The Environmental Impact of the Shift to Cloud Computing

Logan O'Neal*

The University of Tennessee

loneal7@vols.utk.edu

May 15, 2021

Abstract

In recent years, cloud computing services have become increasingly popular primarily due to their scalablity and reliability. Due to this, it is becoming clear that there may exist environmental concerns within this rapidly expanding industry. Datacenters use a large amount of electricity and pose a significant risk to the environment. The question remains as to whether these companies are doing a service or creating environmental hazards by consolidating our computational resources. In order to answer this question, the original effects of individually hosted servers by companies must be compared to the present centralized, remote state. Through computational methods, this paper concludes that (a) the location of datacenters is crucial to their environmental effect, (b) idle machines are a large contributor to the carbon footprint of cloud computing, and (c) hosting companies still have room for growth and novel solutions still exist within this area.

I. Introduction

The modern climate crisis is here, and the evidence is beyond dispute. To save the earth people are coming together more than ever to take steps to reduce their carbon footprint. They are practicing recycling, creating sustainable policy measure, and even buying electric cars. While many people are working to help save our planet by reducing the ways in which they knowingly use resources, there are many ways in which people use resources that they do not even know exist. One of the largest of these secret resources lies deep within almost all supply chains today and is the servers that run the internet. During the last ten years companies have rapidly been migrating from locally hosted servers to the use of large "cloud computing" facilities. These facilities offer many upsides for companies who no longer allocate resources to maintain their own computational infrastructure. When compared to local machines, these resources are faster, more scalable, and safer alternatives to local hosting [1]. Due to the recent market capital-

*A thank you or further information

ization by hosting companies such as Amazon, Google, and GoDaddy (who collectively host around 30% of all websites) cloud options have become the most cost-effective hosting options for many companies [3]. In 2020 companies spent $188 billion on cloud hosting and this number is expected to more than double because of the COVID-19 pandemic where many companies discovered the advantages of remote, cloud-based work models [2]. There are three million data centers in America alone, accounting for over 2% of the total Unites States energy use [3]. A single server, on average, will produce more carbon dioxide than a car and web hosting alone pollutes more than the entire airline industry [3]. This paper looks into the collective impacts of the cloud hosting industry, including the energy efficiency of cloud hosting when compared to locally hosted systems. In addition, this paper provides an overview of the use cases of these systems and qualitatively analyzes the value of these systems in the context of their environmental effects.

II. Methods

This study will use data collected directly from large hosting companies. To get a bigger picture of the total impacts of the global cloud system some extrapolation will be required, but the efficiency of scaled systems will be accounted for based on similar larger systems. In addition, this paper will provide an overview of the most common use cases for these large-scale systems. These use cases will be based on actual data transmitted as it is the best metric available (as opposed to actual computational use, which is harder to quantitatively account for)

III. Hypothesis

Remote, "cloud" based data centers will be found to have a lower net power consumption than locally hosted systems. These data-centers will be located in areas with dense populations and cheap power supplies, potentially making them less environmentally friendly than the local systems which would otherwise be more evenly distributed, using a wider range of power grids

IV. Results

Ideally, this paper will yield insight into the level of impact that the shift to and use of large-scale cloud hosting companies has had and will continue to have on our planet's ecosystem. At the same time, the paper will allow readers to make judgements about whether these impacts are justified based on what these systems are being used for.

V. Experimentation

Two datasets were collected developed for this project. The first was queried directly from Shodan into a set of 1200 of the most used datacenters in the United States. The second was scraped from the United States Energy Information Administration website and includes 12,412,180 individual power sources labeled by power generation source and location. These two sets were cross compared in order to make inferences regarding power consumption by datacenters in the United States.

VI. Discussion

i. Benefits of the Cloud

Cloud computing allows companies to streamline their workflows across work locations while also being more scalable and cost effective than locally hosted systems. The savings that companies create by not purchasing their own datacenters create a savings of about 30% when compared to cloud systems [4]. This effect is even greater for smaller, more inefficient companies who typically reduce power usage at rates upwards of 90% by switching to cloud-based systems [7]. It has also been shown that cloud computing increases a company's overall productivity due to loss of downtime and system interconnection [3].

