Creating Software to Automate Decision Making in Data Replication
(Technical Topic)

Evaluating Claims in Green Computing Using Greenwashing Frameworks
(STS Topic)

A Thesis Project Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Assad Aijazi

Fall, 2019

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Andrew Grimshaw, Professor of Computer Science, Department of Computer Science

Introduction

Data centers, large facilities used to home computational storage and server equipment, form the backbone of the computing world. Companies such as Amazon, Google, and Facebook rely on these data centers to perform computational tasks such as serving user web applications, storing user data, and hosting cloud services (Khan and Zomaya, 2015). The electricity consumed by these data centers is significant. According to a report by the Lawrence Berkeley National Laboratory, data centers are projected to use approximately 73 billion KWh of electricity in the United States in 2020 (Shehabi et. al, 2016, p. ES-3). This large electricity usage has environmental consequences due to the carbon emitted from generating electricity. An NRDC report estimates that electricity usage from data centers in the United States could result in emission of nearly 150 million metric tons of carbon pollution annually by 2020 (Delforge and Whitney, 2014, p. 5).

A potential solution to the environmental consequences of utilizing data centers is to power them using renewable energy. Lancium is a company that is attempting to implement this solution by using wind farms to power the data centers behind its cloud computing service. However, since the power output of these wind farms is variable based on weather, data center operations may have to be temporarily suspended until levels rise again. Having multiple copies of data (i.e. *data replication*) across several data centers may help mitigate some of the negative consequences associated with these suspensions. My technical work will be looking at how some of the decisions in the data replication process can be automated in order to make data replication more feasible.

Lancium is but one example of a larger movement to use computing to tackle environmental challenges. With any product or service that addresses environmental concerns,

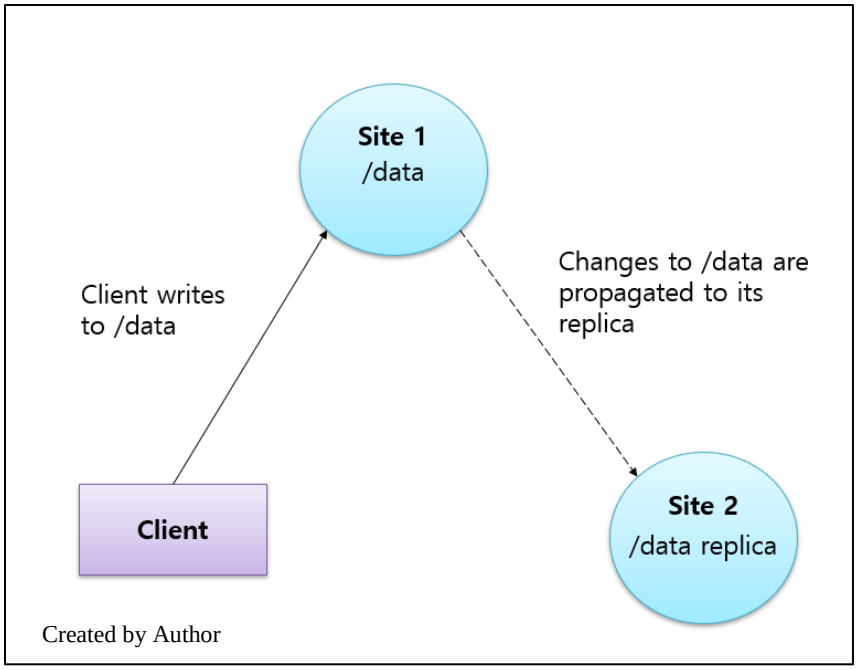
there is a risk of using the label of green as a marketing ploy to mislead consumers. Thus, my STS thesis will address how a framework of so-called *green washing* can be applied to case studies in green computing.

