An improved model for the allocation of carbon emissions among the tenants of cloud services

Master Thesis

Author
Albert Dijkstra
a.dijkstra.38@student.rug.nl

Date
July 24, 2023

First Supervisor:
Vasilios Andrikopoulos

Second Supervisor:
Daniel Feitosa

External Supervisors:
Richard Atherton
Greg Hambleton

Abstract

A lot of current software services use technologies facilitated by cloud computing. Cloud computing produces an increasing amount of emissions. Estimating the carbon footprint of cloud computing can be done by various models. However, it remains unclear which stakeholders bear responsibility for which part of the emissions, leaving the question of accountability unanswered.

In cooperation with BT Global Services, this project continued the development of a model to estimate the carbon emissions of cloud-based software services, and distribute them fairly among stakeholders. The estimate is made by splitting energy consumption into static energy usage, which cloud resources consume while idling, and dynamic energy usage, which is based on the actual usage pattern. Based on this, a lower and upper bound, together with a set of policies for deciding these bounds is defined with respect to which part of the energy consumption the stakeholder is responsible for. This provides the involved stakeholders with more insight on their footprint. On top of this, a new metric which clarifies how energy efficient a tenant uses a service in relation to other tenants is proposed. The use of this metric was evaluated by different stakeholders with respect to its efficacy.

Contents

Chapter 1

Introduction

Over the past few decades, there has been a significant increase in the use of IT equipment. Even though IT enables often-used technologies, it has a substantial environmental impact. This is also reflected in the electricity usage. Various studies have made different estimations of electricity usage. In 2015 it was predicted that data centers will use 3%-13% of global electricity and IT-related equipment between 8% and 51%. That while the global electricity usage of data centers was just 1% in 2010 [1]. Another study in 2021 estimated that data centers would use 2.13% of global energy in 2030. This study also took into account IoT and the end of Moore's law [2].

These variations in reported figures regarding IT energy consumption do not undermine the credibility of these estimates, rather, they reflect the different factors considered and the scope of the studies. Moreover, rapid change in technology[1] has a positive effect on energy usage [3]. Nevertheless, all these studies show the same trend, the energy usage of IT increases each year. This can be explained by the rebound effect [4], which states that when a unit of production can be produced with less amount of units than before, the increased demand can result in countering the potential savings. The more energy-efficient production makes the product of a unit cheaper and increases the demand.

While data centers do contribute to emissions, it would be simplistic to state that data centers only attribute to the amount of emissions emitted. For example, electronic software distribution instead of physical software distribution reduces emissions up to 83% [5]. Also, in other fields where IT supports their operations, they emit up to 7.2 less emissions compared to scenarios without IT implementation [6].

The carbon footprint as defined by Pertsova and widely adopted in the literature is as follows: *"The carbon footprint is a measure of the exclusive total amount of carbon dioxide emissions that are directly and indirectly caused by an activity or is accumulated of the life stages of a product."* [7] However, the carbon footprint is just a part of a larger family, including ecological, energy, carbon and water footprints, called the footprint family [8]. The energy footprint can be translated to the carbon footprint [8]. Moreover, the carbon footprint is more widely adapted. For the scope of this study everything can therefore be combined in the carbon footprint.

Following these premises the chapter is structured as follows: the GHG protocol as the foundation of this work will be explained in Section 1.1. With this information, the case can be summarized in Section 1.2. Then the problem definition to explain the scope of this thesis will be explained in Section 1.3 and Section 1.4 will give the structure of the thesis.

1.1 GHG protocol

Currently, the greenhouse gasses (GHG) protocol is widely recognized and utilized as the standard framework for carbon emissions reporting. It classifies emissions into three scopes, each representing different sources of carbon emissions. The GHG protocol is defined by the World Resources Institute in 2004 [9] and later standardized by ISO [10].

[1] For example virtualization and improvements in energy efficiency at data center level

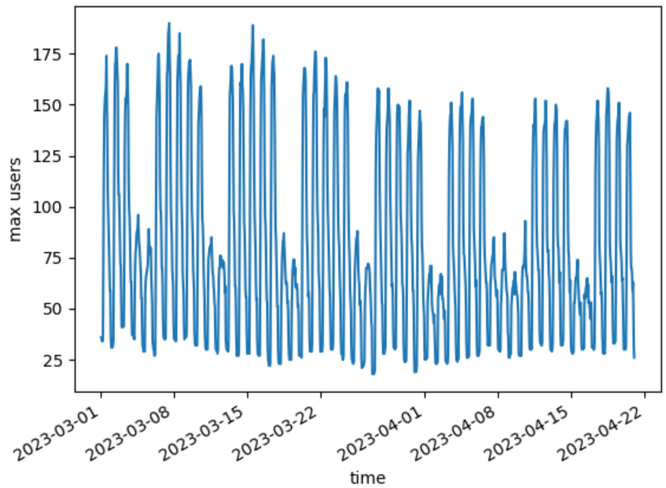


Figure 1.1: Max users per hour

The three scopes are as follows:

Scope 1 These are the direct emissions, emitted by the company itself, such as those emitted by on-site machines and power generators. While data centers typically do not have these emissions, it is worth noting that data centers often have power generators on-site in the event of a blackout.

Scope 2 These are the indirect emissions caused by the consumption of electricity. This is the main focus of this research since devices such as servers, network systems, cooling systems and other such systems consume electricity. Sotos [11] created a guideline for reporting Scope 2 emissions and stated that it is should contain energy consumption in any form.

Scope 3 All emissions that do not fall in Scope 1 and 2 are classified as Scope 3 emissions. Depending on the sector Scope 3 emissions account for either a significant share or most of the emissions [11, 12]. Bhatia et al. [13] created a guide for Scope 3 emission, categorizing it into upstream and downstream emissions. Upstream emissions refer to the emissions based on purchasing a product. If the product had not been bought the entire emissions associated with this product would not have occurred. The demand enables the production and thus the emissions. Downstream emissions are the emissions based on a product being used or consumed. If that product had not been sold, it would not be able to emit emissions. When the product is sold, it enables the emissions made by the customers [14]. For a business, this was divined as upstream emissions are the emissions caused by the creation up to that point. The downstream emissions are enabled by this product. A more detailed explanation can be found in [13] and [14].

1.2 Case Summary

BT Global Services offers cloud services to their clients, including the Cloud Contact Cisco (CCC), which falls under the Software as a Service (SaaS) model. CCC is a customer contact software used by the customer service departments of BT clients. The users of this product are the employees of BT's clients working in call centers. BT keeps track of the maximum amount of users logged in on a half-hour basis, referred to as concurrent users. An example of the number of concurrent users during a period of slightly less than two months can be seen in Figure 1.1.

CCC is hosted on off-premise private cloud deployments worldwide. Different tenants use CCC differently, resulting in different levels of efficiency. BT aims to provide insights into this usage, following the example of data center providers such as Amazon, Microsoft and Google [15–17]. These companies provide users with a dashboard with detailed information on their carbon footprint. These carbon emission estimations of the service are based on the GHG protocol.

There is a significant difference between the BT use case and the service provided by these three companies. BT provides the user with dedicated hardware, whereas the other three provide services based on public cloud infrastructure. The public cloud can be more efficient than the private cloud due to resource sharing, which reduces the amount of emissions through economies of scale. This is similar to taking the bus to produce fewer emissions instead of taking the car to work. Furthermore, it can scale up during peak hours, which makes it more dynamic than having one computer that needs to handle all the peaks [18].

Amazon claims up to 80% reduction of emissions and Microsoft up to 93% [15, 16]. An energy comparison in 2011 supported, in general, public cloud is more energy efficient. However, when moving high quantities of data, it might be more energy efficient to use your own devices. This suggests public cloud is not always more effective [19], but for each use case, one must pick the best suitable solution. Moreover, it is not always to right financial move to make. When using the hardware for 2-3 years, private cloud is cheaper [20, 21]. This is under the assumption that there is at least 70% utilization.

Amazon, Google and Microsoft all three developed a dashboard for customers to track their emissions. BT noticed that their clients want a similar dashboard. This was created by a third party based on the work of Westerhof [22]. However, for this first approach the accountability was left fully to the tenant, something this model will improve upon. This model will be explained in Section 4.1. For this research, this model will be improved to give a more accurate estimation for private cloud emissions and help with allocating them among multiple parties, that is, when tenants and data centers, are involved.

This will also lay the groundwork for tenants to compare their energy efficiency with each other.

1.3 Problem Definition

In 2021 Westerhof [22] introduced and developed a methodology to estimate the total carbon footprint (TCFP) of servers. Continuing his research in 2022 [23], he further improved his model for the private cloud. This model had some limitations, which will be addressed in this thesis. More about this model in Chapter 4. One of the main questions from the tenants in Westerhof's research was how efficient tenants are in comparison with others. This will also be addressed in this research.

