Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

VM reassignment in hybrid clouds for large decentralised companies: A multi-objective challenge

Takfarinas Saber [a,*], James Thorburn [b], Liam Murphy [a], Anthony Ventresque [a,*]

[a] *Lero@UCD, School of Computer Science, University College Dublin, Dublin 4, Ireland*
[b] *IBM Software Group, Toronto, Canada*

H I G H L I G H T S

- VM reassignment in hybrid clouds is a multi-objective challenge.
- Large companies often run decentralised data centres with competing preferences.
- Our approach performs multi-objective reassignments informed by VM re-placements.
- We propose a hybrid algorithm which outperforms all other (meta-)heuristics.

A R T I C L E I N F O

Article history:
Received 29 January 2017
Received in revised form 29 May 2017
Accepted 16 June 2017
Available online xxxx

Keywords:
Large decentralised data centres
Hybrid clouds
VM reassignment
Multi-objective optimisation
Hybrid algorithms
Interval objective

A B S T R A C T

Optimising the data centres of large IT organisations is complex as (i) they are composed of various hosting departments with their own preferences and (ii) reassignment solutions can be evaluated from various independent dimensions. But in reality, the problem is even more challenging as companies can now choose from a pool of cloud services to host some of their workloads. This hybrid search space seems intractable, as each workload placement decision (seen as running in a virtual machine on a server) is required to answer many questions: can we host it internally? In which hosting department? Are the capital allocators of this hosting department ok with this placement? How much does it save us and is it safe? Is there a better option in the Cloud? Etc. In this paper, we define the multi-objective VM reassignment problem for hybrid and decentralised data centres. We also propose H2–D2, a solution that uses a multi-layer architecture and a metaheuristic algorithm to suggest reassignment solutions that are evaluated by the various hosting departments (according to their preferences). We compare H2–D2 against state-of-the-art multi-objective algorithms and find that H2–D2 outperforms them both in terms of quantity (approx 30% more than the second-best algorithm on average) and quality of solutions (19% better than the second-best on average).

1. Introduction

There has been a proliferation of cloud services in the past years, from virtual machines (VMs) of different flavours to out-of-the-box platforms (e.g., ready to use Machine Learning tools). The many benefits of these cloud solutions [1], including but not limited to their cost, have accelerated the adoption of the Cloud for all sorts of companies [2]. However, modern large and often global organisations seem more reluctant to outsource to the Cloud than small and medium companies [2], with only 17% of them reported having 1000+ VMs in the Cloud. Some of the many reasons for this slow process are the complexity of their software systems [3], the types of their workloads [4] and the privacy/security of their data and products [5] — as well as the distribution and segmentation of the data centres (DCs) of these large and global companies [6]: they possess many hosting departments with 'competing' or even 'conflicting' demands and requirements.

However, the hybrid cloud solution [7], i.e., mixing private infrastructure and public cloud services, is now seen as a potential solution for these large companies [2] as it gives them the benefits of both worlds. On one hand, the Cloud provides quick infrastructure provisioning and deployment [8], while on the other hand, companies can still maintain their own infrastructure when exact characteristics of servers [9], performances [10] and reliability [4] are important.

* Corresponding authors.
E-mail addresses: takfarinas.saber@ucdconnect.ie (T. Saber), jthorbur@ca.ibm.com (J. Thorburn), liam.murphy@ucd.ie (L. Murphy), anthony.ventresque@ucd.ie (A. Ventresque).

http://dx.doi.org/10.1016/j.future.2017.06.015
0167-739X/

VM reassignment in data centres is a known challenge [11], especially as real problems are often *multi-objective* [6,12], i.e., solutions are evaluated by decision makers based on different dimensions or objectives. However, most solutions to this problem in the literature use a weak definition of multi-objective [13], i.e., a linear aggregation of objectives [14–16]. Related work also often miss the complexity of large companies, where the infrastructure is not monolithic but distributed over decentralised hosting departments with their own preferences [17,18]. The optimisation of the companies' IT infrastructure is then more challenging as decisions have to take into account the preferences of the capital allocators of each hosting departments and the interests of the company as a whole. Hybrid cloud solutions, while explored extensively in the literature, have not been addressed in the context described above — multi-objective and decentralised. In particular, the various pricing policies [19,20] and their variability make the question of finding good hybrid reassignment solutions even more challenging.

This paper is the first to address the problem of *multi-objective VM reassignment for large and hybrid decentralised data centres*. In this paper: (i) we consider that data centres are decentralised, i.e., capital allocators of hosting departments express their own preferences, (ii) we also consider the different dimensions of the infrastructure optimisation (cost of running the IT infrastructure, reliability, migration cost) as independent and we propose solutions optimising all these objectives together and (iii) we consider hosting some of the VMs in the Cloud as an option.

