that they were planning to use or were already using cloud storage (TwinStrata, 2013).

Generally, data stored in the cloud may be characterized by large capacity, varying data access patterns, soft performance requirements, online access from different geographical locations, and low management overhead (Agarwala et al., 2011). When the application is data intensive, the most

7

important requirements are data durability, availability, access performance, usability, and support for security and privacy (Palankar et al., 2008). However, besides rich functionality, low cost is among the most essential requirements (Agarwala et al., 2011).

Organizations may follow different approaches to maximize the cost-efficiency of cloud storage. First of all, they may store only the provenance for the data and regenerate the rest when needed (Borthakur, 2007; Adams et al., 2009). In this case, in addition to deciding upon trading storage for computing requirements based on a cost-benefit analysis, the organizations also have to consider whether the stored data can be feasibly computed, whether the exact result may be replaced with an acceptable approximation, and whether the legal and security requirements are met (Adams et al., 2009). Different strategies have been proposed in (Yuan et al., 2010a,b, 2011) to find the best trade-off between storage and computational costs by storing the appropriate intermediate data in cloud storage. The need for incorporating the provenance services into cloud storage offerings is also emphasized by Muniswamy-Reddy and Seltzer (2010), who analyze several alternative implementations that collect provenance data and use the cloud as a back end.

Another approach to decreasing the costs of cloud data storage is to use data transformation, such as compression, deduplication, and transcoding (Agarwala et al., 2011). Compression algorithms offer different trade-offs between the decrease in storage volume and the increase in resource consumption (memory, CPU cycles) as well as the additional delays in restore operations (Mao et al., 2014; Agarwala et al., 2011). Data deduplication is a type of data compression that is often used in cloud backup and archiving systems as well as in primary storage for virtual machine servers to reduce the amount of storage space consumed (Mao et al., 2014). Depending on the redundancy requirements, by storing only one single instance of each unique data chunk, storage needs may be reduced by as much as 80% for VM servers, and backup and archiving applications also benefit significantly from data deduplication (Mao et al., 2014; Clements et al., 2009). Current storage systems use erasure codes, such as Reed-Solomon codes, for storing infrequently assessed data (the so-called cold data) to ensure its redundancy (André et al., 2014; Jiekak et al., 2013), whereas frequently accessed data (the so-called hot data) is replicated in multiple disk massives to provide high availability from non-reliable devices (André et al., 2014; Jiekak et al., 2013).

To achieve cost reduction of cloud data centers, data replication mechanisms may be used in conjunction with different energy saving strategies Long et al. (2014). Different algorithms exist to spin down the data nodes from the high energy consumption mode into a lower standby mode when they are inactive (Zhu et al., 2005; Long et al., 2014; Xie, 2008; Weddle et al., 2007). Storage cost may also be reduced by employing efficient audit services to ensure data integrity (Zhu et al., 2012).

The research work outlined above primarily addresses issues pertaining to public cloud storage, whereas relatively little attention has been devoted to hybrid cloud storage and its related costs. In the next subsection, the literature on hybrid cloud infrastructure, with a special focus on hybrid cloud storage, is surveyed.

2.2. Hybrid cloud storage

Hybrid cloud computing infrastructure is a composition of private and public clouds where in-house and public resources are concurrently used in order to enable data and application portability (Armbrust et al., 2010; Mell and Grance, 2011). In a recent survey, over 70% of enterprises have chosen to adopt hybrid cloud infrastructure (RightScale, 2014). In light of these results showing its importance, it is not surprising the hybrid cloud has been researched widely.

Research has been devoted to the two core technical enablers of hybrid cloud computing – virtualization and multi-tenancy – that allow cloud resources to be pooled together to serve multiple clients (Smith and Nair, 2005; Bittencourt and Madeira, 2011; Kabbedijk et al., 2015; Cai et al., 2010). In addition, much research work has focused on partitioning, that is, deciding which applications, application components, or computing jobs must be kept local, and which ones must be migrated to the public cloud (Fan et al., 2011; Huang and Shen, 2015; Tak et al., 2013). In a hybrid environment, different policies might determine if the application or the workload is assigned to the private or public cloud, such as application-specific functionalities and requirements (Khajeh-Hosseini et al., 2011; Fan et al., 2011; Wang et al., 2012; Hajjat et al., 2010; Juan-Verdejo and Baars, 2013), economic, security and privacy implications (Silva et al., 2013), data sensitivity, and high performance requirements (Zhang et al., 2014). Alternatively, the placement may be decided on the fly depending on the current load of the system (Mazhelis, 2012). In this case, organizations can use cloud bursting, the process by

9

which excess load is offloaded to public cloud infrastructure if the workload exceeds a specific threshold.

One of the key research questions in cloud bursting is to determine the workload or portion of a workload that should be offloaded (Fadel and Fayoumi, 2013). It has been shown that the cost-optimal time of using the public cloud is the inverse of the premium charged by the public cloud provider (Weinman, 2012; Mazhelis and Tyrväinen, 2011, 2012). Furthermore, Mazhelis and Tyrväinen (2011, 2012) have shown that data communication overheads as well as the volume discounts set by the public cloud provider affect the cost-optimal time of using the public cloud. Knowing the future workload and the cost-optimal time of using the public cloud, the cost-optimal portion of the workload to serve in-house can be estimated.

Instead of looking at processes that determine a suitable distribution of workloads in advance, a related stream of research concentrates on automatic resource provisioning. This is when application requests are mapped to the distributed physical resources on the fly and the execution of the appications are scheduled on the fly (e.g., Calheiros et al., 2011; Andrikopoulos et al., 2013; Sun et al., 2015; Cerviño et al., 2013; Trummer et al., 2010).

The studies on hybrid cloud storage, in particular, may be exemplified with the model by Lima et al. (2014), which explicitly takes into account latency, uptime, free size, and cost when determining the most appropriate placement of data in a hybrid cloud. In Abu-Libdeh et al. (2010), a system is provided for stripping the data across multiple providers in order to reduce the cost of vendor lock-in and facilitate switching providers, as well as to better tolerate provider outages or failures. In their work, Villari et al. present a solution for distributing the data across many cloud storage providers while enforcing long-term availability, data confidentiality, and data redundancy (Villari et al., 2013, 2014; Celesti et al., 2016). Furthermore, the use of data filtering to reduce intercloud data transmission overheads has been explored by Han et al. (2013), with the aim of improving the cost-efficiency of applications where the performance of a hybrid cloud may not be sufficient due to low bandwidth and high latency of data communications between private and public clouds. However, to a large extent, the available research either infuses the storage-related issues as part of comprehensive hybrid cloud considerations (e.g., Malawski et al., 2013) or focuses on security as a primary design objective (e.g., Dobre et al., 2014).

The relative scarcity of the research on hybrid cloud storage can be explained by the fact that cloud storage services initially relied solely on public

storage infrastructure. Meanwhile, as the results of a recent survey by Enter-
prise Strategy Group indicate, the majority of IT professionals participating
in the survey are extremely (69%) or somewhat (28%) interested in hybrid
cloud storage (McClure, 2014). This suggests that hybrid cloud storage will
likely become the subject of increasing interest to the research community in
the near future.

2.3. Hybrid cloud storage as a form of concurrent sourcing

Hybrid cloud infrastructure in general and hybrid cloud storage in par-
ticular can easily be seen as a form of concurrent sourcing. Concurrent
sourcing refers to the simultaneous use of market contracting and vertical
integration, that is, it means producing as well as buying the same good or
service. This phenomenon has gained increasing attention in recent literature
on organization and strategic management, where it has been labeled as, for
example, tapered integration (Porter, 1980), partial integration (Jacobides
and Billinger, 2006), concurrent sourcing (Parmigiani, 2007; Parmigiani and
Mitchell, 2009; Heide et al., 2013; Mols, 2010), plural sourcing (Jacobides
and Billinger, 2006), and plural governance (Heide, 2003). It has also been
studied from the viewpoint of various theories, including transaction cost eco-
nomics, agency theory, resource-based theory, neoclassical economics, life cy-
cle theory, resource and capability view, theories of multi-profit center firms,
marketing channels, options theory and knowledge-based theory (Mols et al.,
2012; Mols, 2010). For consistency, the term concurrent sourcing will be used
throughout this paper.

The theoretical explanations of concurrent sourcing and the available em-
pirical results shed some light on why some organizations use hybrid cloud
solutions. They suggest, for instance, that volume uncertainty coupled with
the high cost of excess capacity, along with factors such as performance ambi-
guity, technological volatility, and information asymmetry, likely contribute
to the popularity of concurrently using private and public clouds. Further-
more, the open issues that require further studies in concurrent sourcing
literature – including the role of volume uncertainty and variability – are
relevant in the context of the hybrid cloud as well and warrant further inves-
tigation.

When seen through the lens of hybrid cloud research, the concurrent
sourcing literature has two limitations. First, the findings in the extant
literature on concurrent sourcing may not be sufficient for explaining the
concurrent use of private and public cloud infrastructure. This is due to

11

the fact that the concurrent sourcing research may fail to take into account some key aspects of the hybrid cloud, such as the possibility to fulfill QoS or legal requirements (e.g., regarding data availability and confidentiality) with private cloud resources while benefiting from inexpensive external resources to execute the components with less stringent requirements (Juan-Verdejo and Baars, 2013).

Second, it has to be noted that, in the concurrent sourcing literature, the optimal mix of in-house and external resources has been explained in terms of market conditions and firm strategy. However, the hybrid cloud literature also contains a growing body of knowledge that focuses on other factors affecting optimal dynamic allocation of resources and which takes into account various requirements and constraints. These factors can be useful in explaining concurrent sourcing in contexts other than the hybrid cloud, meaning that the hybrid cloud literature can contribute to a deeper general understanding of the concurrent sourcing phenomenon.

Overall, the concurrent sourcing literature offers limited analytical and empirical insight into the role of volume uncertainty in concurrent sourcing. Therefore, there is a need for analytical inquiry focusing on this subject as well as for gathering empirical evidence to validate the findings. Hybrid cloud storage represents a contemporary context for such an inquiry, the specifics of which (e.g., specific forms of demand curves, quality-of-service requirements, or data sensitivity concerns) will provide new insight into concurrent sourcing.

2.4. Cost factors in hybrid cloud infrastructure

From the perspective of neoclassical economics, the use of hybrid cloud infrastructure as a form of concurrent sourcing allows an organization to hedge against the risks of underutilized excess capacity, and therefore minimize the infrastructure-related costs. Different cost-based metrics are available to the decision-makers who are deciding upon the possible adoption of a cloud solution. A summary of these metrics is provided in Table 1 below, along with references to the research where these metrics have been developed or applied.

The research on the cost-efficiency of (hybrid) cloud infrastructure suggests that the cost-efficiency of the hybrid cloud is determined by a variety of cost factors that have a compound non-linear effect on the overall costs. These costs factors, considered in the extant literature on cloud computing,

12

Table 1: Metrics used in cost-based analysis of cloud deployment alternatives

Net Present Value of money	Tak et al. (2013); Brumec and Vrček (2013); Walker et al. (2010); Mastroeni and Naldi (2011); Mazhelis (2012)
Total Cost of Ownership	Klems et al. (2009); Koomey et al. (2007); Martens et al. (2012); Han (2011); Walterbusch et al. (2013); Mazhelis et al. (2012b); Bibi et al. (2012); Brumec and Vrček (2013); Mazhelis et al. (2012b)
Value-at-Risk	Mastroeni and Naldi (2011)
Return On Investment (ROI)	Beaty et al. (2011); Misra and Mondal (2011)

can be grouped into several categories (Table A.3 in the appendix lists the references to the publications where these cost factors have been studied):

1. *Cost factors related to in-house resources:* the cost of acquiring, provisioning and maintaining an in-house data center during its entire lifecycle, including hardware and server costs, software license fees, electricity and labor costs, business premises, as well as the adjacent cost factors related to the strategy and the practices of the organization (e.g., acquisition and forecasting intervals, the degree of data center utilization).

2. *Cost factors related to public resources:* the cost of computations, storage, data communications, load balancing, and other adjacent cost factors, such as the pricing models of cloud providers, charging period, volume discount, market and technological trends.

3. *Cost factors reflecting the interaction between the private and public cloud and/or the use of a private and public cloud concurrently:* partitioning and allocation costs, data communication intensity between the private and public cloud, the threshold for workload re-allocation between in-house and external subsystems.

4. *Cost factors related to the organizational, environmental, or system context:* the usage pattern of the system or the service, the demand

13

growth rate, variability and uncertainty, the type of applications and their requirements, system architecture, size of the organization.

5. *Other cost factors:* Costs related to the decision-making on possible cloud adoption and the selection of a cloud provider; costs of deployment, integration, migration and configuration; support and maintenance cost; training costs; potential losses due to cloud adoption.

Even though a plethora of cost models with various granularities of cost factors have been studied, 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. The work closest to ours is that by Laatikainen et al. (2014), where the role of the acquisition interval in the cost-efficiency of the private versus public cloud has been analyzed. Although the acquisition interval has been studied in the context of selecting between private and public storage, to the best knowledge of the authors no publicly available research focuses on the role of the reassessment interval and volume variation in the cost efficiency of hybrid cloud storage. Therefore, below, a hybrid storage cost model is introduced where the compound effect of the reassessment interval and volume variation is taken into account.

3. Modeling the Cost 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 an 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.

3.1. Assumptions

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

1. First, we assume that the storage demand is a non-decreasing function in time. Indeed, as opposed 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

14

(Laatikainen et al., 2014). As a result, the digital universe as a whole grows 40% a year, according to a recent study by IDC.[1]

2. Second, we 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. 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 assume that each unit of data is atomic in the sense that (i) it bears the same level of criticality,[2] and (ii) it is stored on either the 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 or replicated in another cloud infrastructure. As a result, the interaction between the private and public subsystems can be assumed to be negligible.

