Ad Hoc Networks 25 (2015) 565–580

Contents lists available at ScienceDirect

Ad Hoc Networks

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

On the feasibility of collaborative green data center ecosystems

Anna Agustí-Torra [a], Frederic Raspall [a], David Remondo [a], David Rincón [a,*], Giovanni Giuliani [b]

[a] Dept. of Telematics Engineering, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
[b] HP Italy Innovation Center, Milan, Italy

ARTICLE INFO

Article history:
Received 7 April 2014
Received in revised form 3 October 2014
Accepted 3 November 2014
Available online 20 November 2014

Keywords:
Energy-efficient computing and networking
Energy-aware data centers
Cloud computing
Green service level agreements

ABSTRACT

The increasing awareness of the impact of the IT sector on the environment, together with economic factors, have fueled many research efforts to reduce the energy expenditure of data centers. Recent work proposes to achieve additional energy savings by exploiting, in concert with customers, service workloads and to reduce data centers' carbon footprints by adopting demand-response mechanisms between data centers and their energy providers. In this paper, we debate about the incentives that customers and data centers can have to adopt such measures and propose a new service type and pricing scheme that is economically attractive and technically realizable. Simulation results based on real measurements confirm that our scheme can achieve additional energy savings while preserving service performance and the interests of data centers and customers.

1. Introduction

The advances in virtualization technology of the last years have shaped the evolution of the IT sector towards a model where organizations no longer sustain their IT infrastructure but rent it from third parties. This "cloud" paradigm is attractive for companies and providers as it allows for a quicker deployment of new services, relieves companies from their maintenance, and can significantly cut down companies' and providers' costs. As a result, we are witnessing a rapid deployment of new data centers (DCs) of increasing size, complexity and heterogeneity in service offerings. In parallel, several studies have estimated that the IT sector is responsible for about 1–2% of the greenhouse gas (GHG) emissions worldwide [1]; DCs accounting for 1.3% of the overall energy consumed [2]. The impact of CO_2 emissions on climate change and the rapid concentration of IT services in DCs have raised the concern about their energy sustainability. Moreover, energy consumption accounts for a significant share of DCs' operational costs. Thus, rather than a concern, energy saving has become a necessity for DCs' economic viability. Many regulations are also being established worldwide to limit corporate emissions, increasing the pressure to cap carbon footprint, and to promote the procurement of power with source mixes with a larger portion of renewable energy [3].

Within a DC, energy consumption reduction is tackled at various levels. At the *component* level, for example, by employing more efficient power supplies, multiple spin rate drives, and memories or CPUs with several energy states or clock gating. At the *system* level, by employing Dynamic Frequency and Voltage Scaling (DFVS) techniques to adapt CPU parameters according to load, adjusting performance not to exceed some power limit (*capping*), introducing power *profiles*, or supporting variable-speed fans. At an *architectural* level, by carefully locating racks, server clusters or network interconnects to optimize the effectiveness of cooling and ventilation. Lastly, at the *operations* level, by trying to adapt IT component utilization to workloads. These techniques seek to exploit components' power attributes by tuning them according to utilization (or even shutting down components). Examples include

---

* Corresponding author. Tel.: +34 934137056; fax: +34 934137007.
E-mail address: drincon@entel.upc.edu (D. Rincón).

http://dx.doi.org/10.1016/j.adhoc.2014.11.004
1570-8705/

consolidation of workloads and virtual servers into fewer, more energy-efficient physical servers (so that some servers and network elements can be switched off), or migrating/re-locating these to other DCs with more advantageous conditions due to the weather or the cost of energy [4].