Here is a summary of potential benefits that cloud computing offers to consumers:

1. Cost Savings
2. Security
3. Flexibility
4. Mobility
5. Insight
6. Increased Collaboration
7. Quality Control
8. Disaster Recovery
9. Loss Prevention

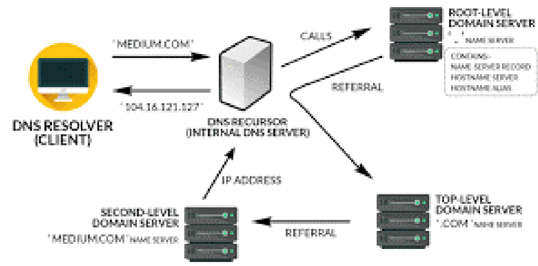

Figure 1: *HTTPS/DNS.*

10. Automatic Software Updates
11. Competitive Edge
12. Environmental Sustainability (Possibly?)

ii. Environmental Concerns

Figure 1 Despite the clear advantages of cloud computing for many companies, the current climate crisis has brought many hosting companies under scrutiny due to the aggressive practices required to keep these large, resource heavy systems running cost effectively. For example, Google was recently brought to the fore front of many tech circles for possible OSHA violations after a report was publishing stating they keep their server rooms at an upwards of 115 degrees Fahrenheit on a regular basis to save money on cooling systems [5]. As a society we are and have always been very disconnected from our information technology ecosystems. The average person probably does not think about the number of servers that are involved to make something as seemingly trivial as going to a website (See Fig. 1). These servers are required to be constantly left on at all times to properly function which is a large part of the reason they consume more energy than the fifth most energy consuming country at 662 billion KwH per year [4].

iii. The Cloud: Problem or Solution?

At face value, while the shift to large, conglomerated cloud systems may seem environmentally harmful, many studies have been published showing the opposite effect. The environmental protections are rooted in the scalability and proximately of these systems where issues can be addressed at the source, as opposed to trying to reduce emissions from every single companies locally hosted network of systems. This is due to the power reduced power consumption per unit of computational work done by these systems. These centers are equipped with state-of-the-art technology than systematically distributes workload across nodes on their networks, water-based cooling systems, and modular efficient storage systems that reduce overall power consumption.

The environmental efficiency of cloud computing centers is ranked using a metric known as PUE or power usage effectiveness. While PUE has been effectively used in the past by companies to maximize their power to output efficiency, it is fundamentally flawed as it does not consider the carbon emissions of the facilities. This means there is little incentive for companies who have access to cheap power sources to search for greener options such as on-site solar or wind farms. Due to this, the problem with cloud computing is not rooted in the methodology of the remote systems themselves, but the means by which these facilities are being powered.

Location is the primary factor in determining where cloud computing facilities will obtain their power supply. Hosting companies tend to seek sites with low power costs that are close to large metropolitan areas (for lower latency). For these reasons, in the US the two most popular options for datacenters are North Carolina and Virginia due to their proximity to New York City and Washington DC [4]. These two states also happen to have outdated and dirty power grids, obtaining only 2-4% of their energy from renewable sources [4].

If datacenters were evenly distributed throughout the US with a similar spread to that of actual companies, the cloud facilities would be the clear environmental winner due to the additional power efficiency measures taken by these companies. The problem with this is the location prioritization by these companies who are seeking cheap power for higher profits and have the freedom to do so because of fast remote connections.

iv. The Current Distribution

There are two primary factors that companies consider when determining the location of a data center: proximity to usersbase and power costs. These two factors are virtually the same for all hosting companies; therefore, the location of these centers tends to be clustered very densely (see Fig. 4). Many of these data-centers are often not data centers at all, but instead old, abandoned company buildings that have been converted and filled with servers. The close proximity of these locations means that they

Table 1: *Data Centers in the World (By Endpoint*)*

Location	Count
US	133,713
HK	63,742
CN	24,300
DE	13,930
ZA	11,437

Table 2: **Words List favored American Companies so distribution is somewhat biased*