4. Finally, the organization is assumed to allocate the storage on a 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, we can consider the cost components comprising the hybrid storage cost model. As overviewed in the previous section, different cost components are relevant for the private and the public subsystems. For the private subsystem, the relevant cost constituents include the costs of hardware and software acquisition, integration, configuration and upgrading, 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 of its growth pattern and its predictability, the time interval between storage acquisitions, and the pricing of the needed equipment, software, and personnel, among

---

[1]http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm
[2]Here, depending on the nature of the organization's business, the criticality may encompass confidentiality, reliability, availability, and other considerations.

other costs.

On the other hand, for public cloud storage, the cost components include usage-dependent costs such as, in the 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) charges its customers based on the maximum storage capacity used in 12-hour intervals.[3]

In the case of hybrid cloud storage, the process of acquiring, provisioning and paying for the necessary storage resources differs between the private and public subsystems. On the one hand, the *private* subsystem's resources cannot be added to instantly when the need arises, because 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. It must periodically estimate its future demand and acquire and deploy the additional resources for the in-house storage infrastructure in advance. The interval between the subsequent resource acquisitions based on the future demand estimates is referred to here as the *reassessment interval*. The cost of the private storage subsystem is incurred at the beginning of each reassessment interval and so it depends on the maximum storage capacity and 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. 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 opposed to the private subsystem, the organization pays for the public subsystem's resources only when they are used and only for the volume of storage in the public subsystem that is actually used.

As stated above, in this paper we assume that the organization 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, the utility premium of the cloud provider, and the length of the reassessment interval, which are the main subjects of the anal-

---

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

16

ysis in this section. The total cost of hybrid cloud storage is also affected by many additional factors, such as the cost of adopting a hybrid infrastructure, 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 (see Section 2), 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 storage subsystems, and hence are left outside of the scope of the analysis.

The remainder of the section is organized as follows. First, in subsection 3.2, we introduce a hybrid cloud storage cost model that captures the role of the reassessment interval in the hybrid cloud costs under the assumption that the growth of the storage demand is well-known or that there is no estimation error. After that, in subsection 3.3, we relax this assumption and include the volume uncertainty in the model in order to assess the impact of the reassessment interval on the hybrid cloud storage costs when the demand is imperfectly estimated.

3.2. General Hybrid Storage Cost Model

Let us define the demand function $s(t) \mapsto \mathbb{R}$ 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, meaning the non-stationary nature of the demand.

Let us consider the total cost of a hybrid storage solution during the reassessment interval of length w as shown in Figure 1. 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 owning the private storage subsystem C_o during the reassessment interval of length w. As described above, at 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
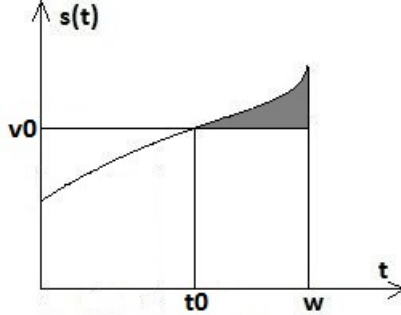
17

Figure 1: Hybrid cloud costs without refinement of reassessment interval

owning in-house capacity C_o can be estimated as

$$C_o = v_0\, p_o(v_0)\, w, \tag{1}$$

where v_0 is the maximum private storage capacity to be used within the next reassessment interval. In case the actual demand $s(t)$ exceeds v_0, the difference $s(t) - v_0$ 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 than 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 deprovisioning the resources (Weinman, 2011a). As a result, some organizations found it significantly less expensive to host their 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

18

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, the 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 and from the public subsystem and (ii) the cost of transferring the cumulated public storage capacity to the private subsystem once the private capacity is increased.

In order to make the analysis tractable, we 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, Equation (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

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

Let us now consider the cost-impact of shortening the reassessment interval. Specifically, let us consider the case when the reassessment interval is refined, that is, when it is divided into two adjacent reassessment intervals $P1$ and $P2$ of the lengths z and $w - z$, respectively (see Figure 2). Let us mark the maximum private storage over the period $P1$ with v_1, and over the period $P2$ with v_2.
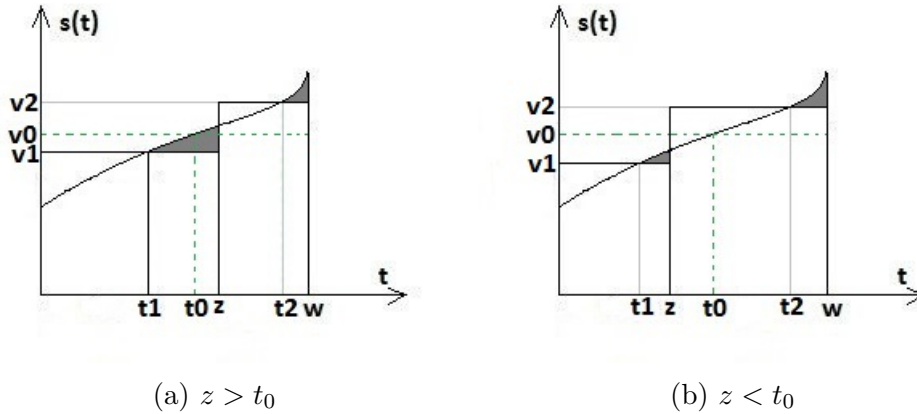

(a) $z > t_0$ (b) $z < t_0$

Figure 2: Hybrid cloud costs with a refinement of the reassessment interval. 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 $P1$ can be expressed as

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

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 $P2$ are

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

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 $P1$ and $P2$:

$$C_{H2} = C_{HP1} + C_{HP2}. \tag{6}$$

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

The cost difference thus can be written as

$$\begin{aligned}
f = C_{H1} - C_{H2} = \\
= p_o\, v_0\, w - u\, p_o\, v_0\,(w - t_0) - p_o\, v_1\, z + u\, p_0\, v_1\,(z - t_1) - p_o\, v_2\,(w - z) + u\, p_0\, v_2\,(w - t_2) \\
+ u\, p_o\, \left(\int_{t_0}^{w} s(t)dt - \int_{t_1}^{z} s(t)dt - \int_{t_2}^{w} s(t)dt \right).
\end{aligned} \tag{7}$$

It can be seen from the equation above 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 take into account the above stated 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 the 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

$$t_0 = \frac{u-1}{u} w, \tag{8}$$

$$t_1 = \frac{u-1}{u} z, \tag{9}$$

$$t_2 = z + \frac{u-1}{u}(w-z). \tag{10}$$

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

$$f = u \, p_o \left(\int_{\frac{u-1}{u} w}^{w} s(t)dt - \int_{\frac{u-1}{u} z}^{z} s(t)dt - \int_{z+\frac{u-1}{u}(w-z)}^{w} s(t)dt \right). \tag{11}$$

Proposition 3.1. *Assuming the allocation of cost-optimal amount of storage capacity to private cloud and no reassessment costs, re-evaluating the storage needs more often is always beneficial costwise, that is, $f > 0$.*

Proof. The proof of the proposition is provided in Appendix B.1. $\qquad\square$

Let us further consider the cost-efficient division of the reassessment interval, by analyzing which division point z allows the cost difference, as reflected in f, to be maximized.

Lemma 3.1. *In the interval $(0, w)$, $f(z)$ has only one extremum point where $\frac{\partial f}{\partial z} = 0$, and this extremum corresponds to the maximum of $f(z)$ in the region $(0, w)$.*

Proof. The proof of the lemma is provided in Appendix B.2. $\qquad\square$

Note that the value of z_{max} depends on the form of the demand function. Let us illustrate the cost-efficient division of the reassessment interval in case of a linear demand function that can be defined as

$$s(t) = a \, t + b, \tag{12}$$

where $a > 0$ (assuming that the demand is monotonically increasing) and $b > 0$ (assuming that the demand is positive at the beginning of the storage period) are real numbers. In this case, the cost difference function f is simplified to

$$f = p_o \, \frac{a \, z \, (u-1) \, (w-z)}{u}. \tag{13}$$

21

Lemma 3.2. *In the case of a linearly growing demand function, the greatest cost-savings can be achieved when $z = w/2$.*

Proof. The proof of the lemma is provided in Appendix B.3. □

However, in the case of an exponentially growing demand function, the largest cost difference is attainable by splitting the reassessment interval into two, with the latter subinterval being shorter.

Lemma 3.3. *In the case of an exponentially growing demand function, the greatest cost-savings can be achieved when $z > w/2$.*

Proof. The proof of the lemma is provided in Appendix B.4. □

Let us investigate the impact of the utility premium on the cost difference function (13). For this, consider the first derivative of the cost difference function with respect to the utility premium, which can be written in the following form:

$$
\begin{aligned}
\frac{\partial f}{\partial u} &= p_o \left(F(t_2) - F(t_0) + F(t_1) - F(z) \right) + u\, p_o \left(\frac{\partial}{\partial u} F(t_2) - \frac{\partial}{\partial u} F(t_0) + \frac{\partial}{\partial u} F(t_0) \right) \\
&= p_o \left(F(t_2) - F(t_0) + F(t_1) - F(z) \right) + u\, p_o \left(\frac{\partial F(t_2)}{\partial t_2} \frac{\partial t_2}{\partial u} - \frac{\partial F(t_0)}{\partial t_0} \frac{\partial t_0}{\partial u} + \frac{\partial F(t_1)}{\partial t_1} \frac{\partial t_1}{\partial u} \right) \\
&= p_o \left(F(t_2) - F(t_0) + F(t_1) - F(z) \right) + u\, p_o \left(s(t_2) \frac{w - z}{u^2} - s(t_0) \frac{w}{u^2} + s(t_1) \frac{z}{u^2} \right).
\end{aligned}
$$
(14)

Referring to Equation (B.1) it can be seen that the first term in the expression above is positive:

$$
p_o \left(F(t_2) - F(t_0) + F(t_1) - F(z) \right) > 0.
$$

Furthermore, the sign of the term

$$
u\, p_o \left(s(t_2) \frac{w - z}{u^2} - s(t_0) \frac{w}{u^2} + s(t_1) \frac{z}{u^2} \right)
$$
(15)

depends on the demand curve $s(t)$ or, more precisely, on how great $s(t_0) \frac{w}{u^2}$ is in relation to $s(t_2) \frac{w-z}{u^2} + s(t_1) \frac{z}{u^2}$.
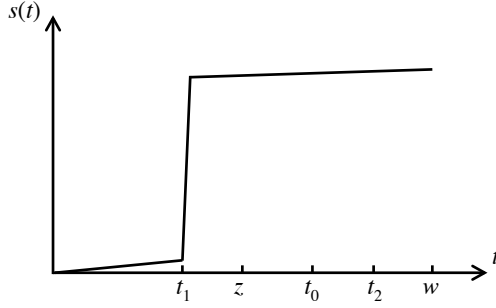
22

Figure 3: Special case of the demand curve that grows very slowly before t_1 and after t_0 but very rapidly in the range $(t_1, t_1 + |\delta|), \delta \to 0$

Let us now investigate the special case when the demand curve grows very slowly before t_1 and after t_0, but very rapidly in the range $(t_1, t_1 + |\delta|), \delta \to 0$ (as shown in Figure 3). In this case, $s(t_2) \approx s(t_0)$ and $s(t_1) \to 0$.

Therefore,

$$p_o \left(F(t_2) - F(t_0) + F(t_1) - F(z) \right) = p_o \left(\int_{t_0}^{t_2} s(t)dt - \int_{t_1}^{z} s(t)dt \right) \to 0,$$

whereas

$$u\,p_o \left(s(t_2) \frac{w - z}{u^2} - s(t_0) \frac{w}{u^2} + s(t_1) \frac{z}{u^2} \right) \approx -\frac{p_o}{u} s_o z.$$

As a result, it follows that

$$\frac{\partial f}{\partial u} \approx -\frac{p_o}{u} s_o z < 0.$$

However, for more conventional demand growth functions, the value of the derivative is likely to be positive, and therefore the observation holds that the greater the utility premium, the greater the cost savings due to the refinement of the reassessment interval. Let us illustrate it with an example of linear demand growth.

Lemma 3.4. *In the case of linearly growing demand, the greater the utility premium, the greater the cost savings due to the refinement of the reassessment interval.*

Proof. The proof of the lemma is provided in Appendix B.5. □

23

Let us introduce the cost of reassessment c_r, that is, a cost associated with each re-assessment event. The reassessment cost includes, among others, the cost of demand estimation for the next reassessment interval and the procurement and deployment of the additional storage resources. Being dependent on the internal practices of the organization, the reassessment cost is difficult to estimate. However, for the sake of simplicity, we will assume that the reassessment cost is the same for each reassessment interval within the same organization and is independent of the volume of the storage, either available or to be purchased.

Let us calculate the total hybrid cost C_{H1} (Equation 4) when the reassessment cost is taken into account:

$$C_{H1} = c_r + p_o\, v_0\, w + u\, p_o\, \Big(\int_{t_0}^{w} s(t)dt - v_0\, (w - t_0)\Big). \tag{16}$$

If the reassessment interval is refined, that is, split into two intervals, the reassessment cost is incurred twice:

$$C_{H2} = 2\, c_r + C_{HP1} + C_{HP2} \tag{17}$$

In this case, the cost difference function f is the following:

$$f = C_{H1} - C_{H2} = -c_r + u\, p_o\, \left(\int_{\frac{u-1}{u}\, w}^{w} s(t)dt - \int_{\frac{u-1}{u}\, z}^{z} s(t)dt - \int_{z+\frac{u-1}{u}\, (w-z)}^{w} s(t)dt\right). \tag{18}$$

Let us refer to proposition 3.1 and define the *cost benefits due to refinement of the reassessment interval* as the following term from (Equation 11):
$u\, p_o\, \left(\int_{\frac{u-1}{u}\, w}^{w} s(t)dt - \int_{\frac{u-1}{u}\, z}^{z} s(t)dt - \int_{z+\frac{u-1}{u}\, (w-z)}^{w} s(t)dt\right).$

Lemma 3.5. *Assuming the allocation of the cost-optimal amount of storage capacity to the private cloud and non-zero reassessment cost, re-evaluating the storage needs more often is beneficial costwise if the cost savings due to reassessment are higher than the cost of reassessment.*

Proof. The proof of the lemma is provided in Appendix B.6. □

Let us now investigate how refining the reassessment interval *recursively* affects the overall costs. Indeed, by reducing the length of the reassessment interval recursively, cost savings can be achieved in line with Proposition 3.1. However, in line with Lemma 3.5, the reassessment cost associated with each reassessment event reduces the cost savings. Thus, a stopping criteria for the

24

recursive reassessment can be defined where the cost of reassessments exceeds the benefits achievable through reassessments. Formally, the stopping criteria can be defined as follows.

