Pattern Recognition Letters 67 (2015) 59–65

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Energy wasting at internet data centers due to fear☆

Alejandro Fernández-Montes [a,*], Damián Fernández-Cerero [a], Luis González-Abril [b], Juan Antonio Álvarez-García [a], Juan Antonio Ortega [a]

[a] Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla, Av. Reina Mercedes s/n, Sevilla, 41012, Spain
[b] Departamento de Econoía Aplicada I, Universidad de Sevilla, Av. Ramón y Cajal s/n, Sevilla, 41018, Spain

ARTICLE INFO

Article history:
Available online 3 July 2015

Keywords:
Costs of fear
Energy efficiency
Data center
Grid computing
Cloud computing

ABSTRACT

The fear experienced by datacenter administrators presents an ongoing problem due to the low percentage of machines that they are willing to switch off in order to save energy. This risk aversion can be assessed from a cognitive system. The purpose of this paper is to demonstrate the extra costs incurred by maintaining all the machines of a data center executing continuously for fear of damaging hardware, degradating the service, or losing data. To this end, an objective function which minimizes energy consumption depending on the number of times that the machines are switched on/off is provided. The risk aversion experienced by these data center administrators can be measured from the percentage of machines that they are willing to switch off. It is shown that it is always the best option to turn off machines in order to reduce costs, given a formulation of the cognitive aspects of the fear experienced by datacenter administrators.

1. Introduction

A data center is a facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and various security devices. Large data centers are industrial-scale operations that can consume as much electricity as a small town and sometimes constitute a major source of air pollution in the form of diesel exhaust.

The main purpose of a data center is to run applications, perform tasks or store data. The many examples of internet and computing services performed by data centers include:

The spread of cloud and grid computing paradigms has increased the size and usage of data centers; today there are thousands of data centers worldwide, which means millions of machines in total.

The majority of these facilities are located in the USA (about 25% of the total energy consumption of data centers worldwide [20]) and to a lesser extent in Europe. However, large companies such as Google locate a number of their data centers in high latitudes near the north pole to minimize cooling costs, which represent almost 40% of total energy consumption of these infrastructures [1].

Energy consumption by data centers has grown in the past ten years to 1.5% of worldwide energy consumption [25]. Major companies have therefore addressed their energy-efficiency efforts to areas such as cooling [7], hardware scaling [8] and power distribution [9], thereby slowing down the growth in power consumption in these facilities in recent years as we can see in Fig. 1, which shows the latest predictions.

In addition to these areas of work, saving energy by switching on/off machines in grid computing environments has been simulated using various energy efficiency policies, such as turning off every machine whenever possible, and turning off a number of machines depending on workload [10].

Although it has been demonstrated that about 30% of energy can be saved by applying these energy-aware policies [11], big companies still prefer not to adopt such policies due to their potential impact on the hardware, the possibility of damaging machines, and the costs associated with this hardware deterioration.

The purpose of this paper is to compute the costs imposed by the risk aversion experienced by data center administrators on switching off machines, and to show that even when taking these fears into consideration, some servers of the data center should still be turned off to minimize energy consumption and overall costs.

1.1. Cognitive systems modeling emotions

In psychology [33], emotion is a subjective, conscious experience characterized primarily by psycho-physiological expressions, biological reactions, and mental states. It is influenced by hormones and neurotransmitters, such as dopamine, noradrenaline, serotonin, oxytocin, cortisol, and gamma-aminobutyric acid. Furthermore,

☆ This paper has been recommended for acceptance by Lledó Museros.

* Corresponding author. Tel.: +34 954 559 769; fax: +34 954 557 139.
E-mail address: afdez@us.es (A. Fernández-Montes).