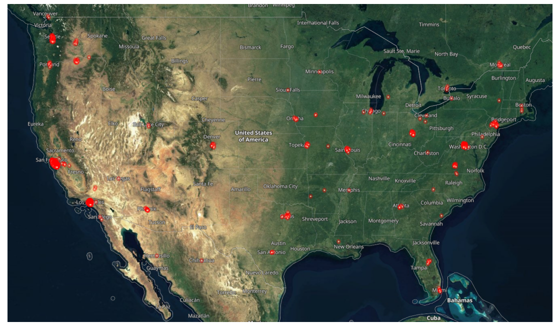

Figure 2: *United States Data-Centers*

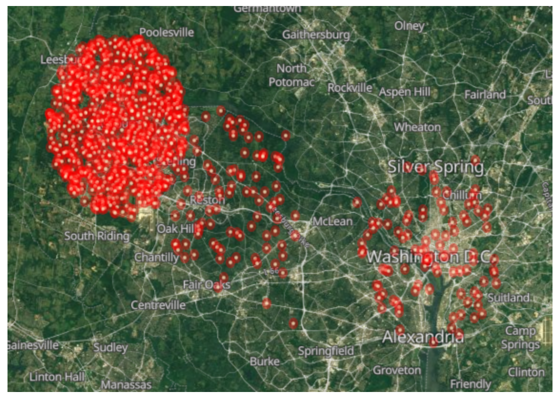

Figure 3: *Virginia Data-Centers*

tend to use the same power grid and since the companies favor cheaper power options, they tend to be in areas with non-renewable grid systems.

While no publically available datasets of the location of every datacenter in the US were found for use in this paper, there are other means of locating these facilities. If you are unfamiliar, in cyber security "port scanning" is when a potential attacker of security specialist will scan a network in order to index all available connected devices. Shodan works similarly to this, but on a massive scale scanning large portions on the internet on a daily basis. This is useful for this study as it can be used to locate datacenters which will contain certain keywords in their domain names such as "aws" or "gcloud" and so on. After complining a words list of these keywords for each companis base domain, the query was fed into the Shodan API which returned a dataset of over 300,000 datacenters, 100,000 of which were in the United States.

For example, Leesburg and Ashburn Virginia house a very large cluster of cloud computing facilities due to the cities proximity to Washington, DC. At the same time, both of these cities are on the Virginia power grid which is overwhelmingly based of on non-renewable energy sources (See Fig. 5).

Figure 2

Figure 3

v. Datacenter Energy Consumption

To determine the environmental effects of datacenters we must first consider how each center generates its power. This is a somewhat complex problem due to the vast number of datacenters and power plants in the United States and must be solved computationally. For this reason, this study involved the creation of a model that pairs these power plants and datacenters using the K-Nearest Neighbors equation with a K-value of 1 to determine the closest possible power plant (which was then assumed to be the power source for the plant). A K - value of 2 was also tested, but the differences were determined to be non-statistically significant. According to this approach, the number one fuel source for datacenters in the United States is natural gas (NG), followed by solar energy (SUN) and distillate fuel oil (DFO) (See Figure). This result is somewhat unsurprising when considering the overall distribution of fuel sources in the United States which follows the same order if overall power output is not a factor in the model.

When compared to the total distribution of power generation sources in the United States it become clear that there is significant incentive for a datacenter to be placed in an area with cheaper power options such as natural gas and distillate fuel options.