Let C_{H1}, C_{H2}, ... , C_{Hn} be the total hybrid costs when the reassessment interval is divided into $1, 2, ..., n$ intervals. Let $f_2 = C_{H1} - C_{H2}$, $f_3 = C_{H2} - C_{H3}, ..., f_n = C_{H(n-1)} - C_{Hn}$ be the corresponding benefits due to reassessment. It can be shown that if $n \to \infty$ then $f_n \to 0$, that is, at some point, $f_n < c_r$, making further refinements economically inferior. Thus, given the constant reassessment cost c_r, the refinements are economically justifiable as long as $f_i > c_r$.

3.3. Cost Model Taking into Account Demand Forecasting Errors

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

$$\hat{s}(t) = s(t)\left(1 + \varepsilon\right), \tag{19}$$

where the estimation error $\varepsilon = \varepsilon(t_y - t_x)$ is a function of the length of the forecasting horizon $t_y - t_x$, 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 volume variation: volume uncertainty, meaning the inaccuracy with which the demand volumes are predicted.

Several assumptions need to be made about the estimation error function. The estimation error may be additive or multiplicative depending on the application. In this study, 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. In addition, the estimation error increases with the forecasting horizon even if the estimable storage demand exhibits little change. Accordingly, the error is assumed to be a non-constant and increasing function of time, although no specific functional form is assumed. Lastly, the demand function is assumed to have negligible or no bias.

The estimation error contaminates the estimates of required storage ca-

25

pacity v:

$$\hat{v}_0 = s(t_0)(1 + \varepsilon_0), \text{where } \varepsilon_0 = \varepsilon(t_0);$$
$$\hat{v}_1 = s(t_1)(1 + \varepsilon_1, \text{where } \varepsilon_1 = \varepsilon(t_1);$$
$$\hat{v}_2 = s(t_2)(1 + \varepsilon_2), \text{where } \varepsilon_2 = \varepsilon(t_2 - z).$$

Importantly, the error in the estimates of v_i also spreads into the "effective" value of t_i, denoted as $\hat{t}_i$, where $i = \{0, 1, 2\}$. For instance, if v_0 is overestimated ($\varepsilon_0 > 0$), it effectively means that the public cloud storage will start to be used later than originally envisioned, $\hat{t}_0 > t_0$. Having denoted the error function impacting t_i as ξ, we can express the "effective" values of t_i as follows:

$$\hat{t}_0 = t_0(1 + \xi_0), \text{where } \xi_0 = \xi(t_0),$$
$$\hat{t}_1 = t_1(1 + \xi_1), \text{where } \xi_1 = \xi(t_1),$$
$$\hat{t}_2 = z + (t_2 - z)(1 + \xi_2) = z + \frac{u - 1}{u}(w - z)(1 + \xi_2), \text{where } \xi_2 = \xi(t_2 - z).$$

Several notes shall be made. First, ξ is also assumed to act as a multiplicative error, in line with ε. Second, the errors ξ and ε are covarying, so if $\varepsilon > 0$, then $\xi > 0$, and vice versa, as demonstrated in Figure 4 below. Finally, it is important to observe that $\hat{s}(t_i) = s(\hat{t}_i)$.
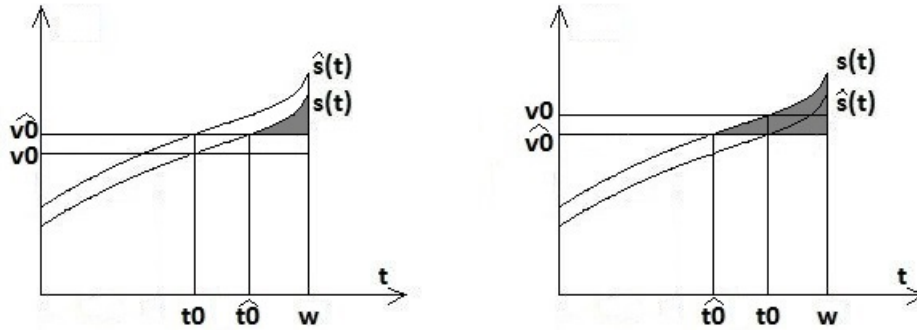
(a) Demand overestimation ($\varepsilon_i > 0, i = 1..3$, and $\xi_j > 0, j = 1..3$)

(b) Demand underestimation ($\varepsilon_i < 0, i = 1..3$, and $\xi_j < 0, j = 1..3$)

Figure 4: Dependency between the errors ξ and ε

Taking into account the estimation errors introduced above, for the cost-optimal storage allocation as specified in Equations (8-10), the cost difference function f can be rewritten as

$$f = p_o s(t_0)(1 + \varepsilon_0)w + u p_o \left[\int_{\hat{t}_0}^{w} s(t)dt - s(t_0)(1 + \varepsilon_0)(w - \hat{t}_0) \right]$$

$$- p_o s(t_1)(1 + \varepsilon_1)z - u p_o \left[\int_{\hat{t}_1}^{z} s(t)dt - s(t_1)(1 + \varepsilon_1)(z - \hat{t}_1) \right]$$

$$- p_o s(t_2)(1 + \varepsilon_2)(w - z) - u p_o \left[\int_{\hat{t}_2}^{w} s(t)dt - s(t_2)(1 + \varepsilon_2)(w - \hat{t}_2) \right]. \quad (20)$$

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

$$f = p_o s(t_0)(1 + \varepsilon_0)w\xi_0(u - 1) + u p_o \left[\int_{t_0}^{w} s(t)dt - \int_{t_0}^{\hat{t}_0} s(t)dt \right]$$

$$- p_o s(t_1)(1 + \varepsilon_1)z\xi_1(u - 1) - u p_o \left[\int_{t_1}^{z} s(t)dt - \int_{t_1}^{\hat{t}_1} s(t)dt \right]$$

$$- p_o s(t_2)(1 + \varepsilon_2)(w - z)(u - 1)\xi_2 - u p_o \left[\int_{t_2}^{w} s(t)dt - \int_{t_2}^{\hat{t}_2} s(t)dt \right], \quad (21)$$

or, after regrouping,

$$f = f^* + \left(p_o s(t_0)(1 + \varepsilon_0)w\xi_0(u - 1) - u p_o \int_{t_0}^{\hat{t}_0} s(t)dt \right)$$

$$- \left(p_o s(t_1)(1 + \varepsilon_1)z\xi_1(u - 1) - u p_o \int_{t_1}^{\hat{t}_1} s(t)dt \right)$$

$$- \left(p_o s(t_2)(1 + \varepsilon_2)(w - z)\xi_2(u - 1) - u p_o \int_{t_2}^{\hat{t}_2} s(t)dt \right), \quad (22)$$

where $f^* = u p_o \left[\int_{t_0}^{w} s(t)dt - \int_{t_1}^{z} s(t)dt - \int_{t_2}^{w} s(t)dt \right]$ is the value of the cost difference function in case the estimation of demand is free of estimation error, as specified in Equation (11).

27

Based on the equation above, the difference $f - f^*$ can be expressed as

$$\Delta = f - f^* = \left[p_o s(t_0)(1 + \varepsilon_0) w \xi_0 (u - 1) - u p_o \int_{t_0}^{\hat{t}_0} s(t) dt \right]$$
$$- \left[p_o s(t_1)(1 + \varepsilon_1) z \xi_1 (u - 1) - u p_o \int_{t_1}^{\hat{t}_1} s(t) dt \right]$$
$$- \left[p_o s(t_2)(1 + \varepsilon_2)(w - z) \xi_2 (u - 1) - u p_o \int_{t_2}^{\hat{t}_2} s(t) dt \right]$$
$$= \alpha(\varepsilon_0, \xi_0) - \alpha(\varepsilon_1, \xi_1) - \alpha(\varepsilon_2, \xi_2), \tag{23}$$

where $\alpha(\varepsilon_0, \xi_0)$, $\alpha(\varepsilon_1, \xi_1)$, and $\alpha(\varepsilon_2, \xi_2)$ represent the three 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., $|\varepsilon_0| > |\varepsilon_2|$, $|\varepsilon_0| > |\varepsilon_1|$, $|\xi_0| > |\xi_2|$, $|\xi_0| > |\xi_1|$), then it is likely that $\alpha(\varepsilon_0, \xi_0) \gg \alpha(\varepsilon_1, \xi_1)$ and $\alpha(\varepsilon_0, \xi_0) \gg \alpha(\varepsilon_2, \xi_2)$, 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 becomes smaller. Let us illustrate it with the special case of the linearly growing demand function.

In the case of linear growth specified by the demand function in Equa-

tion (12), the cost difference f is in the form

$$
\begin{aligned}
f = f^* &+ p_o(at_0 + b)(1 + \varepsilon_0)w(u - 1)\xi_0 \\
&- p_o(at_1 + b)(1 + \varepsilon_1)z(u - 1)\xi_1 \\
&- p_o(at_2 + b)(1 + \varepsilon_2)(w - z)(u - 1)\xi_2 \\
&+ up_o \times [-\frac{a}{2}t_0^2\xi_0(\xi_0 + 2) - b\xi_0 t_0 + \frac{a}{2}t_1^2\xi_1(\xi_1 + 2) + b\xi_1 t_1 \\
&+ \frac{a}{2}\left(z^2 + (t_2 - z)^2(1 + \xi_2)^2 + 2z(t_2 - z)(1 + \xi_2) - t_2^2\right) + b(z + (t_2 - z)(1 + \xi_2) - t_2)],
\end{aligned}
\tag{24}
$$

which can be rewritten as

$$
\begin{aligned}
f = f^* &+ p_o(u - 1) \\
&\times \left[\xi_0 w \left[(a\frac{u - 1}{u}w + b)(1 + \varepsilon_0) - w(\xi_0 + 2)\frac{a(u - 1)}{2}\frac{}{u} - b\right]\right. \\
&\quad -\xi_1 z \left[(a\frac{u - 1}{u}z + b)(1 + \varepsilon_1) - z(\xi_1 + 2)\frac{a(u - 1)}{2}\frac{}{u} - b\right] \\
&\quad \left. -\xi_2(w - z)\left[(a(z + \frac{u - 1}{u}(w - z)) + b)(1 + \varepsilon_2) - \frac{a}{2}\left(\frac{u - 1}{u}(w - z)(\xi_2 + 2) + 2z\right) - b\right]\right]
\end{aligned}
\tag{25}
$$

or, equally, as

$$
\begin{aligned}
f = f^* + p_o(u - 1) \times &\left[\xi_0 w \left[a\frac{u - 1}{2u}w(2\varepsilon_0 - \xi_0) + b\varepsilon_0\right]\right. \\
&\quad -\xi_1 z \left[a\frac{u - 1}{2u}z(2\varepsilon_1 - \xi_1) + b\varepsilon_1\right] \\
&\quad \left. -\xi_2(w - z)\left[a\frac{u - 1}{2u}(w - z)(2\varepsilon_2 - \xi_2) + (az + b)\varepsilon_2\right]\right].
\end{aligned}
\tag{26}
$$

Observe that, for the linear growth function, it holds that

$$
\xi_i = \varepsilon_i(1 + \frac{b}{at}) = k_i\varepsilon_i,
\tag{27}
$$

where $k_i = 1 + \frac{b}{at}$ is a function of t, $i = 1..3$. Further, let the errors ε_1 and ε_2 be expressed as functions of ε_0, i.e. $\varepsilon_1 = c_1\varepsilon_0$ and $\varepsilon_2 = c_2\varepsilon_0$, where c_1 and

29

c_2 are real-valued coefficients. Then, the equation above can be rewritten as

$$f = f^* + p_o(u - 1)\varepsilon_0^2 \times \left[k_0 w \left[a \frac{u - 1}{2u} w(2 - k_0) + b \right] \right.$$
$$- k_1 c_1 z \left[a \frac{u - 1}{2u} z c_1 (2 - k_1) + c_1 b \right]$$
$$\left. - k_2 c_2 (w - z) \left[a \frac{u - 1}{2u} (w - z) c_2 (2 - k_2) + c_2 (az + b) \right] \right].$$
$$(28)$$

Thus, if we assume that c_1 and c_2 are independent of ε_0, then $f - f^*$ is a quadratic function of ε_0 with an extremum at $(0, 0)$. Among other factors, the sign of $f - f^*$ depends on a and c_2: if $c_2 \leq 1$ (i.e., if the errors are non-declining functions of the length of the forecasting horizon), then it is likely that $f - f^* > 0$. However, when $c_2 > 1$ and $a \gg 0$, then the last term in the equation above likely dominates, resulting in $f - f^* < 0$.

4. Numerical Experiment: Simulating an Archival System

In the previous section, an artifact in the form of a hybrid cloud storage cost model has been introduced and analytically investigated, with the aim of revealing its inherent properties. This section expands our effort at evaluating this model by means of numerical simulations that take into account the context of a real-world organization.

