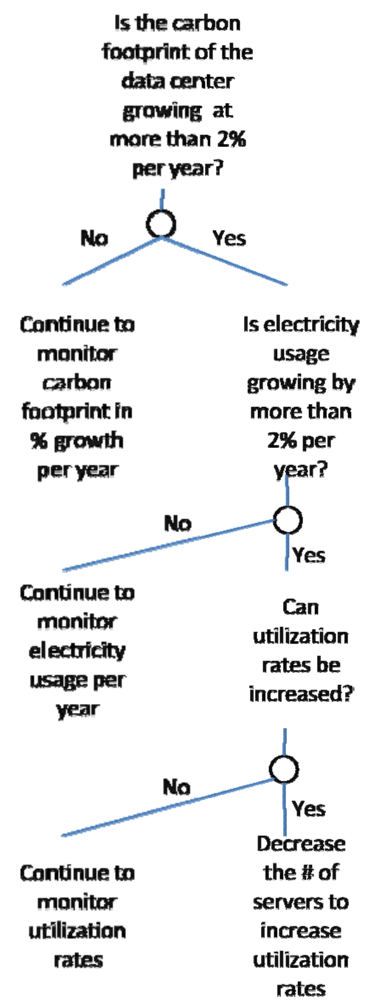
OPTIMIZING STRATEGIES FOR SERVER UTILIZATION IN DATA CENTERS

Michael Pawlish[#], Aparna S. Varde[*] and Stefan A. Robila[*],
[#] Department of Earth and Environmental Studies
[*]Department of Computer Science
Montclair State University, Montclair, NJ, USA[1]

In our research, we propose a decision support system (DSS) for the greening of data centers. We employ decision trees and case based reasoning (CBR) in the analysis which in effect serves to model and manage a complex system such as a data center. A significant aspect of data centers is the utilization of servers which forms an important part of our work and will be the focus here. Data centers are fueled by the incessant demand for information and continuous uninterrupted service by users. Increasingly, the data centers that house the information required to run modern society are becoming more complex to meet users' demands. With the numerous servers of different age and manufacturer that make up data centers, there is a natural tendency over time for server sprawl that use resources less efficiently. Server sprawl is the allocation of more servers to run additional applications. To combat server sprawl, our research suggests the monitoring of utilization rates on servers to encourage higher efficiency rates. Our recommendation will lower the number of servers in the data center that will most likely have the biggest impact on reducing electricity use to both power the servers and eliminate the energy cost to cool the hot air output of the additional servers. By running the servers at an increased utilization rate, the data center will reduce complexity by having fewer servers that run at a higher efficiency rate. This policy of optimizing servers at greater utilization rates will reduce server sprawl, with the following advantages:

- Lower electricity usage that translates to cost savings and a reduced carbon footprint
- Decrease management costs to maintain and service additional servers
- Free up floor space on inefficient or phantom servers
- Provide greater efficiency in the use of resources, e.g., less server sprawl equates to lower cooling costs, decreased electrical bills and less use of natural resources to build additional servers



As a motivating real-world example, we have been monitoring the electrical consumption of our university data center over a two year period, and there has been a steady increase in electrical consumption as the number of servers in our data center increases. According to the data center manager at our campus the server room is approximately 60 to 65% full, which still provides a large percentage for server expansion; however the demands on cooling the data center are becoming stressed. Because of the growth of our data center over the past years, there is a need to add an additional air-conditioner to maintain the desired temperature of approximately 68°F and a relative humidity of approximately 45%. Our data center currently has three large air-conditioners that have been running at full capacity to maintain the desired temperature and humidity level. A fourth large air-conditioner will be added next year to meet the growing needs of our data center, which implies higher costs in addition to the adverse environmental impacts such as increased carbon footprint due to greater power consumption. Likewise, there are issues in other institutions, which pose challenges in optimization. There are several aspects to be studied here such as monitoring the carbon footprint with respect to server utilization.

The carbon footprint of an organization is the estimated total of the output of carbon dioxide into the atmosphere from primarily burning fossil fuels to supply the power for in this case the data center. The formula to calculate the carbon footprint is as follows;

*Carbon footprint = Electricity usage/year * U.S. CO_2 national emissions average[2] *1 metric ton/2,204.6 lbs*

To calculate the carbon footprint of our data center we substitute in our values as follows:

*Carbon footprint = 788,400 kWh/year * 1.34 lbs/kWh * 1 metric ton/2,204.6lbs = 479.21 metric tons/year*

The electrical usage value of 788,400kWh/year is a combined value of the energy use of our servers and air-conditioning system for 2010. The value 1.34 lbs/kWh is the national average of US CO_2 emissions according to Schulz (2009) [2]. A metric ton conversion ratio is used because CO_2 emissions are commonly expressed in the international community in metric tons.

A data center is thus a complex system to maintain, considering various issues. One of these is also the diversity of manufacturers. For example, in our data center, we currently have around 350 servers of different manufacturers and vintage. Servers in most institutions are typically upgraded every three to four years. An important goal of our research therefore is to examine the utilization rates of such a variety of servers in data centers to determine which servers are being under-utilized. We envision a next generation data center that balances the number of servers with a higher utilization rate to decrease the need for additional power, cooling and floor space. Accordingly, we build models using decision trees and CBR as the basis for DSS development. As we propose optimal strategies for data center server utilization, these will be depicted using CBR for case-specific analysis decision trees for more general analysis. We provide an example of a decision tree in the adjoining figure to analyze some aspects of server utilization.

We claim that this data analysis and DSS will be beneficial in terms of providing support to assist data centers managers and other users in making decisions to optimize strategies for better utilization of servers. Our research would have the broader impact of greening the environment.

1 This work is supported by a PSE&G grant ($218,000) awarded from Jan 2011 to Dec 2012 to Dr. S. Robila, Dr. A. Varde & Dr. M. Weinstein (Montclair State Univ.)

2 Schulz, G. (2009) *The Green and Virtual Data Center,* Auerbach Publications, Boca Raton, Fl., Pg. 11.