Technical Project: Creating Software to Automate Decision Making in Data Replication

Lancium utilizes a network of geographically-dispersed data centers in order to host its computing grid. These data centers are powered using electricity generated using nearby wind farms. The amount of electricity that can come from these farms vary greatly depending on how much wind there is on a particular day. As a result, data centers may need to scale down jobs and tasks that they are executing if there is not enough power to support them. Thus, there is potential that data on a particular data center may not be available for use by other clients if that particular data center is suspended.

A potential solution that will ensure data availability is to utilize data replication. Replicating data involves having multiple copies of data located at different data centers at different geographical locations. When one copy is modified, the changes are propagated to all of the other copies. An example of how data replication works can be found in Figure 1. Having data replicants at different locations means that if data required by a client is unavailable at one particular location, the data can still be accessed via a replicant at a different location. Another reason to replicate data at different sites is to get a performance boost. A client may need to access data from a site that is geographically far away from where they are located, which will most likely make the access time of the data slower. In the scenario where the data needs to be accessed multiple times, a replicant can be made to a closer site so that the time for each access will be faster.

Figure 1: Data Replication Example



In this example, folder /data exists at Site 1 and a data replica of /data exists at Site 2. When the client modified /data, the changes are also sent to Site 2. If Site 1 were to be unavailable, the client could still access the contents of /data via the replica located at Site 2.

However, data replication is limited due to storage constraints. Since there is a fixed amount of memory at each data center, it is impossible to create replicants of all data. As a result of this constraint, there are several decisions that need to be made when replicating data, such as when to create a data replicant, what data to replicate, and at which site to store that replicant. Additionally, it may also be required to delete extra data replicants due to storage constraints. When deciding what data replicants to delete, it is very important to make sure that if a data replicant is deleted from a specific site, there is at least one other copy that exists on another site; otherwise the data will be deleted permanently. All of these decisions are hard to make manually, especially in real time. Thus, my technical project will involve writing software to automate this decision making by monitoring factors such as data usage and available memory. The end goal will be that data replication will be done automatically when it makes sense to do so based on the information available.

STS Topic: Evaluating Claims in Green Computing Using Greenwashing Frameworks

Placing my technical work in a larger context, Lancium and its work can be seen as an example of what is known as the *green computing* movement. Green computing can be defined as "the study of designing, manufacturing, using, and disposing of computing devices in a way that reduces environmental impact" (Khan et al, 2019, p. 38). The earliest example that is often attributed to the green computing movement is the launch of the Energy Star program by the Environmental Protection Agency in 1992; Energy Star is a government-backed label that products can obtain assuming they meet some government-set standards for energy efficiency (ENERGY STAR Overview, n.d.). According to Bud E. Smith, there are several motivations for companies to become engaged with green computing: boosting the company's reputation as a marketing effort, cutting costs by wasting less energy and fewer resources, and pursuing altruistic goals of leaving less of a negative impact on the environment (Smith, 2019).

The motivation of using green computing as a marketing tactic can become problematic because it lends itself to what is known as *greenwashing.* Greenwashing is when "environmental claims are used for products that are not inherently environmentally friendly", or when "green advertisements present confusing truths that lack substantive information about the real environmental attributes of their products" (Schmuck et al., 2018, p. 127). Greenwashing has negative consequences; in addition to customers being mislead on the actual environmental impact of the products they use, greenwashing can also have the effect of undermining consumers' confidence in green advertising (Chen and Chang, 2013).

While the negative impacts of greenwashing are known, identifying greenwashing in specific cases can be difficult. To aid in greenwashing identification, Schmuck, Matthes, and Naderer have postulated a framework that draws upon the affect-reason-involvement (ARI)

model (Schmuck et al., 2018). The ARI model suggests that two different mechanisms of persuasion drive attitude formation: rational cognition and affect (Buck et al., 2004). The purpose of this framework is to describe key characteristics of messages in greenwashing ads and their effects on consumers' attitude formation (Schmuck et al., 2018). Another useful framework for identifying greenwashing is the FTC Guides for the Use of Environmental Marketing Claims. The purpose of these so called "Green Guidelines" is to "outline general principles that apply to all environmental provide guidance regarding many specific environmental benefit claims marketing claims", "explain how reasonable consumers likely interpret each such claim, describe the basic elements necessary to substantiate it, and present options for qualifying it to avoid deception" (The Green Guides: Statement of Basis and Purpose, 2012, p. 1-2). Combining these two frameworks creates a means to evaluate whether specific examples of statements or claims can be deemed as greenwashing.