Figure 4

Figure 5

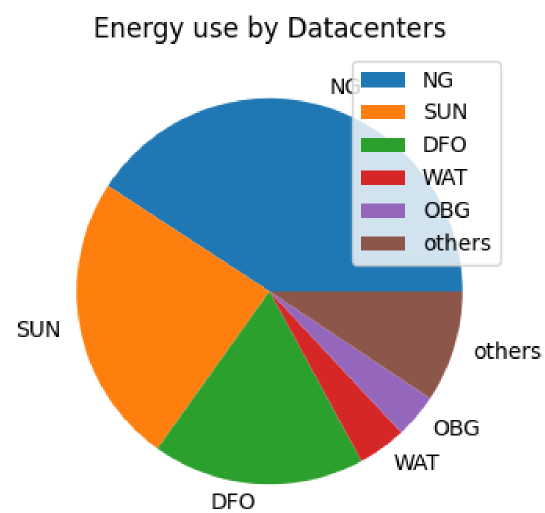

Figure 4: *Energy Source Datacenters.*

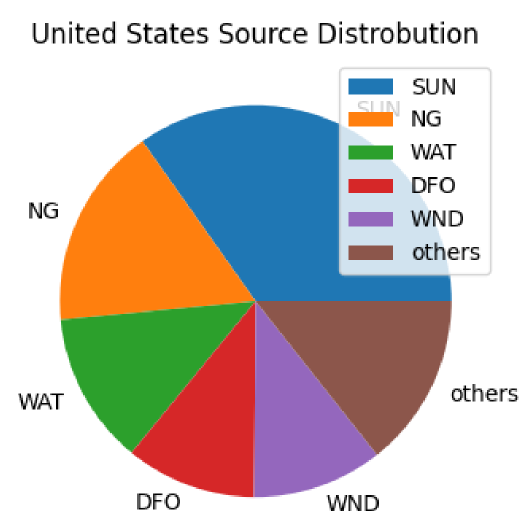

Figure 5: *US Source Distribution.*

vi. Power Cost Comparison

Figure 6

Despite the increasing availability of solar energy in the most technologically developed areas in the United States, the areas where many of these companies are located, datacenters tend to rely on oil and natural has as fuel sources. The reasoning behind this choice becomes clear when comparing the costs of the

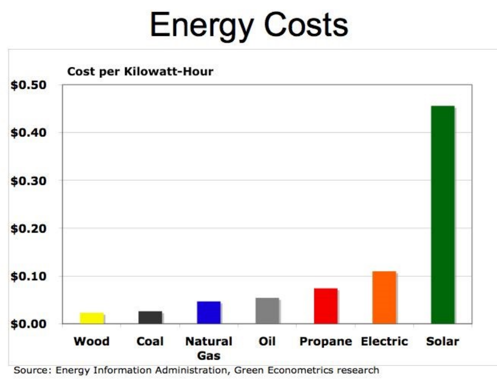

Figure 6: *Energy Cost by Source.*

different energy sources. Solar energy is approximately five to six times as expensive as oil and natural gas when a price analysis is conducted on the entire energy supply chain [11]. While an overall power comparison between company use by actual Watt-hour versus a remote site usage would be beneficial to this study, this is not effective due to the adoption of increasingly low power chipsets and virtualization technologies which make this comparison much too complex [12].

vii. Computational Use Cases

When considering the environmental effects of cloud computing and large server hosting companies it is important to take into account the computational use cases of these systems. Not all tasks are equal, while some positively contribute the to the greater connection of individuals, SMS texting for example, some are much more trivial. For example, in recent years lots of cloud facilities, especially those sold to individuals, have been rented for cryptocurrency mining. This proof-of-work process to generate cryptocurrency has no intrinsic value and only contributes to the wealth of the individual running the machine. The electricity that is essentially wasted in this process is a significant contributor to the total energy use of these facilities. Affording to Goodkind, in 2018 "each 1USD of Bitcoin value created was responsible for 0.49USD in health and climate damages in the US and 0.37USD in China" [10]. It has been reported that the bitcoin mining industry alone could resut in a 2 degree Cesius global warming by the year 2050 if it mining continues at the present rate [15].

viii. Is Cloud Better or Worse?

Based on the analysis and experimental modeling conducted leading up to this point, we can see that economic incentives do in fact lead hosting companies to make environmentally detrimental choices. These choices include the selection of database locations with cheaper power options and the over-development of the hosting sites themselves. This over-development is largely due to an extremely competitive market within the cloud computing sector and the magnitude of the issue is largely unknown due to company privacy. One study took a sample of 333 servers and found that 50% where completely unused on a day to day basic, despite consuming near the same amount of idle power as the in-use server. Over three-quarters of the servers in the study were found to use less than 10% of their computational capacity [14]. In addition, the industry itself is riddled with environmental violations. "Amazon was cited with more than 24 violations over a three-year period in Northern Virginia, including running some of its generators without a basic environmental permit" [14]. So, while cloud solutions do solve a lot of problems through consolidation and replication, they still have some very detrimental issues of their own. It could even be argued that locally hosted systems make companies more conscious of their computational resources which leads them do better optimize their systems based on individual needs.