4.1. Design of Numerical Experiments

Numerical simulation is a kind of 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 is designed to maximize the amount of relevant information from a limited number of simulation runs (Hunter et al., 1978). In order to resemble reality, the 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 and 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 a real-world organization that maintains and develops a large-scale storage solution whose storage demand and its growth can be considered to be representative. Second, a long-time trace of storage massives in use at NCAR allows the historic developments of storage needs to be observed. Finally, as opposite to commercial organizations that keep their infrastructure details 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 on the organization's website.[4] The NCAR's archival systems have their roots in the mid-1960s. Over the years, a number of developments were made to accommodate the growing needs for 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.[5] At present, the archival system represents a combination of the new tape libraries of High Performance Storage System (HPSS),[6] and the legacy tape libraries maintained by a subcontractor.[7] 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.[8]

For the purposes of this study, we use the storage metrics with monthly granularity that were kindly provided by NCAR. In Figure 5, the growth profile of the NCAR's archival system during the period 1 September 1986–1 April 2014 is shown. As evidenced by the figure, the demand for data storage exhibited exponential growth during these years, rising from 2TB in 1986 to over 30PB in 2014.

---

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

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

[6]`https://www2.cisl.ucar.edu/docs/hpss`

[7]`http://www.nar.ucar.edu/2009/CISL/1comp/1.3.6.amstar.php`
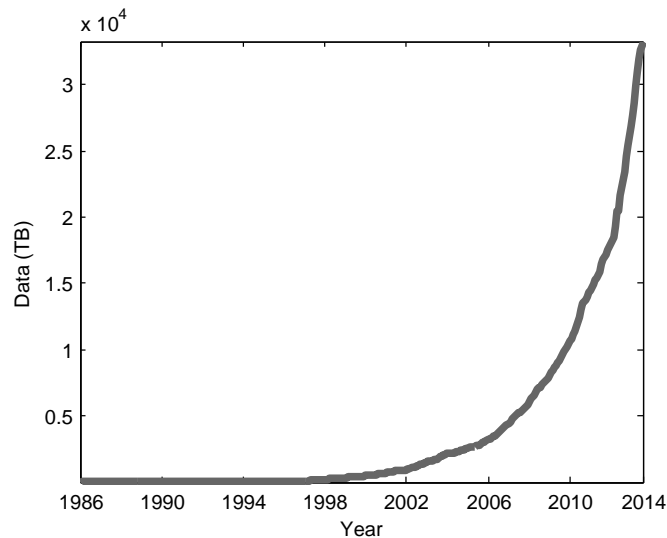
[8]`https://www2.cisl.ucar.edu/resources/glade`

Figure 5: Growth of the NCAR archiving storage during 1986-2014

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

Assuming that Reduced Redundancy Storage (RRS) is used as a public storage equivalent,[9] 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. Note that, for simplicity, the request pricing has not been taken into account, because the contribution of the request-based charges to the overall cost is rather modest in the case of the archival solutions.

Instead of RRS, Amazon Glacier could have been used as an inexpensive public tape storage equivalent that only costs \$0.01 to store 1GB for a month. However, significant costs are incurred for transferring the data out of the service because, in addition to the data transfer fee above, the transfer may incur a significant retrieval fee that depends on the desired retrieval rate.

---

[9]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.

Deleting files stored for less than three months incurs fees as well. All this makes the use of Glacier economically inefficient in cases where the data is stored for short periods of time, as is 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, in 2011 Backblaze was reportedly spending $94 563 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, the software solutions for managing the storage (e.g., IBM Tivoli Storage Manager) and related services are also likely to be needed, thus increasing the cost of the storage solution further. However, we 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 a six-month period cost: (i) $1 539 if the data is stored in-house using Backblaze's type of storage, and (ii) $22 878 if the data is stored in Amazon Reduced Redundancy Storage and transferred at the end of the storage period; this results in a utility premium value of $22 878/$1 539=14.9.

Storing the same volume of data on tape will cost (i) $2 550 if the in-house tape storage is used as described in (Reine and Kahn, 2013), and (ii) $16 786 if Amazon Glacier is used instead,[10] thus resulting in a utility premium of $16 786/$2 550=6.6.

A couple of issues should be noted at this point. The costs of the private storage solutions may be underestimated. First, additional labor costs

---

[10]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.

33

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.[11] We will therefore explore a set of different values of utility premium in the range of $u \in [4; 20]$.

It should also be mentioned that the reassessment cost (i.e., the cost associated with estimating the future demand for the next reassessment interval and acquiring and deploying additional in-house storage resources) greatly depends on the internal practices of the organization; its value therefore is difficult to estimate. Due to this, and also because this cost is likely to be insignificant in the case under consideration when compared with the overall storage costs for the case organization, in the numerical experiments below we assume the reassessment cost to be zero.

Finally, it should be pointed out 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 assume that approximately the same decline rate applies to both the private and the public portions. Likewise, the time value of money is not taken into account in the cost estimates, because, within a single reassessment interval, the present value of money changes insignificantly and has a limited effect on the overall costs.

In the simulation below, we compare the costs incurred by an organization facing the growing demand for storage as experienced by NCAR (i) in the case that the organization is re-estimating its storage needs and acquiring additional in-house resources on a yearly basis, and (ii) in the case when the organization is conducting the reassessment twice a year at different time points measured in months, $z \in \{2, 3, \ldots, 11\}$.

4.2. Results

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

---

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

4.2.1. Known demand for storage

We 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 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, so the median value of $u = 2.88$ is therefore used in the cost calculations unless explicitly specified otherwise.

In Figure 6, the total yearly costs of hybrid cloud storage are shown for the reassessment intervals 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 the figure shows, the total hybrid storage costs are lower if the organization reassesses its storage needs more often, that is, once every six months instead of once a year. This is in line with Proposition 3.1, which claims that the more frequent re-evaluation of storage needs to be cost-beneficial.

According to Lemma 3.1, the cost saving function $f(z)$ has a single extremum in the interval $(0, w)$, which corresponds to the maximum of $f(z)$ in the region $(0, w)$. This is visible in Figure 7, where the cost savings are portrayed as a function of the refinement point z. Furthermore, in line with Lemma 3.2, for linearly growing demand function, the greatest cost saving is expected when $z_{max} = w/2$, whereas for exponentially growing demand function, in line with Lemma 3.3, the value of z_{max} shifts to the right, $z_{max} > w/2$.

As can be seen from the graphs in Figure 7, the greatest cost savings are achieved when the company reassesses its storage needs at the middle of the original reassessment period, in other words, if the refinement is done at $z = 6$ months given the original reassessment interval of $w = 12$ months. Thus, albeit the storage demand does exhibit an exponential growth and hence $z_{max} > w/2$ is expected, in practice, $z_{max} = w/2$ predicted for the linear growth is observed. This can be explained by the fact that, within
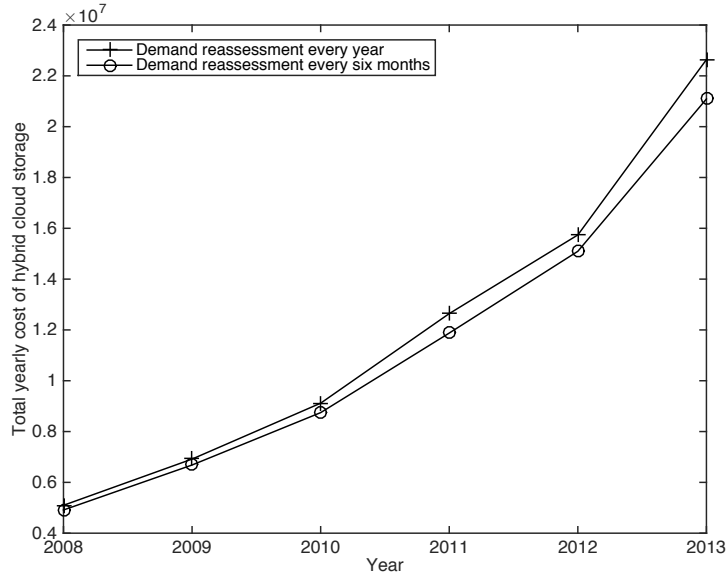
Figure 6: The total yearly cost of a hybrid cloud storage for reassessment intervals of six and twelve months

a single year, the growth rate is relatively low (circa 0.03 in average), and hence the growth can be relatively well approximated with a linear function.

Finally, let us turn to the effect of the utility premium u on the cost savings. As stated in Lemma 3.4, the cost savings due to the refinement of the reassessment interval increase with the value of u. In order to investigate this dependency, Figure 8 plots the dependency between the cost savings function f and the utility premium.

The graphs in the figure reflect the overall regularity expressed in Lemma 3.4: in the cost-optimal allocation of storage to the private and public cloud, the more expensive the public cloud is compared to the private cloud, the greater are the cost savings that can be achieved with reassessing the storage needs more often. Meanwhile, as can be seen from the figure, for some subregions of utility premium values the cost savings may remain constant or even decrease temporarily (e.g., consider the case of $z = 9$ and $8 \leq u \leq 17$). Such temporal declines are caused by the rounding of the time of public cloud storage use. Specifically, small changes in u induce a small change in the cost-optimal time of using the public cloud resources (i.e., $w - t_i$). However, since $w - t_i$ is calculated with monthly granularity and hence needs to be
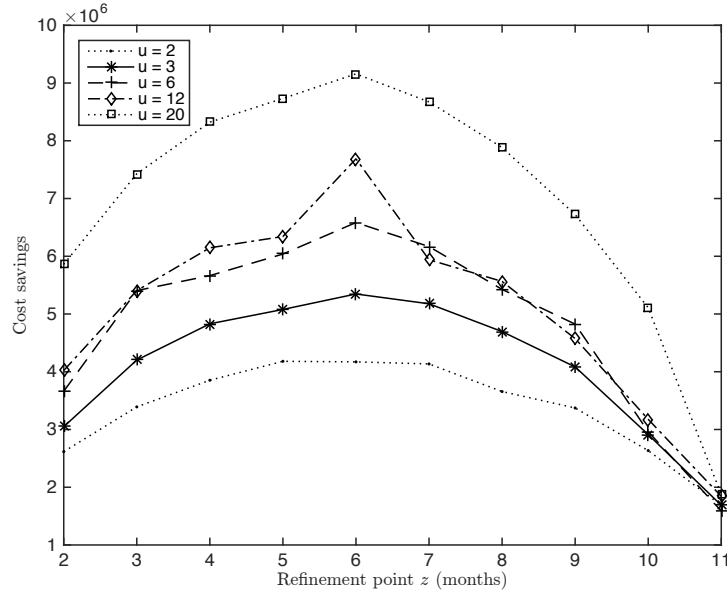
Figure 7: Cost savings due to the refinement of the reassessment interval for different times of refinement

rounded to the nearest month, this results in suboptimal values of $w - t_i$, and consequently may result in a cost saving that is (temporally) decreasing with u. Note that the effect of this rounding is also visible in Figure 7, where, for $u = 2$, $z_{max} = 5$, while $z_{max} \geq 6$ is expected.

4.2.2. 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 in both the 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
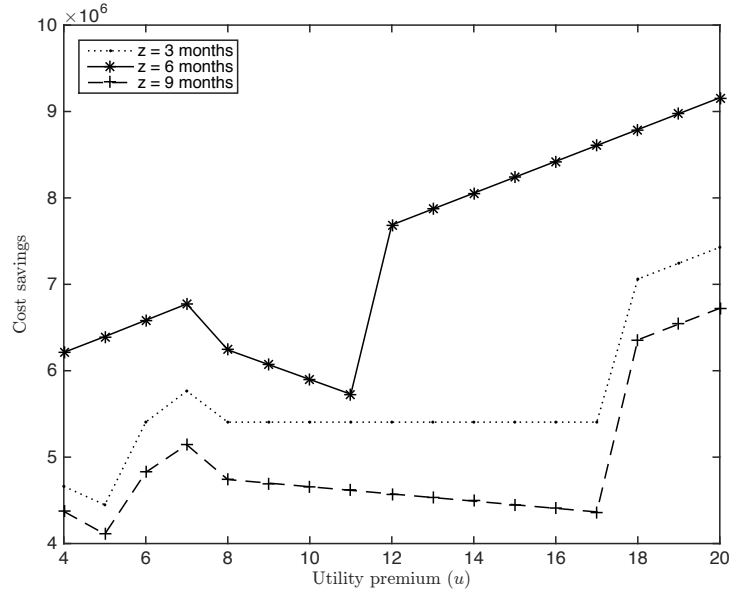
37

Figure 8: Cost savings due to the refinement of the reassessment interval for different levels of utility premium u

estimate the parameters of an exponential growth function. The forecasted and the original data for yearly reassessment interval are shown in Figure 9. As can be seen from the figure, the forecasted demand curve largely follows the original demand, although there are periods when the demand is under- or overestimated. In the results presented next, we have excluded the data for the first year because there was no historical data to base the forecast on. We have also excluded the data for the final year (2014) because the available data for that year were incomplete.

As was analytically shown in the previous section, the effect of the refinement of the reassessment interval on the cost savings depends on whether the forecasting inaccuracy decreases with the refinement. Indeed, as Figure 10 and Table 2 show, the change in the cost savings Δ 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 (in 2001 and 2002), whereas for the years when the error increases ($\xi_0 - \xi_2 > 0$), the cost difference de-
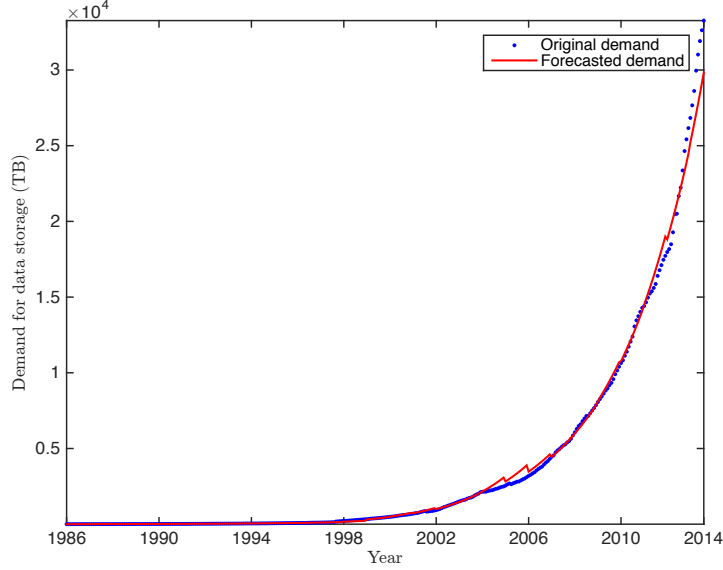
38

Figure 9: Forecasted and real storage growth

clines (in 2011 and 2012). Thus, in line with the analytical considerations
in subsection 3.3, it can be observed that, while the refinement of the re-
assessment interval does cut the cost of 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. Otherwise, it decreases.

