The Journal of Systems and Software 122 (2016) 180–201

Contents lists available at ScienceDirect

The Journal of Systems and Software

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

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

ARTICLE INFO

Article history:
Received 23 September 2015
Revised 3 May 2016
Accepted 3 September 2016
Available online 13 September 2016

Keywords:
Hybrid cloud storage
Volume variation
Reassessment interval
Acquisition interval
Plural governance
Concurrent sourcing

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.

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. Furthermore, the hybrid cloud infrastructure can be seen as an instantiation of concurrent sourcing, which is a simultaneous use of market con-

* Corresponding author.
E-mail addresses: gabriella.laatikainen@jyu.fi (G. Laatikainen), mazhelis@jyu.fi (O. Mazhelis), pasi.tyrvainen@jyu.fi (P. Tyrvainen).

http://dx.doi.org/10.1016/j.jss.2016.09.008
0164-1212/

tracting 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 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 benefits in the context of cloud storage resources as well.

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