ix. Centralization is the Key

The greatest strength of the transition to cloud-based computing is centralization. This centralization is attractive to users because it allows companies to know exactly how much resources they are using and budget accordingly. In addition, the same data can be used to determine exactly how much of an environmental effect their usage has on the environment. According to Google, who has done a great deal of research in this area, a single search requires 0.0003 kWh of energy and procures 0.2g of CO_2 [17]. Through this type of precision, environmental chargeback programs could be implemented into cloud systems. While legislative incentives may need to be put into place to force companies to implement these measures, we know that this type of program is feasible. A program such as this would been to be equitable, accurate, and scalable so that the monitoring itself would not be too taxing on the systems. This type of monitoring already occurs in the cloud space for pricing so this solution is easily within reach. Figure 7

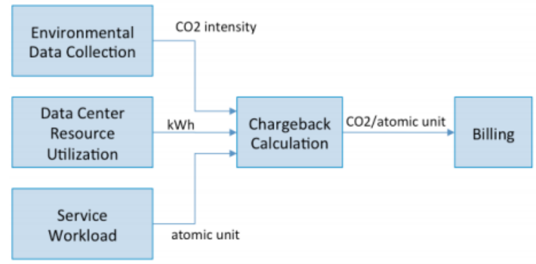

Figure 7: *Environmental Charge-back Program*

VII. The Solution

While the most clear and impactful solution to the environmental issues associated with cloud computing is legislative, there are also some technical solutions available. Due to the adaptability and scalability of these remote systems and recent advances in distributed computing technology, some experts have proposed a renewable energy-based load distribution system that would prioritize remote systems using renewable energy. At the International Conference on Distributed System Platforms and Open Distributed Processing researchers Yanwei Zhang, Yefu Wang, and Xiaorui Wang proposed a system called "GreenWare", a play on the term middleware which is a software technology that handles server requests. In their paper, the research point out that Google hosts 500,000 servers in each of its data centers which are already equipped with long range distribution capabilities. Currently these distribution algorithms are proximity based, but this algorithm could be adapted to consider environmental factors [9]. The research also proved that this distribution system could be made cost effective by including weather as

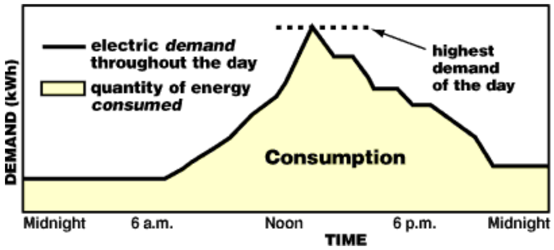

Figure 8: *Power Consumption and Time*

a factor in the algorithm for solar and wind based locations [9].

Figure 8

In addition to implementing a middleware that considers the power source of the host server, load balancing technology can be implemented to manage task runtimes and prioritize times with a lower electrical demand on the power grid. This will save money since many grids offer cheaper rates when overall consumption is lower. In regions of the United States with wholesale energy markets, prices are updated every 15 minutes based on total consumption [12]. Most datacenters are in areas with developed, leading edge power grids, with wholesale options.

Figure 9

Some researchers have proposed a "meta-scheduler" system that would globalize computational loads and give all nations access to clean grid systems. Such a system would rely upon a dynamically calculated COP (power usage efficient) value and require a great deal of cooperation between hosts. The primary argument behind this approach is that it would both minimize carbon emissions while maximizing profits at the expense of less energy efficient hosting companies. For this reason,

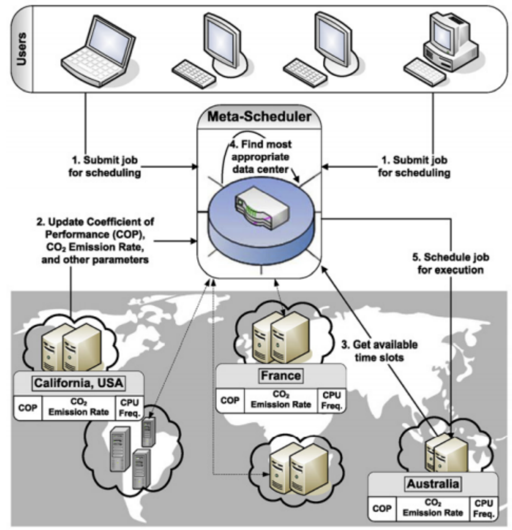

Figure 9: *Meta-Scheduler*

legislative policy would be required in order to generate buy in from less environmentally efficient companies [13].