There is also an interesting phenomenon worth mentioning. In the case
that (i) the demand function is growing stepwise, (ii) the stepwise growth
co-occurs with t_i, and (iii) a small overestimation of demand is present, then
the presence of the overestimation has no impact on the estimation of t_i.
Specifically, given two consecutive steps of the demand function s_x and s_{x+1}
occurring at t_i, and given a small estimation error ε_i s.t. $s_x(1 + \varepsilon_i) < s_{x+1}$,
the cost impact of the overestimation can be expressed as:

$$p_o s_x(1+\varepsilon_i)w - u p_o s_x(1+\varepsilon_i)(w - \hat{t}_i) = p_o s_x(1+\varepsilon_i)(w - u(w - \frac{u-1}{u}w)) = 0.$$

As a result, whenever such conditions occur, the corresponding term in Equa-
tion (23) nullifies (i.e., $\alpha(\varepsilon_i, \xi_i) = 0$), which may have a decisive effect on the
sign of Δ. This is the case, for instance, for 2004 and 2008, when $\alpha(\varepsilon_0, \xi_0) = 0$,
resulting in $\Delta < 0$, as well as for 1992 and 2010, when $\alpha(\varepsilon_2, \xi_2) = 0$, resulting

39

Table 2: Estimation errors and the change in cost savings

| 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	-16.17
1994	0.104	0.074	0.067	0.50	0.50	0.50	-144.51
1995	0.091	0.069	0.049	0.38	0.50	0.50	183.90
1996	0.045	0.036	0.032	0.25	0.25	0.25	123.04
1997	-0.013	0.036	-0.027	-0.13	0.25	-0.50	90.20
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
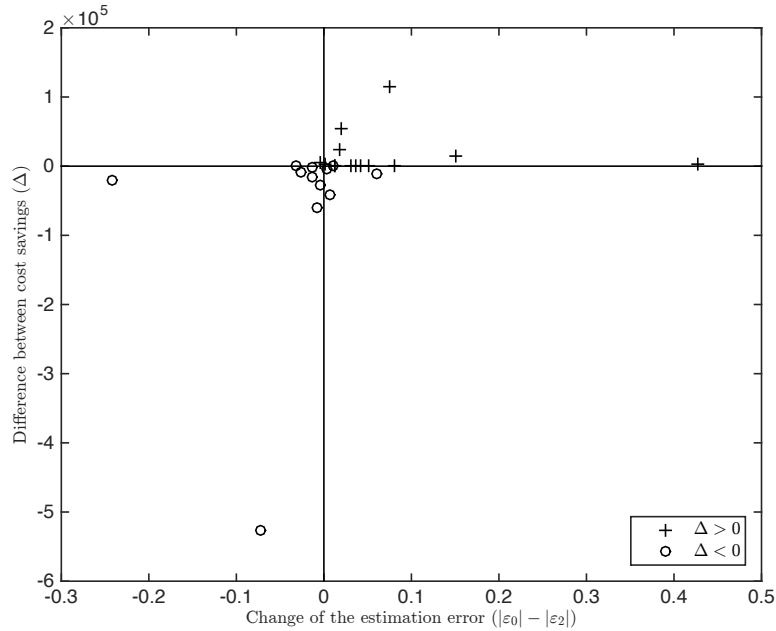
Figure 10: Change in the cost savings (Δ) vs. the change in the demand estimation error ($|\varepsilon_0| - |\varepsilon_2|$)

in $\Delta > 0$.

5. Discussion

Neoclassical economics can provide an apt characterization for the case of concurrent sourcing in the cloud storage domain. In it, the technology is commonly available with little asset specificity, partner behavior is predictable and so requires little protection against supplier opportunism, performance is predictable, and one of the central problems is how to operate at optimal scale and scope under volume uncertainty. Harrigan (1986), for instance, suggests that a mix of internal production and external suppliers is a low-risk strategy when demand is erratic and uncertain. Likewise, Carlton (1979) argues that it is advantageous to integrate in order to save costs for the high probability component of demand and use external suppliers for the low-probability demand.

In fact, the cloud storage domain represents a case which Parmigiani hypothesized but did not confirm in her study (Parmigiani, 2007). That is, a greater scope of economies for both the firm and its suppliers to produce the

41

good was hypothesized to encourage the firm to concurrently source part of ₉₆₅ the demand. On the one hand, making everything internally would require prior investment based on estimated demand and create extra costs for unused capacity, while unpredictable volumes raise costs and hurt performance (cf. Wagner and Bode, 2006). On the other hand, the premium charged by the external suppliers for the surplus capacity is high: The standard ₉₇₀ neoclassical economic explanation for concurrent sourcing involves hedging against demand uncertainty. In this case, a firm can keep its internal plant at full production by using suppliers to handle fluctuating additional volumes, thereby running more efficiently due to having this flexibility in capacity (Adelman, 1949; Carlton, 1979; Porter, 1980). This position assumes a ro-₉₇₅ bust spot market with a large number of qualified external suppliers vying for the firm's business, although these suppliers will have higher base costs (Adelman, 1949). The actual prices they charge the firm may be even higher, due to the risk they are bearing by having unused capacity during slack times and by not knowing when the 'low probability' demand will occur (Carl-₉₈₀ ton, 1979). Indeed, suppliers may charge premiums for lower volumes and short lead times since they know they are merely 'overflow outlets' for the firm (Harrigan, 1986; Hill, 1994). Firms may be willing to pay these premiums rather than invest in additional, and potentially underutilized, capacity. (Parmigiani, 2007) In fact, unlike many other previous studies summarized in ₉₈₅ (Mols, 2010) and (Parmigiani, 2007), the neoclassical theory alone seems to provide the most applicable explanation for the concurrent sourcing problem in hybrid cloud storage.

Prior literature has shown that concurrent sourcing in the context of cloud resources, referred to as the hybrid cloud, can reduce costs by combining ₉₉₀ in-house processing and storage capacity with premium-priced public cloud capacity. This paper has shown that the cost of hybrid cloud storage in concurrent sourcing may be reduced even more by refining the reassessment interval. Furthermore, the magnitude of the cost cut depends on two distinct dimensions of volume variation: on the non-stationary (demand variability) ₉₉₅ and on the non-deterministic nature of the demand volume (demand volume uncertainty). The findings are the following:

- For demand variability: The maximum cost cut is achieved when the refinement is at the middle of the sourcing period, for linear growth, and the cost cut grows with the utility premium.

₁₀₀₀ - For demand volume uncertainty: If the refinement allows the forecast-

ing inaccuracy to be reduced, then the economic benefit of the refinement increases. Otherwise, it decreases.

Note that the results on demand variability are specific to hybrid cloud storage and they do not necessarily hold true, for example, in the context of hybrid cloud computing resources. This is due to the monotonically growing nature of the storage demand, which makes the variability decline if the reassessment interval gets shorter. Meanwhile, the results regarding demand volume uncertainty are generally applicable to hybrid cloud computing resources as well, and are likely to be applicable to the other domains where concurrent sourcing is used.

Based on the analytical findings (see Lemma 3.5), re-estimating future storage needs more often and acquiring additional in-house resources accordingly reduces the total hybrid cost, assuming no additional cost associated with the reassessment. However, this additional cost – the reassessment cost – reduces the cost benefits. It follows that refining the reassessment interval can be recursively continued and cost-benefits achieved until the cost associated with demand estimation and additional in-house resource provisioning exceed these benefits.

Based on the numerical example with NCAR's archival system data, moving from twelve-month cycles in storage capacity acquisition of hybrid cloud storage resources to six-month cycles would decrease the annual costs by about $1M, representing about 5% of the annual $23M in costs for 2013. Meanwhile, when compared to the costs of acquiring new resources only, the cost benefit of shortening the acquisition interval grows to about 15% of the acquisition expenses. In this example, the demand estimation errors were relatively small, in most years resulting in an impact on the savings volume that was between plus or minus $ 50K (i.e., less than 3%). However, due to the nature of a business, the estimation error can easily be much higher than in this example.

We should note that the demand for storage in the case of NCAR exhibits annual growth of 40%, a figure in line with the general growth trend reported for digital storage (IDC, 2014). Therefore, the results are expected to be applicable to other organizations engaging in the adoption of hybrid cloud storage. However, care should be taken when extrapolating the findings above to other domains of concurrent sourcing, where the specifics of these domains, including the growth trend and its predictability as well as the utility premium values, should be taken into account.

The case considered in this paper further connects concurrent sourcing to the literature on strategic flexibility (Sanchez, 1995), especially that regard-
ing resource flexibility (Sanchez, 2004) and real options (Brydon, 2006) as well as relates to transaction cost economics (Williamson, 1985). As long as we assume no extra cost from repeating the capacity estimation and acqui-sition cycle more often, the faster cycle provides an option to minimize the 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 $1M for halving the cycle for the case orga-nization. However, even though in an organization of this size, the benefits exceed the costs, for a small organization the savings could easily be smaller than the resource acquisition costs, thus recommending the use of an annual capacity acquisition cycle.

6. Conclusions

The core benefit of cloud computing can be attributed to the business flexibility achievable by converting capital IT expenditures to on-demand operational expenditures. As compared with traditional in-house IT infras-tructure, this provides both a low-cost option to scale a business and the ability to make frequent and rapid changes in business models. This flexibil-ity has been essential to the emergent trend to utilize cloud-based capacity, transform the IT function to cloud-compatible systems, and utilize agile net-worked business models. In order to deliver such flexibility, public cloud providers have to be capable of guaranteeing scalability for services whose demand grows by factors of 100 or even 1000 in a few months. In response, these providers may request utility premiums as high as 2 to 20 times the in-house costs.

The hybrid cloud solutions combining fixed in-house cloud resources and flexible public cloud resources provide cost-optimal solutions when the vol-ume 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 the public cloud for the peak demand only. Significantly, 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.

44

This paper contributes to the cloud storage economics literature by analyzing hybrid cloud storage, which combines in-house storage and public cloud storage. A general hybrid storage cost model was constructed to analyze the cost benefits of using hybrid cloud storage in the presence of volume variation. Specifically, these cost benefits were analyzed in the presence of demand variability, as manifested in seasonal changes, and in the presence of volume uncertainty, as manifested in the volume estimation errors. This analysis shows that shortening the reassessment interval and more frequent acquisition of private cloud storage capacity allows the volume variability to be reduced, yielding a reduction of the overall costs. We further showed that splitting an in-house resource acquisition interval into equal subintervals maximizes the cost saving, assuming that the demand needs grow linearly. The analytical part was validated with a numerical example from a conventional storage organization. Namely, the data from NCAR's archival system showed that cutting the resource acquisition cycle from twelve months to six months would provide 15% acquisition cost savings, with the assumption that there would be no costs for speeding up the storage acquisition cycle.

Importantly, this paper sheds some light on the economic viability of organizational transformation towards cloud adoption through re-engineering or replacing an organization's information systems to become hybrid cloud-enabled. Indeed, the economical viability of cloud transformation can be questioned for a number of reasons: the renewal of information systems incurs costs, using cloud-enabled software includes a performance penalty (5–15%), and the use of public cloud offerings is associated with high utility premiums. Such a high premium is tolerable for small firms with no in-house IT capability, but for larger enterprises with in-house IT capabilities the use of the resources available in-house may prove less expensive in the longer term.

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., the combination of in-house private cloud and a limited volume of public cloud). In the numerical example from a conventional storage organization, cutting the annual resource acquisition cycle to six months would provide 15% savings on cloud storage costs, assuming no costs for speeding up the storage acquisition cycle and executing it twice a year. This cost saving may represent an incentive for an enterprise to acquire the capability needed for concurrent sourcing in cloud environments, that is, for adopting a 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 could motivate an enterprise to carry out a cloud transformation.

These results encourage enterprises to enable the use of a hybrid cloud approach through conformance to standards and the development of in-house competences, and thus promote the development of capability for performing cost-efficient cloud storage acquisition. From the perspective of Schlagwein et al. (2014), the results support the trend toward ensuring technological cloud readiness in enterprises. This readiness can support immediate or long-term migration of applications to the cloud as well as enable flexible, short-term contracts with cloud providers while allowing enterprises to retain internal capabilities in their IT functions and become competent IT brokers able to integrate external and internal IT resources. In other words, they become able to support hybrid cloud storage solutions.

While this paper showed that the cost benefit of flexibility in concurrent sourcing motivates an enterprise like the case organization to adopt a hybrid cloud approach which requires in-house cloud capability and cloud transformation of incumbents, the case could be somewhat different in 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, leading to a reduction in the cost of the transformation. For small firms the overhead of establishing in-house information systems, maintaining in-house servers, and so on may also be a capital-intensive cost factor that can or even must be avoided. They may therefore have the tendency to use only the public cloud until the storage demand has increased substantially and the cost benefit of a hybrid cloud approach overrules the capital expanditure and inflexibility of in-house storage. To further understanding of this practice, we suggest that future research address the economic view on the flexibility and premium costs of the public cloud-only approach in comparision to the cost-optimal hybrid cloud solution in quickly growing small enterprises with high volume variation.