http://dx.doi.org/10.1016/j.patrec.2015.06.018
0167-8655/

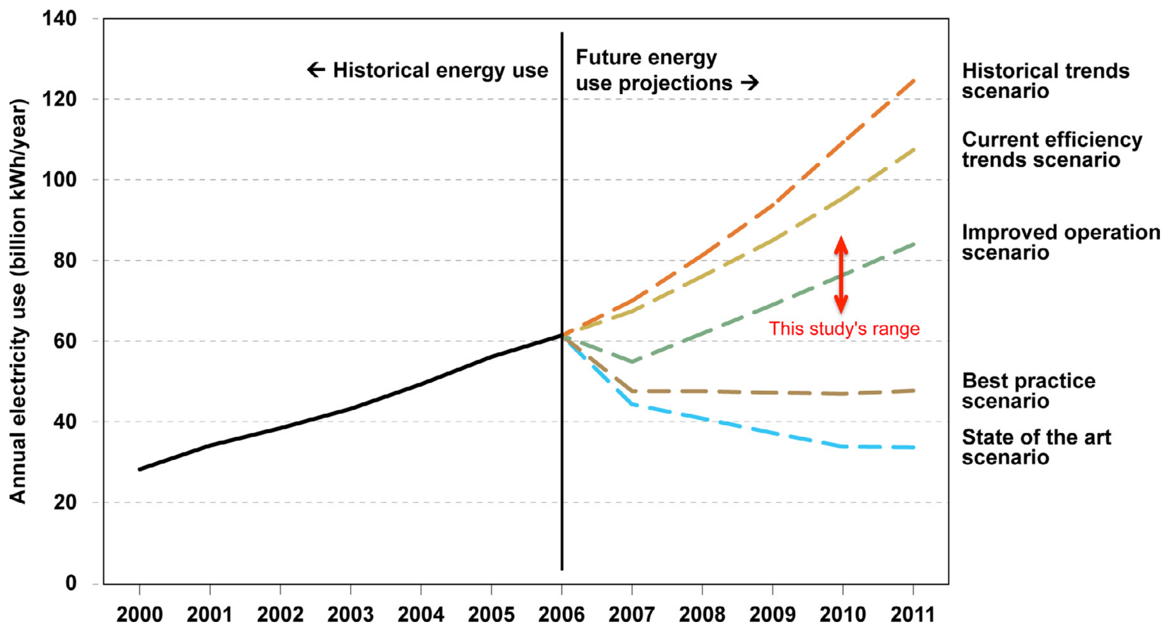

Fig. 1. Data center energy consumption worldwide [24].

neurologists [14] have made progress in demonstrating that emotion is as, or more, important than reason in the process of making decisions. Modeling emotions is a problem tackled from diverse knowledge areas: robot-based systems [6], music [30], videogames and virtual worlds [15] and domain-independent systems [16]. Moreover, emotion recognition systems [18] are on the rise in effective computing research. Data can be obtained from diverse sources: physiological signals (electromyogram, blood pressure, skin conductance, respiration rate and electroencephalogram rate), speech and facial expressions. Focusing on the emotional fear, it appears in response to a specific and immediate danger or a future specific unpleasant event. It can be measured and detected through biosignals such as irregular heart and respiration rate [5,19], visual signals (head gestures, nods and shakes) [17] and facial feature information [34]. Several studies [21] using optogenetic techniques have shown how aversive experiences trigger memories and suggest that combined hebbian and neuromodulatory processes interact to engage associative aversive learning.

Our interest in this paper is to model a function that quantifies the costs of the fear experienced by a datacenter operator on deciding whether a machine must be switched off. According to Michael Tresh, formerly a senior official at Viridity, a company that delivers energy-optimization to data centers: *"Data center operators live in fear of losing their jobs on a daily basis, because the business won't back them up if there's a failure."* The startup 'Power Assure' which is focused on energy management, marketed a technology that enables commercial data centers to safely power down servers when they are not needed, but, as the manager of energy efficiency programs at the utility, Mary Medeiros McEnroe, explains that, even with aggressive programs to entice its major customers to save energy, Silicon Valley Power, a not-for-profit municipal electric utility, failed to persuade a single data center to use that technology. *"It's a nervousness in the I.T. community that something isn't going to be available when they need it"* [13]. Moreover, Power Assure, was dissolved in october 2014. Its technology was based on algorithms that enabled optimal server capacity and application needs to be calculated and to automatically shut off unnecessary capacity or spin up more capacity based on actual application demand. Jennifer Koppy, research director for data center management at International Data Corporation (IDC), said Power Assure's energy management technology was *"extremely forward-looking … they had a superb idea, but I don't think the market is ready yet."*

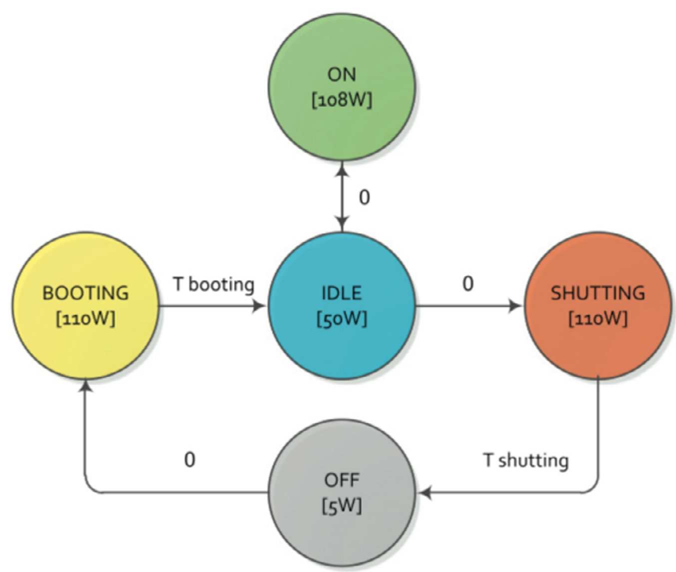

Fig. 2. Life cycle of a data center server [10].

2. Problem analysis

It makes sense that one of the most effective ways to achieve considerable energy savings is to turn off computers that are not being used. Although this idea is generally accepted by users, and hence most personal computers are turned off at night or during periods of low usage, it is seldom implemented in data centers or at enterprise level.

Although the average server utilization within data centers is very low (typically between 10% and 50% [4]), very few companies prefer to turn off the machines that are not in use rather than leaving them in an idle state. While idle servers consume half the energy of those in a state of intensive use [24], this remains a high direct and indirect energy cost due to the increased need for cooling. The several different states through which a machine can pass are shown in Fig. 2. In this state diagram the average power consumption of a common server per CPU in each state is also shown, and the time needed to