This leads to the following research questions:

Main research question *How can different tenants of the same software service hosted in the cloud with different usage profiles be compared to each other in terms of carbon emissions?*
Different tenants have different amounts of concurrent users. Just using the total carbon footprint would not give an accurate depiction since the tenants work on different scales. To answer this question multiple sub-questions have to be answered first.

Research question 1 *How to fairly allocate the emissions among different stakeholders?*
In Westerhof's model, all emissions were allocated to the tenants. It is more accurate to give the different stakeholders partial responsibility. But how can this be done fairly? I.e. all stakeholders can take responsibility for the part they have control over.

Research question 2 *When is the software services more efficient with respect to the TCFP considering the number of concurrent users?*
Parts of the infrastructure are shared between the users. Does this mean it will decrease the total carbon footprint per user?

Research question 3 *How much difference does the location of the data centers make?*
These factors can have a huge impact whether for the carbon intensity and the Power Usage Effectiveness (PUE). Where the PUE is the total power used by the data center divided by the amount of usage by the IT equipment.

1.4 Thesis Structure

In this first chapter, an introduction is given on the topic. Chapter 2 delves deeper into the subject by summarizing the background and related work. Moving forward, Chapter 3 presents the case study with BT Global Services, in which the problem statement, research methods and functional and non-functional requirements are defined. Additionally, this chapter explains the data collection procedure and how the model will be evaluated. In Chapter 4 Westerhof's model will be summarized, focusing on the improvements made to adapt it to the available data and explaining the utilization of this refined model. The details for the model implementation, the technologies used and the output of the implemented model are described in Chapter 5. In Chapter 6 the model is applied to six tenants of the CCC service, and the resulting findings are presented. Chapter 7 introduces a method to compare different tenants with each other. This method is then evaluated through a survey filled in by employees of BT and the result and implications of this are then discussed. In Chapter 8 the thesis is summarized and the impact is discussed. The chapter will end with future work.

Chapter 2

Background & Related Work

The start of the Renewable Energy Directive (RED) in 2009 [24] created an incentive for research in green energy. This is part of a broader picture where data centers are becoming more efficient and the reduction of energy consumption is gaining importance. In this chapter, we will provide a wide overview of this research. In Section 2.1, we will present some important definitions. The standardization efforts will be summarised in Section 2.2. Life Cycle Analysis, which considers the entire life of a product, will be explained and summarised in Section 2.3, as it closely relates to the problem defined in this thesis. Section 2.4 explores energy-efficient techniques to minimize power consumption. In Section 2.5, we will explain the different ways of estimating the carbon footprint.

2.1 Definitions

There are three important topics used throughout the literature: PUE, private and public cloud and X as a Service.

PUE An often-used metric in the context of data centers is the Power Usage Effectiveness (PUE) and its inverse, data center efficiency (DCE). PUE is the total power used by the data center divided by the amount of usage by the IT equipment. The closer to 1.0, the more efficient a data center is [25]. It is defined as in Equation 2.1.

$$PUE = \frac{TotalFacilityPower}{ITequipmentpower} \tag{2.1}$$

The DCE is defined as in Equation 2.2.

$$DCE = \frac{1}{PUE} = \frac{ITequipmentpower}{TotalFacilityPower} \tag{2.2}$$

PUE can be used as a way to evaluate the efficiency of a data center. Over the past few years, the PUE of data centers has been decreasing. It is important to note that data centers are doing this voluntarily [26].

Despite the wide adoption of PUE, there has been criticism [27]. When all the servers are idling the PUE is better than if idling servers are turned off. Thus, the use of this metric should be done with care.

Private and public cloud In this research, it is important to make a clear definition of private and public cloud. NIST defines private cloud as something used exclusively for a single organization [28].

Private cloud can be both off-premise, also called outsourced, and on-premise. On-premise private cloud applies to all private clouds implemented at the premise of the customer, while outsourced private cloud where the hosting is outsourced to a hosting company. Both solutions have their positive and negative sides, which depending on the use case [29].

Public cloud on the other hand is used by the general public [28].

In [28], they also define the different service models.

X as a service Software as a Service (SaaS) is where the software is provided as a service. The client does not manage anything, but simply uses the software. This is the focus of this thesis.

Platform as a Service (PaaS) is where a platform is provided as a service. The client gets access to a platform to deploy their applications, but does not manage or control the underlying cloud infrastructure.