Appendix A. Cost Factors in Hybrid Cloud Infrastructure

46

Table A.3: Cost factors reported in the literature on hybrid cloud infrastructure

Cost factor	References
Cost factors related to the use of in-house resources	
hardware (servers, network devices, lifetime of hardware, replacement costs of different hardware components)	Khajeh-Hosseini et al. (2012, 2011); Risch and Altmann (2008); Kashef and Altmann (2012); Tak et al. (2011); Kondo et al. (2009); Opitz et al. (2008); Greenberg et al. (2008); Koomey et al. (2007); Han (2011); Brumec and Vrček (2013); Bibi et al. (2012); Mastroeni and Naldi (2011); Walker et al. (2010); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Beaty et al. (2011); Laatikainen et al. (2014); Mazhelis et al. (2012a); Gonzalez et al. (2013)
electricity (cooling, lightning, electronic devices)	Khajeh-Hosseini et al. (2012, 2011); Risch and Altmann (2008); Kashef and Altmann (2012); Tak et al. (2011); Kondo et al. (2009); Armbrust et al. (2010); Opitz et al. (2008); Greenberg et al. (2008); Koomey et al. (2007); Han (2011); Bibi et al. (2012); Mastroeni and Naldi (2011); Walker et al. (2010); Mazhelis et al. (2012b); Beaty et al. (2011); Laatikainen et al. (2014); Mazhelis et al. (2012a); Gonzalez et al. (2013)
software costs (basic server software, middleware, application software)	Andrikopoulos et al. (2013); Kashef and Altmann (2012); Risch and Altmann (2008); Kondo et al. (2009); Opitz et al. (2008); Tak et al. (2013); Bibi et al. (2012); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Beaty et al. (2011); Laatikainen et al. (2014); Mazhelis et al. (2012a)
	Continued on next page

47

Cost factor	References
labor costs (software and hardware maintenance, other support)	Kashef and Altmann (2012); Tak et al. (2011); Kondo et al. (2009); Opitz et al. (2008); Koomey et al. (2007); Han (2011); Bibi et al. (2012); Mastroeni and Naldi (2011); Walker et al. (2010); Beaty et al. (2011); Laatikainen et al. (2014); Mazhelis et al. (2012a); Gonzalez et al. (2013)
business premises (air conditioner, rack, cabling, facility, internet connectivity, land, interest during construction, architectural and engineering fees, security, taxes, lifetime of data center, tier level of functionality, usable uninterruptible power supply (UPS) output, the electrically active floor area, insurance, etc.)	Kashef and Altmann (2012); Tak et al. (2011); Kondo et al. (2009); Armbrust et al. (2010); Hajjat et al. (2010); Truong and Dustdar (2010); Greenberg et al. (2008); Koomey et al. (2007); Opitz et al. (2008); Han (2011); Bibi et al. (2012); Turner and Seader (2006); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012a)
in-house resource utilization degree	Opitz et al. (2008); Mazhelis (2012); Tak et al. (2013); Greenberg et al. (2008); Koomey et al. (2007)
acquisition interval (time period between two acquisitions of additional in-house resources)	Mazhelis (2012); Laatikainen et al. (2014); Mazhelis et al. (2012a)
forecasting, provisioning, de-provisioning and demand monitoring interval	Weinman (2011c,a)
Cost factors related to the use of public resources	

Cost factor	References
computation costs (running virtual machine hours/CPU hours/server usage)	Khajeh-Hosseini et al. (2012, 2011); Risch and Altmann (2008); Andrikopoulos et al. (2013); Kashef and Altmann (2012); Tak et al. (2011); Kondo et al. (2009); Hajjat et al. (2010); Truong and Dustdar (2010); Hernández et al. (2013); Martens et al. (2012); Mian et al. (2012); Han (2011); Martens and Teuteberg (2012); Brumec and Vrček (2013); Kratzke (2012); Bibi et al. (2012); Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Weinman (2011b,a); Truong and Dustdar (2010); Agarwala et al. (2011); Gonzalez et al. (2013); Adams et al. (2009)
storage costs	Khajeh-Hosseini et al. (2012, 2011); Risch and Altmann (2008); Andrikopoulos et al. (2013); Kashef and Altmann (2012); Kondo et al. (2009); Armbrust et al. (2010); Hajjat et al. (2010); Truong and Dustdar (2010); Hernández et al. (2013); Martens et al. (2012); Mian et al. (2012); Martens and Teuteberg (2012); Kratzke (2012); Mastroeni and Naldi (2011); Walker et al. (2010); Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Truong and Dustdar (2010); Laatikainen et al. (2014); Mazhelis et al. (2012a); Agarwala et al. (2011); Ruiz-Alvarez and Humphrey (2011, 2012); Gonzalez et al. (2013); Adams et al. (2009)

Cost factor	References
cost of input/output requests	Khajeh-Hosseini et al. (2012, 2011); Risch and Altmann (2008); Andrikopoulos et al. (2013); Kashef and Altmann (2012); Kondo et al. (2009); Martens et al. (2012); Martens and Teuteberg (2012); Kratzke (2012); Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Truong and Dustdar (2010); Mazhelis et al. (2012a); Agarwala et al. (2011); Gonzalez et al. (2013); Adams et al. (2009)
cost of data in/out	Khajeh-Hosseini et al. (2012, 2011); Risch and Altmann (2008); Andrikopoulos et al. (2013); Kashef and Altmann (2012); Kondo et al. (2009); Armbrust et al. (2010); Hajjat et al. (2010); Truong and Dustdar (2010); Opitz et al. (2008); Martens et al. (2012); Mian et al. (2012); Martens and Teuteberg (2012); Kratzke (2012); Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Truong and Dustdar (2010); Mazhelis et al. (2012a); Agarwala et al. (2011); Gonzalez et al. (2013); Adams et al. (2009)
cost of message queuing service, such as Amazon Simple Queue Service (decoupling the components of a cloud application)	Hernández et al. (2013)
cost of load balancing	Hernández et al. (2013); Mazhelis (2012)
cost of Domain Name System (DNS) web service, such as Amazon Route 53	Hernández et al. (2013)
pricing models and utility premium of the cloud provider	Mazhelis (2012); Mazhelis and Tyrväinen (2012); Weinman (2011b,a); Laatikainen et al. (2014); Mazhelis et al. (2012a); Agarwala et al. (2011)
charging/billing period of the cloud provider	Mazhelis (2012); Laatikainen et al. (2014); Mazhelis et al. (2012a)

Cost factor	References
data compression method (incl. compression ratio, compression and decompression time)	Agarwala et al. (2011)
geographical redundancy, auditing and monitoring systems	Gonzalez et al. (2013)
volume discount	Brumec and Vrček (2013); Mazhelis (2012); Mazhelis and Tyrväinen (2012)
Costs factors depending on the interaction between the private and public cloud	
allocation and partitioning costs	Tak et al. (2013); Martens and Teuteberg (2012)
split between the private and public portions of the infrastructure	Mazhelis (2012)
intensity of data communication between private and public portion of the cloud	Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012)
time of using the public portion of the cloud	Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012); Weinman (2011b,a)
Cost factors related to the organizational/environmental/system context	
system and service usage pattern (demand, usage duration and intensity, workload intensity and variance in workload intensity, infrastructure resource requirements, number of users, number of requests, data access frequency)	Andrikopoulos et al. (2013); Kondo et al. (2009); Klems et al. (2009); Kratzke (2012); Bibi et al. (2012); Mastroeni and Naldi (2011); Mazhelis (2012); Mazhelis and Tyrväinen (2011, 2012); Mazhelis et al. (2012b); Weinman (2011c,b,a); Laatikainen et al. (2014); Mazhelis et al. (2012a); Tak et al. (2013); Risch and Altmann (2008); Truong and Dustdar (2010); Kashef and Altmann (2012); Klems et al. (2009); Misra and Mondal (2011); Agarwala et al. (2011); Adams et al. (2009)
storage growth rate	Tak et al. (2013); Brumec and Vrček (2013); Mastroeni and Naldi (2011); Laatikainen et al. (2014); Mazhelis et al. (2012a); Gonzalez et al. (2013)

Cost factor	References
demand predictability	Mazhelis (2012); Weinman (2011c,a); Laatikainen et al. (2014); Mazhelis et al. (2012a); Kashef and Altmann (2012); Klems et al. (2009); Adams et al. (2009)
application type (Sequential or multi-threaded program, Parallel/MPI programs on multiple machines, workflows), application complexity, performance changes, possible security vulnerability, various time delay	Truong and Dustdar (2010); Tak et al. (2013)
application requirements (e.g. runtime environment, database technology, software, load-balancing and redundancy requirements, security, data availability, reliability, scalability, Quality of Service, data sensitivity, work criticality, likelihood of reuse the data)	Klems et al. (2009); Laatikainen et al. (2014); Mazhelis et al. (2012a); Misra and Mondal (2011); Gonzalez et al. (2013); Adams et al. (2009)
system architecture (loadbalancing, autoscaling, processing, storage and backup tier structure, service dependency)	Kratzke (2012); Truong and Dustdar (2010)
enterprise size / size of IT resources	Walker et al. (2010); Misra and Mondal (2011)
technological advances/trends (e.g. growing disk capacity)	Brumec and Vrček (2013); Mastroeni and Naldi (2011); Adams et al. (2009)
pricing/market trends	Walker et al. (2010); Mastroeni and Naldi (2011); Mazhelis (2012); Mazhelis and Tyrväinen (2012); Adams et al. (2009)
Other cost factors pertaining to the lifecycle of a cloud-based system	

Cost factor	References
cost of strategic decision making (incl. identifying the application and infrastructure requirements, technology suitability analysis and stakeholder impact analysis) on adopting the hybrid cloud	Martens et al. (2012); Khajeh-Hosseini et al. (2012); Mazhelis et al. (2012b); Brumec and Vrček (2013)
cost of evaluation and selection of service provider, SLA analysis and negotiation costs	Martens et al. (2012); Martens and Teuteberg (2012); Brumec and Vrček (2013); Mazhelis et al. (2012b); Adams et al. (2009)
costs related to implementation, configuration, customization, integration and migration (incl. cost of migrating an application to cloud on infrastructure level, cost of software porting to the programming API exposed by a cloud, cost related to the transition period when both legacy and cloud environment co-exist)	Sun and Li (2013); Tak et al. (2013); Martens et al. (2012); Martens and Teuteberg (2012); Bibi et al. (2012); Mazhelis et al. (2012b); Beaty et al. (2011); Gonzalez et al. (2013)
costs related to support and maintenance (hardware, software, ex post administration and coordination of the sourcing contracts and SLAs)	Martens et al. (2012); Martens and Teuteberg (2012); Bibi et al. (2012); Walker et al. (2010); Mazhelis et al. (2012b); Gonzalez et al. (2013)
costs related to user training	Martens et al. (2012); Bibi et al. (2012); Mazhelis et al. (2012b)
losses/benefits related to change in QoS (e.g. productivity gains or losses, confidentiality loss, availability loss, integrity loss due to system failure, downtime, security incidents)	Martens et al. (2012); Martens and Teuteberg (2012); Mazhelis et al. (2012b); Beaty et al. (2011); Mian et al. (2012); Khajeh-Hosseini et al. (2011); Misra and Mondal (2011); Adams et al. (2009)
cost of vendor lock-in, cost of switching providers	Abu-Libdeh et al. (2010); Mastroeni and Naldi (2012); Ruiz-Alvarez and Humphrey (2011, 2012)

Table A.3 – continued from previous page

Cost factor	References
cost of backsourcing or discarding	Martens et al. (2012)

Appendix B. Proofs of the propositions

Appendix B.1. Proof of Proposition 3.1

Proof. Let $F(t)$ be an antiderivative of $s(t)$. In this case, Equation (11) can be rewritten as follows:

$$f = u\,p_o\,(F(w) - F(\frac{u-1}{u}\,w) - F(z) + F(\frac{u-1}{u}\,z) - F(w) + F(z + \frac{u-1}{u}\,(w-z)))$$

$$= u\,p_o\,(F(z + \frac{u-1}{u}\,(w-z)) - F(\frac{u-1}{u}\,w) + F(\frac{u-1}{u}\,z) - F(z))$$

$$= u\,p_o\,(\int_{\frac{u-1}{u}\,w}^{z+\frac{u-1}{u}\,(w-z)} s(t)dt - \int_{\frac{u-1}{u}\,z}^{z} s(t)dt). \tag{B.1}$$

Having introduced an auxiliary function $g(t) = s(t + \frac{u-1}{u}\,(w-z))$, Equation (B.1) can be further rewritten in the form

$$f = u\,p_o\,(\int_{\frac{u-1}{u}\,z}^{z} g(t)dt - \int_{\frac{u-1}{u}\,z}^{z} s(t)dt). \tag{B.2}$$

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. $\square$

Appendix B.2. Proof of Lemma 3.1

Proof. Based on Equation (B.1), the first derivative of f with respect to z can be obtained:

$$\frac{\partial f}{\partial z} = u\,p_o\,\left(\frac{\partial}{\partial z}F(t_2) + \frac{\partial}{\partial z}F(t_1) - \frac{\partial}{\partial z}F(z)\right) = u\,p_o\,\left(\frac{1}{u}s(t_2) + \frac{u-1}{u}s(t_1) - s(z)\right). \tag{B.3}$$

It can be shown that, in the interval $(0, w)$, f has only one extremum point where $\frac{\partial f}{\partial z} = 0$.