My STS research will thus focus on how specific case studies in the green computing can be evaluated for greenwashing using the aforementioned frameworks. Some examples of case studies that I could consider would be the Energy Star program mentioned earlier, and public company claims and statements, such as Google's claim that they have achieved carbon neutrality (Google Environmental Report, 2019). Additional sources of case studies to be evaluated can also come trade journals and magazines. While it will be difficult to make any generalizations about greenwashing in the entire green computing movement just from a few specific examples, this research can be a good first step to understanding how exactly how prevalent greenwashing exists within the scope of the green computing movement evaluated in this research.

Conclusion

With the completion of my technical work, the anticipated deliverable is software that can help automate decision making for performing data replication across Lancium's data centers. With the completion of my STS research, the anticipated deliverable is an increased understanding of the role greenwashing plays within the green computing movement. The completion of both these tasks will contribute to the resolution of the issue of electricity usage by data centers. Accomplishing my technical work will be a step further in ensuring that the technical challenges that arise from Lancium's wind-powered data centers are resolved. Accomplishing my STS research, meanwhile, will help provide a broader framework for how companies like Lancium can successfully advertise their accomplishments in sustainable computing without misleading consumers.

References

Buck, R., Anderson, E., Chaudhuri, A., & Ray, I. (2004). Emotion and reason in persuasion: Applying the ARI model and the CASC Scale. *Journal of Business Research*, *57*(6), 647–656.

Chen, Y.-S., & Chang, C.-H. (2013). Greenwash and Green Trust: The Mediation Effects of Green Consumer Confusion and Green Perceived Risk. *Journal of Business Ethics*, *114*(3), 489–500. https://doi.org/10.1007/s10551-012-1360-0

Delforge, P., & Whitney, J. (2014). Data center efficiency assessment—Scaling up energy efficiency across the data center industry: Evaluating key drivers and barriers. NRDC, Anthesis.

ENERGY STAR Overview. (n.d.). Retrieved from Energy Star website: https://www.energystar.gov/about

Google Environmental Report 2019. (2019, September). Retrieved from Google Sustainability website: https://sustainability.google/reports/environmental-report-2019/

Khan, S. N., Aljaberi, M. A., & Muammar, S. (2019). Success factors model for green computing implementations. *International Journal of Technology Management & Sustainable Development*, *18*(1), 37–54.

Khan, S., & Zomaya, A. (2015). *Handbook on Data Centers*. Retrieved from https://doi-org.proxy01.its.virginia.edu/10.1007/978-1-4939-2092-1

Schmuck, D., Matthes, J., & Naderer, B. (2018). Misleading Consumers with Green Advertising? An Affect—Reason—Involvement Account of Greenwashing Effects in Environmental Advertising. *Journal of Advertising*, *47*(2), 127–145. https://doi.org/10.1080/00913367.2018.1452652

Shehabi, A., Smith, S., Sartor, D., Brown, R., Herrlin, M., Koomey, J., … Lintner, W. (2016). United States Data Center Energy Usage Report. Retrieved from Lawrence Berkeley National Laboratory website: http://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf

Smith, B. E. (2013). *Green Computing* (1st ed.). Retrieved from https://doi.org/10.1201/b15098

The Green Guides: Statement of Basis and Purpose. (2012, October). Retrieved from Federal Trade Commission website: https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-issues-revised-green-guides/greenguidesstatement.pdf