Infrastructure as a Service (IaaS) is where the infrastructure is a service and the client has control over the operation system and applications, but has no control over the underlying cloud infrastructure. [28]

Thus, SaaS has the highest level of abstraction, while IaaS has the lowest.

2.2 Standardization efforts

Multiple standards have been proposed for the reporting of carbon footprint and its related emissions. With the exception of the software carbon intensity (SCI), they all build on the GHG protocol. SCI calculates the rates of carbon emissions for a software system. SCI should help developers make informed choices about which tools, approaches, architectures and services they use in the future [30]. This is however a too detailed method for this use case as it tries to estimate the carbon intensity per part of the software, for example an API call. Moreover, the authors of the SCI whitepaper did not stated how to determine these values. They do propose to normalize the data per certain unit, for example users. SCI's goal is to eventually lead to a greater focus on sustainability. As part of the European Green Deal, large corporations have to disclose, among others, their sustainability [31], later more companies were also required to do the same [32]. "This helps investors, civil society organizations, consumers and other stakeholders to evaluate the sustainability performance of companies, as part of the European green deal." [32]

Extensions for the GHG protocol

Some extensions have been made to the GHG protocol. Two similar extensions are the Clean Energy Emission Reduction (CLEER) protocol [33] and the Science Based Targets initiative (SBTi) [34]. The primary goal of both is to reduce the emissions of businesses. CLEER gives insight into how much emission a business can reduce by providing clean energy actions. SBTi helps corporations reduce their emissions by settings targets.

The current usage of the GHG protocol has a problem in reporting. When IT is outsourced to the cloud, their respective Scope 1 and 2 emissions move to Scope 3 emissions. As scope 3 emissions are voluntary to report, they can therefore be hidden [35]. This poses a challenge in obtaining a comprehensive view of the total emissions.

A way to make Scope 3 more insightful, is to split Scope 3 into Scope 3 and 4. Where Scope 3 contains the entire supply chain and Scope 4 includes delivery, use and end of life [36]. Essentially splitting the upstream and downstream emissions into different scopes. This split would give a more detailed report, but this proposal has not seen a wide adoption yet.

Armstrong et al. [37] extended the GHG protocol to help data centers identify their emissions. Multiple factors are identified that are important for the final calculation. Such as fixed and variable emissions[1] and the allocation steps for the layers, from data centers to users. They include all the possible emissions in their analysis to give an in-depth estimate. They identify two different methods for obtaining these values. Top-down and bottom-up. In the Top-down approach, the total emissions of the entire data center are calculated first and then allocated to individual components. Bottom-up is the other way around, as it calculates the individual parts and then combines them.

2.3 Life-Cycle Analysis

Life-Cycle Analysis (LCA) is the analysis of the environmental impact of a product, or in other words, from 'cradle to grave' [38]. ISO defines four stages for LCA [39].

1. *Goal and scope definition* determines the level of detail, depth and breath & methods used.

2. *Life cycle inventory analysis* (LCI) is the inventory of the material or/and energy used for the product.

[1] More about this later

3. *Life cycle impact assessment* translates the LCI results to the environmental impact.
4. *Interpretation phase* is where the results are reported and a conclusion is drawn. Based on this also recommendations can be made.

2.3.1 Methods for Life-Cycle Inventory Analysis

An integral part of LCA is LCI. For that reason and extended research has been done in this area. LCI can be done in many different ways. There can be three main categories [40] defined. Suh and Huppes define an extra, often-overlooked category. Each of these has positive and negative sides. To categorize the different works, we looked at the categorization of both [40] and [41]. The following categories can be made:

Process analysis Process analysis or bottom-up approach first defines the entire supply chain. This is the most commonly used model [41]. This can be used to determine the footprint of the product for the consumer.

This can be calculated in two different ways, which, if done correctly, gives the same result. Most often used is the infinite geometric progression which is also used by Lenzen et al. [42]. The other way is the most often overlooked method, with the use of a matrix representation [41].

Environmentally extended input-output analysis Environmentally extended input-output analysis (EEIOA) is a methodology based on that all processes are interlinked. With the use of process analysis this is not taken into account and gives truncated data. It is based on the Leontief model which was originally meant for financial data and money flows [43]. Emissions can be treated similarly to monetary flows, which makes it possible to adapt the model. This model works under the assumption that each industry consumes and produces its output for multiple industries. This is modeled with the use of a matrix.