Let us rewrite the partial derivative in the form:

$$\frac{\partial f}{\partial z} = u \, p_o \, s(z) \left(\frac{1}{u} \frac{s(t_2)}{s(z)} + \frac{u-1}{u} \frac{s(t_1)}{s(z)} - 1 \right). \tag{B.4}$$

It can be seen that $\frac{\partial f}{\partial z} = 0$ iff the condition holds that

$$\frac{1}{u} \frac{s(t_2)}{s(z)} + \frac{u-1}{u} \frac{s(t_1)}{s(z)} = 1. \tag{B.5}$$

Observe that if $z \to 0$, then $t_1 \to 0$ and $s(t_1) \approx s(z)$. Furthermore, since $s(t_2) > s(z)$, it can be easily seen that

$$\frac{1}{u} \frac{s(t_2)}{s(z)} + \frac{u-1}{u} \frac{s(t_1)}{s(z)} > 1, \tag{B.6}$$

and therefore $\frac{\partial f}{\partial z} > 0$.

Similarly, if $z \to w$, then $t_2 \to w$ and $s(t_2) \approx s(w) \approx s(z)$. In this case, since $s(t_1) < s(w)$, it can be easily seen that

$$\frac{1}{u} \frac{s(t_2)}{s(z)} + \frac{u-1}{u} \frac{s(t_1)}{s(z)} < \frac{1}{u} + \frac{u-1}{u} \frac{s(w)}{s(z)}, \tag{B.7}$$

and therefore

$$\frac{1}{u} \frac{s(t_2)}{s(z)} + \frac{u-1}{u} \frac{s(t_1)}{s(z)} < 1, \tag{B.8}$$

and it follows that $\frac{\partial f}{\partial z} < 0$.

Given a monotonically increasing demand function $s(t)$, the left part of the equation (B.5) is a monotonically decreasing function in the range $(\frac{1}{u} + \frac{u-1}{u} \frac{s(w)}{s(z)}, q)$, where $\frac{1}{u} + \frac{u-1}{u} \frac{s(w)}{s(z)} < 1$ and $q > 1$. Therefore, there exists a single value $z_{\max}$ in the interval $(0, w)$ satisfying Equation (B.5). Furthermore, since the derivative changes its sign, $f > 0$ (based on Proposition (3.1)) and since a single extremum point exists at $z_{\max}$, it follows that this extremum corresponds to the maximum of f in the region $(0, w)$.

□

Appendix B.3. Proof of Lemma 3.2

Proof. Taking the derivative of the function in Equation (13) with respect to the refinement point z, we get

$$\frac{\partial f}{\partial z} = p_o \frac{a \, (u-1) \, (w - 2z)}{u}. \tag{B.9}$$

55

Taking the second derivative of the function Equation (13) with respect to the refinement point z, we get

$$\frac{\partial^2 f}{\partial z^2} = -2\, p_o\, \frac{a\,(u-1)}{u}. \tag{B.10}$$

Since $p_o > 0$, $a > 0$ and $u > 1$, it follows that $\frac{\partial^2 f}{\partial z^2} < 0$ and f is concave. The critical point z_{max} where $\frac{\partial f}{\partial z} = 0$ is a maximum point:

$$z_{max} = \frac{w}{2}. \tag{B.11}$$

Because the function f is positive for every $0 < z < w$, at this point the cost savings are the greatest.

1170 $\hspace{11.5cm}\square$

Appendix B.4. Proof of Lemma 3.3

Proof. Consider the exponentially growing demand function in the form

$$s(t) = e^{a\,(\tau+t)+b}, \tag{B.12}$$

where τ is the beginning of the reassessment interval prior to refinement, and where $a > 1$ (if the demand is monotonically increasing) and $b > -1$ (if the demand is positive at the beginning of the storage period) are real numbers.

Given the demand function above, the equation for z_{max} specified in Equation (B.5) can be rewritten as

$$\frac{1}{u}e^{\frac{u-1}{u}a(w-z)} + \frac{u-1}{u}e^{-\frac{1}{u}az} = 1. \tag{B.13}$$

Observe that for $z = w/2$, the left part of the equation above simplifies to

$$\frac{1}{u}e^{-\frac{aw}{2u}}\left(e^{\frac{aw}{2}} + u - 1\right) > 1. \tag{B.14}$$

1175 Thus, the left part of Equation (B.13) in the region $(w/2, w)$ is monotonically decreasing from $\frac{1}{u}e^{-\frac{aw}{2u}}\left(e^{\frac{aw}{2}} + u - 1\right) > 1$ to $\frac{1}{u} + \frac{u-1}{u}\frac{s(w)}{s(z)} < 1$. Therefore, there exists a value of $z_{\max} \in (w/2, w)$, satisfying Equation (B.13). $\hspace{0.5cm}\square$

Appendix B.5. Proof of Lemma 3.4

Proof. Let us calculate the derivative of the cost difference function in Equation (13) with respect to the utility premium u:

$$\frac{df}{du} = p_o \, \frac{a \, z \, (w - z)}{u^2}. \tag{B.15}$$

Since $p_o > 0$, $a > 0$ and $0 < z < w$, it follows that the function is monotonically increasing. In addition, since the function f is positive for every $0 < z < w$, the cost savings are increasing as the utility premium increases.

$\square$

Appendix B.6. Proof of Lemma 3.5

Proof. It can be easily seen that $f > 0$ iff

$$u \, p_o \left(\int_{\frac{u-1}{u} \, w}^{w} s(t)dt - \int_{\frac{u-1}{u} \, z}^{z} s(t)dt - \int_{z + \frac{u-1}{u} \, (w-z)}^{w} s(t)dt \right) > c_r. \tag{B.16}$$

The left part of the equation reflects the benefits achievable through the refinement of the reassessment interval, in line with Equation 18. Thus, re-evaluating the storage needs more often reduces the overall hybrid storage cost if the cost benefits due to reassessment are higher than the cost of reassessment.

$\square$

References

Abu-Libdeh, H., Princehouse, L., Weatherspoon, H., 2010. RACS: a case for cloud storage diversity. In: Proceedings of the 1st ACM Symposium on Cloud Computing. ACM, pp. 229–240.

Adams, I. F., Long, D. D., Miller, E. L., Pasupathy, S., Storer, M. W., 2009. Maximizing efficiency by trading storage for computation. In: Proc. of the Workshop on Hot Topics in Cloud Computing.

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: Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, pp. 436–443.

Altmann, J., Kashef, M. M., December 2014. Cost model based service placement in federated hybrid clouds. Future Generation Computer Systems 41, 79–90.

André, F., Kermarrec, A.-M., Le Merrer, E., Le Scouarnec, N., Straub, G., van Kempen, A., 2014. Archiving cold data in warehouses with clustered network coding. In: Proceedings of the Ninth European Conference on Computer Systems. EuroSys '14. ACM, New York, NY, USA, pp. 21:1–21:14.

Andrikopoulos, V., Song, Z., Leymann, F., 2013. Supporting the migration of applications to the cloud through a decision support system. In: IEEE Sixth International Conference on Cloud Computing (CLOUD). 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., Zaharia, M., Apr. 2010. A view of cloud computing. Commun. ACM 53, 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., Frangi, J.-P., 2002. Design and analysis of numerical experiments to compare four canopy reflectance models. Remote Sensing of Environment 79 (1), 72 – 83.

Beaty, K., Naik, V., Perng, C. S., Nov 2011. Economics of cloud computing for enterprise it. IBM Journal of Research and Development 55 (6), 12:1–12:13.

Bibi, S., Katsaros, D., Bozanis, P., 2012. Business application acquisition: on-premise or saas-based solutions? IEEE Software 29 (3), 86–93.

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

58

Borthakur, D., 2007. The hadoop distributed file system: Architecture and design. The Apache Software Foundation, http://hadoop.apache.org/docs/r1.2.1/hdfs_design.pdf.

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

Brumec, S., Vrček, N., 2013. Cost effectiveness of commercial computing clouds. Information Systems 38 (4), 495–508.

Brydon, M., 2006. Evaluating strategic options using decision-theoretic planning. Information Technology and Management 7 (1), 35–49.

Cai, H., Wang, N., Zhou, M. J., 2010. A transparent approach of enabling saas multi-tenancy in the cloud. In: Proceedings of the 2010 6th World Congress on Services. IEEE Computer Society, pp. 40–47.

Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A., 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), 23–50.

Carlton, D. W., 1979. Vertical integration in competitive markets under uncertainty. The Journal of Industrial Economics 27 (3), pp. 189–209.
URL http://www.jstor.org/stable/2098317

Celesti, A., Fazio, M., Villari, M., Puliafito, A., 2016. Adding long-term availability, obfuscation, and encryption to multi-cloud storage systems. Journal of Network and Computer Applications 59, 208 – 218.
URL http://www.sciencedirect.com/science/article/pii/S1084804514002288

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

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

Clements, A. T., Ahmad, I., Vilayannur, M., Li, J., et al., 2009. Decentralized deduplication in san cluster file systems. In: USENIX Annual Technical Conference. pp. 101–114.

Colombo, S. P., Rizzo, C. L., 2009. Numerical Simulation Research Progress. Nova Science Publishers.

Dobre, D., Viotti, P., Vukolić, M., 2014. Hybris: Robust hybrid cloud storage. In: Proceedings of the ACM Symposium on Cloud Computing. SOCC '14. ACM, New York, NY, USA, pp. 12:1–12:14.

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

Fan, C.-T., Wang, W.-J., Chang, Y.-S., 2011. Agent-based service migration framework in hybrid cloud. In: 13th International Conference on High Performance Computing and Communications (HPCC). IEEE, pp. 887–892.

Freytag, P. V., Kirk, L., 2003. Continuous strategic sourcing. Journal of Purchasing and Supply Management 9 (3), 135 – 150.

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

Gonzalez, J., Perez, J. C., Sosa-Sosa, V., Rodriguez Cardoso, J. F., Marcelin-Jimenez, R., 2013. An approach for constructing private storage services as a unified fault-tolerant system. Journal of Systems and Software 86 (7), 1907–1922.

Greenberg, A., Hamilton, J., Maltz, D. A., Patel, P., December 2008. The cost of a cloud: research problems in data center networks. SIGCOMM Comput. Commun. Rev. 39, 68–73.

Gregor, S., Hevner, A. R., 2013. Positioning and presenting design science research for maximum impact. MIS Quarterly 37 (2), 337–356.

Gregory, R. W., Beck, R., Keil, M., 2013. Control balancing in information systems development offshoring projects. MIS Quarterly 37 (4), 1211–1232.

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

Hajjat, M., Sun, X., Sung, Y.-W. E., Maltz, D., Rao, S., Sripanidkulchai, K., Tawarmalani, M., 2010. Cloudward bound: planning for beneficial migration of enterprise applications to the cloud. ACM SIGCOMM Computer Communication Review 40 (4), 243–254.

Han, H., Lee, Y. C., Choi, S., Yeom, H. Y., Zomaya, A. Y., 2013. Cloud-aware processing of MapReduce-based OLAP applications. In: Proceedings of the Eleventh Australasian Symposium on Parallel and Distributed Computing-Volume 140. Australian Computer Society, Inc., pp. 31–38.

Han, Y., 2011. Cloud computing: case studies and total cost of ownership. Information technology and libraries 30 (4), 198–206.

Harrigan, K. R., 1986. Matching vertical integration strategies to competitive conditions. Strategic Management Journal 7 (6), 535–555.

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

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

Hernández, S., Fabra, J., Álvarez, P., Ezpeleta, J., 2013. Cost evaluation of migrating a computation intensive problem from clusters to cloud. In: Economics of Grids, Clouds, Systems, and Services. Springer, pp. 90–105.

Hevner, A. R., March, S. T., Park, J., Ram, S., 2004. Design science in information systems research. MIS Quarterly 28 (1), 75–105.

Hill, T., 1994. Manufacturing Strategy. Richard D. Irwin, Burr Ridge, IL.

Huang, K.-C., Shen, B.-J., 2015. Service deployment strategies for efficient execution of composite saas applications on cloud platform. Journal of

61

Systems and Software 107, 127 – 141.
URL   `http://www.sciencedirect.com/science/article/pii/`
`S0164121215001156`

₁₃₃₀ Hunter, W. G., Hunter, J. S., George, E., 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 ₁₃₃₅ research & analysis by idc, executive summary, IDC, available from http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm (retrieved March 10, 2015).

Jackson, J., March 2014. Price war! Amazon cuts cloud costs to counter Google. IDG News Service, Computerworld, available from ₁₃₄₀ `http://www.computerworld.com/article/2489105/cloud-computing/` `price-war--amazon-cuts-cloud-costs-to-counter-google.html`, last 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. ₁₃₄₅ Organization science 17 (2), 249–261.

Jiekak, S., Kermarrec, A.-M., Le Scouarnec, N., Straub, G., Van Kempen, A., 2013. Regenerating codes: A system perspective. ACM SIGOPS Operating Systems Review 47 (2), 23–32.

Juan-Verdejo, A., Baars, H., 2013. Decision support for partially moving ₁₃₅₀ applications to the cloud: the example of business intelligence. In: Proceedings of the 2013 international workshop on Hot topics in cloud services. ACM, pp. 35–42.

Kabbedijk, J., Bezemer, C.-P., Jansen, S., Zaidman, A., 2015. Defining multi-tenancy: A systematic mapping study on the academic and the ₁₃₅₅ industrial perspective. Journal of Systems and Software 100, 139 – 148.
URL   `http://www.sciencedirect.com/science/article/pii/`
`S0164121214002313`

Kashef, M. M., Altmann, J., 2012. A cost model for hybrid clouds. In: Economics of Grids, Clouds, Systems, and Services. Springer, pp. 46–60.

Khajeh-Hosseini, A., Greenwood, D., Smith, J. W., Sommerville, I., 2012. The cloud adoption toolkit: supporting cloud adoption decisions in the enterprise. Software: Practice and Experience - Special Issue on Software Architectures and Application Development Environments for Cloud Computing 42 (4), 447–465.

Khajeh-Hosseini, A., Sommerville, I., Bogaerts, J., Teregowda, P., 2011. Decision support tools for cloud migration in the enterprise. In: IEEE International Conference on Cloud Computing (CLOUD). IEEE, pp. 541–548.