Lasty, a far more novel and bleeding edge solution has been proposed to phase out cloud computing facilities completely. This methods attempts to solve the computational waste problem of cryptocurrencies where the mining processs holds no intrinsic value. Projects such as CPUcoin offer remote computing options that rely on decentralized cryptocurrency networks where users are paid in cryptucurrency to solve actual, real world problems. The project itself already has over 140,000 users in 80 different countries and is growing rapidly. In addition, per dollar the project provides 120 times more computational power than Amazon Web Services [16].

In summary, the available solutions to environmentally conscious cloud computing are:

1. The implementation of a novel mathemat-

ical model used to make current systems more energy efficient, prioritizing cost effectiveness and low carbon emissions.

2. Optimized task scheduling policies which maximize profit for hosting companies (in hopes that the profit differences will offset the costs of environmental efficacy)

3. The implementation of physical energy efficient technologies within data centers (this is most likely already accomplished due to the profit incentives of such technology)

4. The decentralization of remote computing as a whole to a block-chain based approach

i. Future Considerations

While it has been mathematically proven that scheduling technologies can both reduce costs and carbon emissions, there are a few factors that have yet to be examined. There is a potential for the efficiency of the original meta-scheduling model to be increased by completely turning off certain servers based on usage [13].

References

[1] Loobrok, R. (2019, March 19). The Environmental Impact of Server Hosts. International Conference on Multimedia, pp. 473-480, June 1996.

[2] Pettey, C. (n.d.). Cloud shift impacts all it markets.

[3] Bernheim, L. (2019, November 07). 20+ largest web hosting companies in 2021.

[4] Conley, John. "The Economics of Cloud Computing." ResearchGate, Korean Economic Review, Feb. 2016.

[5] 23, Rich Miller | Mar. "Too Hot for Humans, But Google Servers Keep Humming." Data Center Knowledge, 23 Mar. 2012.

[6] FreeCodeCamp.org. "An Introduction to HTTP: Domain Name System Servers." FreeCodeCamp.org, FreeCodeCamp.org, 21 Aug. 2018.

[7] Bajdor, Paula. "The Environmental Benefits of Cloud Computing." Research Gate, 2016 5th IEEE Conf. on Advanced Logistics Transport (ICALT), June 2016.

[8] "Power Generation in Virginia By Source." Virginia Performs, US Energy Information Administration State Energy Data System (EDS), Feb. 2014.

[9] Yefu, Yanwei, et al. "GreenWare: Greening Cloud-Scale Data Centers to Maximize the Use of Renewable Energy." Springer Link, ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing, 2013.

[10] A. L. Goodkind, B. A. Jones, and R. P. Berrens, "Cryptodamages: Monetary value estimates of the air pollution and human health impacts of cryptocurrency mining," Energy Res. Social Sci., vol. 59, Jan. 2020.

[11] Pfeffer, Kathlyn. Derwi, 28 Oct. 2019, www.derwi.tk/9109/comparison-chart-cost-per-kilowatt-hour-of-different-energy-sources.html.

[12] "Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a Multi-Electricity-Market Environment." IEEE Xplore.

[13] Garg, Saurabh Kumar, et al. "Environment-Conscious Scheduling of HPC Applications on Distributed Cloud-Oriented Data Centers." ScienceDirect, Journal of Parallel and Distributed Computing, Jan. 2011.

[14] Glanz, James. "Power, Pollution and the Internet." The New York Times, The New York Times, 23 Sept. 2012.

[15] "Uncovering the Environmental Impact of Cloud Computing." Earth.Org - Past | Present | Future, 12 Oct. 2020.

[16] Barger, Sean, et al. "CPUcoin Whitepaper." CPUcoin, CPUcoin, Jan. 2021.

[17] Curry, Edward, et al. "An Environmental Chargeback for Data Center and Cloud Computing Consumers." Edward Curry, Digital Enterprise Research Institute (DERI).