Example: As a motivating example, let us suppose a large distributed company with different hosting departments (or VCs, for Virtual data Centres), each of them managed by capital allocators (CAs) who have their own preferences regarding the placement of VMs on their servers. Some CAs may want to reduce the energy footprint of their VC, while others see the licensing cost as more important; or some CAs view the reliability of their infrastructure as crucial (e.g., if they run critical systems), while other CAs consider response time as essential and are mindful of VMs' colocation. When reassigning its VMs, the company can also choose one of the multiple cloud solutions out there and pay the cost (which may only be known partially in advance, i.e., an interval of possible prices) of the deployment of VMs. Either way, possible reassignments have to respect constraints of the IT infrastructure and satisfy the various preferences of CAs.

Fig. 1 shows possible VM reassignments (from a to g) in two dimensions: (i) reliability and (ii) cost of running some of the VMs in the Cloud (the lower the better for both). Note that the cloud cost is not always known in advance so the values in this dimension are often intervals (except for solution e which does not outsource any VM). The broken segments with black dots represent the non-dominated reassignments, i.e., those that are better than the others in at least one particular objective (the good ones). For example, solution f is worse than b on both objectives and is not considered as one of the good reassignments (hence the white dots on f). Notice that while solution d has a worse reliability than c and potentially (given the overlapping intervals of c and d) a worse cloud cost, it is possible that d eventually gets a better cloud cost and we keep d in the list of non-dominated solutions at this stage. At the end, we end up with five good reassignments: a, b, c, d and e. These solutions can then be evaluated locally by decision makers, such as: "a and b have a good reliability but also have high cloud costs, whereas c's reliability is worse, but still not as worse as d's and brings a good reduction in the public cloud cost despite its large interval. Solution e does not cost anything in terms of cloud cost but has the worst reliability amongst them — so we favour c and keep d as a backup plan for when VM prices rise too much".

H2–D2: In this paper, we propose *H2–D2*, a multi-objective VM reassignment system for large and hybrid decentralised DCs (H2–D2 stands for Hybrid algorithm for Hybrid Decentralised Data centres). H2–D2's search space is composed of the different hosting departments of the data centres of a large company and the different public cloud solutions. H2–D2 suggests reassignments of VMs to either the hosting departments or the public cloud locations depending on various objectives: cost of running servers in the hosting departments, migration cost of the VMs, reliability of the servers and cost of hosting VMs in the Cloud. The non-dominated possible reassignments are then suggested to the company's decision makers who can navigate them and choose one solution over another based on their preference and current focus.

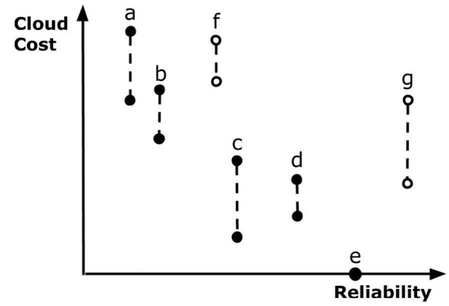

Fig. 1. A decision space in two dimensions (reliability and cost of VMs hosted in the Cloud). The price variability in the Cloud means the solutions are actually in the form of intervals. The (potentially) non-dominated solutions are the broken lines with black dots.

H2–D2 is an adaptation of our previous work [6] to the more challenging (and realistic) cloud environment. We propose a new model/problem definition, with new constraints and a new objective (cloud cost) — this particular objective is an interval objective, the cost of migrating workload to the Cloud being difficult to estimate precisely. This makes the problem significantly different and more challenging than the one we addressed previously.

We compare the performance of H2–D2 against systems with various reassignment algorithms on a realistic data set. We show that H2–D2 outperforms all of them both in terms of quantity (number of non-dominated solutions): H2–D2 gets 29.99% more solutions than the second-best on average, and quality (hypervolume): H2–D2 is ∼20% better than the second-best on average.

After describing the related work (Section 2), we formally define the *multi-objective VM reassignment problem for hybrid and decentralised data centres* (Section 3). This formal problem definition includes a large number of constraints and four objectives, i.e., directions in the search space considered independent. Then we describe our solution, which is based on an architecture for large decentralised data centres [6] but adds the critical and challenging public cloud element (Section 4). This new element makes the problem more complex and H2–D2 has to take into account: (i) individual CAs' preference, (ii) price fluctuation of cloud services and (iii) optimisation of the global infrastructure (at the general managers' level). We then compare H2–D2 against various reassignment algorithms on a realistic data set – inspired from a challenge proposed by Google and modified to make it more realistic for the complex context we address here (Section 5). We show that H2–D2 outperforms (Section 6) all of them both in terms of number of non-dominated solutions (H2–D2 gets ∼30% more solutions than the second-best on average) and quality of these solutions (H2–D2 achieves ∼20% better hypervolume in comparison to the second-best on average).

2. Related work

This section is a short literature survey of three areas: (i) machine/VM reassignment, (ii) VM reassignment in Decentralised and hybrid DCs and (iii) multi-objective VM reassignment.