Klein, A., March 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/`, last retrieved on 29.9.2014.

Klems, M., Nimis, J., Tai, S., 2009. Do clouds compute? a framework for estimating the value of cloud computing. In: Designing E-Business Systems. Markets, Services, and Networks. Springer, pp. 110–123.

Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P., 2009. Cost-benefit analysis of cloud computing versus desktop grids. In: Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on. IEEE, pp. 1–12.

Koomey, J., Brill, K., Turner, P., Stanley, J., Taylor, B., 2007. A simple model for determining true total cost of ownership for data centers. Uptime Institute White Paper, Version 2.

Kotlarsky, J., Scarbrough, H., Oshri, I., 2014. Coordinating expertise across knowledge boundaries in offshore-outsourcing projects: The role of codification. MIS Quarterly 38 (2), 607–627.

Kratzke, N., 2012. Cloud computing costs and benefits. In: Cloud Computing and Services Science. Springer, pp. 185–203.

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

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

Lacity, M., Khan, S., Willcocks, L., 2009. A review of the it outsourcing literature: insights for practice. The Journal of Strategic Information Systems 18 (3), 130–146.

Lacity, M. C., Solomon, S., Yan, A., Willcocks, L. P., 12 2011. Business process outsourcing studies: a critical review and research directions. Journal of Information Technology 26 (4), 221–258.

Leong, L., Toombs, D., Gill, B., Petri, G., Haynes, T., May 2014. Magic quadrant for cloud infrastructure as a service. Tech. rep., Gartner.

Lima, D., Moura, B., Oliveira, G., Ribeiro, E., Araujo, A., Holanda, M., Togawa, R., Walter, M., May 2014. A storage policy for a hybrid federated cloud platform: A case study for bioinformatics. In: Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on. pp. 738–747.

Long, S., Zhao, Y., Chen, W., 2014. A three-phase energy-saving strategy for cloud storage systems. Journal of Systems and Software 87, 38 – 47.
URL http://www.sciencedirect.com/science/article/pii/S0164121213002148

Malawski, M., Figiela, K., Nabrzyski, J., 2013. Cost minimization for computational applications on hybrid cloud infrastructures. Future Generation Computer Systems 29 (7), 1786 – 1794.

Mao, B., Jiang, H., Wu, S., Fu, Y., Tian, L., 2014. Read-performance optimization for deduplication-based storage systems in the cloud. ACM Transactions on Storage (TOS) 10 (2), 6.

Martens, B., Teuteberg, F., 2012. Decision-making in cloud computing environments: A cost and risk based approach. Information Systems Frontiers 14 (4), 871–893.

Martens, B., Walterbusch, M., Teuteberg, F., 2012. Costing of cloud computing services: a total cost of ownership approach. In: 45th Hawaii International Conference on System Science (HICSS). IEEE, pp. 1563–1572.

Mastroeni, L., Naldi, M., sept. 2011. Long-range evaluation of risk in the migration to cloud storage. In: IEEE 13th Conference on Commerce and Enterprise Computing (CEC). pp. 260 –266.

Mastroeni, L., Naldi, M., 2012. Pricing of insurance policies against cloud storage price rises. ACM SIGMETRICS Performance Evaluation Review 40 (2), 42–45.

Mazhelis, O., 2012. Costs of using hybrid cloud infrastructure: Towards a general framework. In: Software Business. Springer, pp. 261–266.

Mazhelis, O., Fazekas, G., Tyrvainen, P., 2012a. Impact of storage acquisition intervals on the cost-efficiency of the private vs. public storage. In: 5th International Conference on Cloud Computing (CLOUD). IEEE, pp. 646–653.

Mazhelis, O., Tyrväinen, P., 2011. Role of data communications in hybrid cloud costs. In: Proceedings of the 37th EUROMICRO Conference on Software Engineering and Advanced Applications.

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

Mazhelis, O., Tyrväinen, P., Tan, K. E., Hiltunen, J., 2012b. Assessing cloud infrastructure costs in communications-intensive applications. In: Cloud Computing and Services Science. Springer, pp. 277–293.

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

Mell, P., Grance, T., September 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/`.

Mian, R., Martin, P., Zulkernine, F., Vazquez-Poletti, J. L., 2012. Estimating resource costs of data-intensive workloads in public clouds. In: Proceedings of the 10th International Workshop on Middleware for Grids, Clouds and e-Science. MGC '12. ACM, ACM, New York, NY, USA, pp. 3:1–3:6.

Misra, S. C., Mondal, A., 2011. Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding return on investment. Mathematical and Computer Modelling 53 (3), 504–521.

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

Mols, N. P., Hansen, J. R., Villadsen, A. R., 2012. Plural governance: The effect of internal production on supplier performance. Industrial Marketing Management 41 (5), 874–885.

Muniswamy-Reddy, K.-K., Seltzer, M., 2010. Provenance as first class cloud data. ACM SIGOPS Operating Systems Review 43 (4), 11–16.

Nufire, T., July 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/`, last retrieved on 29.9.2014.

Opitz, A., König, H., Szamlewska, S., 2008. What does grid computing cost? Journal of Grid Computing 6 (4), 385–397.

Palankar, M. R., Iamnitchi, A., Ripeanu, M., Garfinkel, S., 2008. Amazon s3 for science grids: a viable solution? In: Proceedings of the 2008 international workshop on Data-aware distributed computing. ACM, pp. 55–64.

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

Parmigiani, A., 2007. Why do firms both make and buy? an investigation of concurrent sourcing. Strategic Management Journal 28 (3), 285–311.

Parmigiani, A., 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), 1065–1091.

Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S., Dec. 2007. A design science research methodology for information systems research. J. Manage. Inf. Syst. 24 (3), 45–77.

Porter, M., 1980. Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press, New York.

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

Reine, D., Kahn, M., May 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`, last retrieved on 29.9.2014.

RightScale, 2014. Rightscale 2014 state of the cloud report. Available online at `http://www.rightscale.com/lp/2014-state-of-the-cloud-report`, last retrieved on 29.9.2014.

Risch, M., Altmann, J., 2008. Cost analysis of current grids and its implications for future grid markets. In: Altmann, J., Neumann, D., T., F. (Eds.), Proceedings of the 5th international workshop on Grid Economics and Business Models. GECON '08. Springer-Verlag, Berlin, Heidelberg, pp. 13–27.

Ruiz-Alvarez, A., Humphrey, M., 2011. An automated approach to cloud storage service selection. In: Proceedings of the 2nd international workshop on Scientific cloud computing. ACM, pp. 39–48.

Ruiz-Alvarez, A., Humphrey, M., 2012. A model and decision procedure for data storage in cloud computing. In: The 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). IEEE, pp. 572–579.

Sako, M., Chondrakis, G., M. Vaaler, P., 2013. How do firms balance making and buying? the case of legal services sourcing by fortune 500 companies. ForthComingAvailable from `https://test-intranet.law.ox.ac.uk/ckfinder/userfiles/files/Paper%20-%20Sako%20Chondrakis%20Vaaler(1).pdf`, retrived on 8.8.2014.

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

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

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

Shifrin, M., Atar, R., Cidon, I., 2013. Optimal scheduling in the hybrid-cloud. In: IFIP/IEEE International Symposium on Integrated Network Management (IM 2013). IEEE, pp. 51–59.

Silva, L. A. B., Costa, C., Oliveira, J. L., 2013. A common {API} for delivering services over multi-vendor cloud resources. Journal of Systems and Software 86 (9), 2309 – 2317.
URL http://www.sciencedirect.com/science/article/pii/S0164121213001052

Smith, J. E., Nair, R., 2005. The architecture of virtual machines. Computer 38 (5), 32–38.

Son, I., Lee, D., Lee, J.-N., Chang, Y. B., 2014. Market perception on cloud computing initiatives in organizations: An extended resource-based view. Information and Management 51 (6), 653 – 669.
URL http://www.sciencedirect.com/science/article/pii/S0378720614000597

Sun, K., Li, Y., 2013. Effort estimation in cloud migration process. In: Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on. IEEE, pp. 84–91.

Sun, Y., White, J., Eade, S., Schmidt, D. C., 2015. Roar: A qos-oriented modeling framework for automated cloud resource allocation and optimization. Journal of Systems and Software, –.
URL http://www.sciencedirect.com/science/article/pii/S0164121215001715

Tak, B. C., Urgaonkar, B., Sivasubramaniam, A., 2011. To move or not to move: The economics of cloud computing. In: Proceedings of the 3rd USENIX conference on Hot topics in cloud computing. USENIX Association, pp. 5–5.

Tak, B. C., Urgaonkar, B., Sivasubramaniam, A., 2013. Cloudy with a chance of cost savings. IEEE Transactions on Parallel and Distributed Systems 24 (6), 1223–1233.

¹⁵⁵⁰ Trummer, I., Leymann, F., Mietzner, R., Binder, W., 2010. Cost-optimal outsourcing of applications into the clouds. In: The Second International Conference on Cloud Computing Technology and Science (CloudCom). pp. 135 –142.

Truong, H.-L., Dustdar, S., 2010. Composable cost estimation and moni-¹⁵⁵⁵ toring for computational applications in cloud computing environments. Procedia Computer Science 1 (1), 2175–2184.

Turner, W. P., Seader, J. H., 2006. Dollars per kW plus dollars per square foot are a better datacenter cost model than dollars per square foot alone. The Uptime Institute White Paper.

¹⁵⁶⁰ TwinStrata, 2013. A snapshot into cloud storage adoption. TwinStrata White Paper.

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

van de Water, H., van Peet, H. P., 2006. A decision support model based on the analytic hierarchy process for the make or buy decision in manu-¹⁵⁶⁵ facturing. Journal of Purchasing and Supply Management 12 (5), 258 – 271.

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

Villari, M., Celesti, A., Fazio, M., Puliafito, A., 2014. Evaluating a file frag-¹⁵⁷⁰ mentation system for multi-provider cloud storage. Scalable Computing: Practice and Experience 14 (4), 265–277.

Villari, M., Celesti, A., Tusa, F., Puliafito, A., 2013. Advances in Service-Oriented and Cloud Computing: Workshops of ESOCC 2013, Málaga, Spain, September 11-13, 2013, Revised Selected Papers. Springer Berlin ¹⁵⁷⁵ Heidelberg, Berlin, Heidelberg, Ch. Data Reliability in Multi-provider Cloud Storage Service with RRNS, pp. 83–93.
URL http://dx.doi.org/10.1007/978-3-642-45364-9_8

Wagner, S. M., Bode, C., 2006. An empirical investigation into supply chain vulnerability. Journal of Purchasing and Supply Management 12 (6), 301–¹⁵⁸⁰ 312.

Walker, E., Brisken, W., Romney, J., April 2010. To lease or not to lease from storage clouds. Computer 43, 44–50.

Walterbusch, M., Martens, B., Teuteberg, F., 2013. Evaluating cloud computing services from a total cost of ownership perspective. Management Research Review 36 (6), 613–638.

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

Wang, W.-J., Lo, Y.-M., Chen, S.-J., Chang, Y.-S., 2012. Intelligent application migration within a self-provisioned hybrid cloud environment. In: Computer Science and Convergence. Springer, pp. 295–303.

Weddle, C., Oldham, M., Qian, J., Wang, A.-I. A., Reiher, P., Kuenning, G., Oct. 2007. Paraid: A gear-shifting power-aware raid. Trans. Storage 3 (3). URL http://doi.acm.org/10.1145/1289720.1289721

Weigelt, C., Sarkar, M., 2012. Performance implications of outsourcing for technological innovations: managing the efficiency and adaptability trade-off. Strategic Management Journal 33 (2), 189–216.

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

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

Weinman, J., January 8 2011b. Mathematical proof of the inevitability of cloud computing, working paper, available from http://www.joeweinman. com (last retrieved on February 28, 2012).

Weinman, J., January 7 2011c. Time is money: The value of "on-demand", working paper, available from http://www.joeweinman.com (last retrieved on February 28, 2012).

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

Willcocks, L., Lacity, M., 2012. The New IT Outsourcing Landscape: From Innovation to Cloud Services. Palgrave Macmillan.

Williamson, O., 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), 105–125.

Xie, T., 2008. Sea: A striping-based energy-aware strategy for data placement in raid-structured storage systems. IEEE Transactions on Computers 57 (6), 748–761, cited By 49.
URL `http://www.scopus.com/inward/record.url?eid=2-s2.0-44049099115&partnerID=40&md5=587ac364edae1bfdbd3b07c7b2d6bf40`

Yuan, D., Yang, Y., Liu, X., Chen, J., 2010a. A cost-effective strategy for intermediate data storage in scientific cloud workflow systems. In: Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on. IEEE, pp. 1–12.

Yuan, D., Yang, Y., Liu, X., Chen, J., 2011. On-demand minimum cost benchmarking for intermediate dataset storage in scientific cloud workflow systems. Journal of Parallel and Distributed Computing 71 (2), 316–332.

Yuan, D., Yang, Y., Liu, X., Zhang, G., Chen, J., 2010b. A data dependency based strategy for intermediate data storage in scientific cloud workflow systems. Concurrency and Computation: Practice and Experience 24 (9), 956–976.

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

Zhu, Q., Chen, Z., Tan, L., Zhou, Y., Keeton, K., Wilkes, J., Oct. 2005. Hibernator: Helping disk arrays sleep through the winter. SIGOPS Oper. Syst. Rev. 39 (5), 177–190.
URL `http://doi.acm.org/10.1145/1095809.1095828`

Zhu, Y., Hu, H., Ahn, G.-J., Yau, S. S., 2012. Efficient audit service outsourcing for data integrity in clouds. Journal of Systems and Software

85 (5), 1083 – 1095.

1645   URL   http://www.sciencedirect.com/science/article/pii/
S0164121211003189