While DC-holistic approaches can significantly cut down power consumption, further savings and lower environmental impact may be achieved by widening the scope of the solution to span the entire energy consumption chain. In this direction, [5] proposes a concerted strategy toward both energy and CO_2 emission reduction where the three parties in the DC *ecosystem* interact; namely, the DC, its customers, and the DC's energy provider (EP). The approach relies on two ideas: first, a higher *flexibility* in the service agreements between the parties; second, their dynamic *collaboration*. *Flexibility* is supported by the inclusion of energy-related (*green*) clauses in the Service Level Agreements (SLAs), established between the DC and its customers, and the power Supply–Demand Agreements (SDAs), established between DCs and EPs. *Collaboration* allows a party to request the other to adjust its behavior so that some energy/power consumption objective is met, and is implemented by the exchange of messages between a DC and its EP or the DC and its customers. Thus, to avoid resorting to fossil energy sources (e.g. diesel engines) due to a peak in the total energy demanded, an EP can instead request one or more of its customer DCs to temporarily reduce their power consumption; if a DC accepts the request, it will adjust the performance of the hosted IT applications to reduce its energy consumption (e.g. by consolidating virtual machines (VM) in fewer physical servers, delaying the execution of tasks, etc.). In case of energy surplus, the performance of the IT services provided by the DC can be upgraded (by increasing the physical resources allocated to them, advancing the execution of scheduled jobs, or accepting workloads from other DCs).

The idea of reducing power usage in a demand-response (DR) fashion when supply is scarce is not new. Ref. [6] describes a use case where a company reduces its power demand by automatically adjusting lighting, thermal settings or rack power distribution units (PDUs) upon an EP request. While this form of collaboration or those in [5,7] can effectively lead to an environment-friendly DC ecosystem, all the different parties must have an incentive to adopt them –aside from environmental stewardship–, as this may critically impact their businesses.

Refs. [5–7] suggest the incentives to be financial, based on *rewards* or *discounts*. EPs will make discounts to DCs that diminish their power demands. As this may require changing the operating conditions of applications, DC *customers should agree in advance* to reductions in performance and be compensated with lower tariffs accordingly. Otherwise, DCs would have little interest in collaborating with EPs: the economic loss that resulted from compensating customers for SLA violations could outweigh EP's rewards and cause customers to choose other providers. Still, even if DC customers adhered to flexibilize the committed performance levels, finding a suitable reward/pricing scheme is hard given the multilaterality of the DC ecosystem: customers' discounts might not pay off the performance degradation experienced. On the other hand, while a DC would see its energy expenditures reduced, it could still see its profit diminished due to lower revenues resulting from the use of lower tariffs during energy saving episodes. We conclude that, for any energy saving collaborative framework as those described to succeed, *it must permit reaching some equilibrium point where customers are satisfied with the service received (relative to its price), and where DCs and EPs do not see their profits depressed.* A second condition is that the technical requirements to implement it must be simple enough for DCs to adopt it. In this regard, we note that, while [5] establishes a framework and methods to deploy such a paradigm in a technically realizable manner, these do not warrant the above requirement by themselves. This paper seeks to contribute in this direction by proposing a service model and pricing scheme that can be attractive for users, profitable for DCs and EPs and technically feasible, while promoting energy savings in a collaborative manner. We discuss benefits that its adoption could bring to all the parties and assess its advantages and disadvantages with results obtained with a simulator (contrasted with measurements from real DCs), fed with real and synthetic service demands and load patterns, real server parameters and typical energy costs.

This paper is organized as follows. Section 2 briefly describes the components of the DC ecosystem. Section 3 summarizes related work, including ideas in [5] that contextualize and motivate this work. In Section 4 we discuss potential barriers that may preclude the adoption of collaborative measures as the ones discussed. Section 5 presents the motivation, assumptions and idea behind our proposal, and its expected advantages. Section 6 identifies the major variables that need to be considered to measure such advantages and presents how we study their evaluation. Sections 7 and 8 discuss our experimental work: the tool employed, how it works, its parameters and the results obtained. Section 9 presents general conclusions of this research, recommendations concerning the feasibility of the proposal and future directions.

2. Background and terminology

2.1. Data centers (DCs)

A data center is an infrastructure built to provide IT services such as massive data storage, CDN, web, e-mail, server hosting, enterprise-class applications, or on-demand computing. Such services often consist of several components like databases, front-ends, application servers or middleware, increasingly being deployed as virtual machines. The business model varies, but typically a DC is either a service within a large corporation, or a business by itself that resells some of its infrastructure to third parties or "tenants".

While varying in size, purpose and structure, DCs typically consist of large farms of servers executing services for end users. Connectivity is provided by high-speed network equipment. Servers execute application software related to the offered services (e.g. web, e-mail, social network or video) as well as management software (e.g. backup or antivirus). Servers may host several VMs,