Hybrid analysis Hybrid analysis is a combination of the two above. In the past two decades, a lot of research has been done in this field. However, one challenge is the lack of clear and precise terminology, which can undermine practitioners' confidence [40]. [40] identified four different hybrid methods and [41] identified three different hybrid methods. Each of these hybrid models falls somewhere on a spectrum between process analysis and EEIOA. More so, there is a debate regarding the superiority of hybrid methods over process-based analysis. Yang et al. claim that hybrid does not necessarily produce better results than process-based [44]. This was disputed by Pomponi and Lenzen as they disagree with the scenario used. They argue a more realistic scenario would favor hybrid LCA [45].

2.3.2 Sharing Responsibility

Double counting refers to a situation where emissions are counted twice. For example, there exists Company A which both consumes and produces for Company B. In the supply chain for Company A it has to count itself, thus resulting in a double counting problem. This is a significant problem if multiple participants want to be held responsible within the supply chain [42, 46]. Process analysis is often used to hold different stakeholders accountable. However, hybrid LCI can be improved to address the double-counting problem [47]. Caro et al. [48] argue that when a company aims to offset its emissions, it should not allow for double-counting. However, when providing incentives to suppliers, double counting can be allowed. When for firm A the greatest potential for emission reduction lies within Firm B, firm A should focus its efforts on this point. This strategy can only work if double-counting is allowed in a pro-forma fashion, creating opportunities for joint improvement.

Apart from this study, most studies try to estimate emissions without double counting. Apart from the emissions emitted by its own process, it can also be argued that both the producer and consumer are responsible for the emissions. Without consumer demand, producers would not create products, and without producers, consumers would not have products to use [42]. This line of thought leads to the concept of upstream and downstream emissions as is done with the Scope 3 emissions as explained in Section 1.1. However, LCA generally assumes consumer-based responsibility [49]. This can give a skewed view. An example would be the global trade which consists of approximately 18% of the world's emissions. Norway, which is a very emission-low country as a consumer, exports among other huge quantities of oil. As a producer, they are therefore still responsible for a significant amount of emissions.

This can also be expanded in a production line with multiple producers and consumers. The product or service at the end causes joint carbon production. Two different models were proposed to deal with this challenge. The first model makes use of a social planner. The social planner is a third party that allocates the emissions a company in this production line is responsible for. The second model makes use of a carbon leader, which takes full responsibility for the emissions. The other companies then pay a certain price to the carbon leader [48].

2.3.3 Data centers

LCA can also be used to determine the emission from a data center. A guideline was developed to determine the LCA for a data center by Aggar et al [50]. For this, they identified the boundaries of the data center to report, as well as the average lifetime for different equipment. Moreover, they also had to include other significant contributions to the LCA such as the geographical area. Lastly, they also defined the different LCA stages of a data center. By combining all these factors, it becomes possible to assess the environmental impact of a data center [50]. Whitehead et al. [51] performed a life-cycle assessment of a UK data center. They used a hybrid model to accurately depict the emissions. However, they did have to use proxy data and advocated for more reliable data.

2.4 Carbon footprint minimization techniques

Research shows that on average somewhere around 10% to 50% [52–54] of servers are underutilized. Idle servers consume up to 50% of peak power consumption [53, 55, 56], in some cases, it is even estimated to be as high as 95% [57]. In 2015 Google reported that they have an average idleness of 30% [54]. Due to the significant impact of server idleness, extensive research is done to optimize this. The techniques can be split into different categories. One way to categorize it broadly is to split the techniques into hardware and software techniques [58]. A more detailed way of categorizing is by using eight different categories: software level, hardware level, non-technical, miscellaneous techniques, thermal management and cooling, bio-inspired, power-aware management and VM consolidation [59]. Which gives a more in-depth view of the different options.

In the following section, we will go over some important contributions. A more extensive survey on the topic was done in 2018 by Gill and Buyya [60].

2.4.1 Virtual Machine Placement

A significant distinction between public and private cloud is the utilization of hardware. A big part of this is virtual machine placement. Public cloud optimizes hardware utilization in such a way that it improves efficiency. [2] In the following paragraphs different ways to achieve efficient utilization will be explained.

Temperature The cooling of equipment is a significant contribution to the energy usage of data centers. At lower temperatures, the cooling equipment does not have to work as hard. The temperature distribution in a data center can be optimized in such a way to cause a reduction in the energy consumption. This can be done by distributing the workload. An example of this is an optimization of the weekend effect. The weekend effect is when everybody submits jobs on Friday and expects them to be finished by Monday. On a first come first served basis everything will be done as quickly as possible. This could also be distributed over the entire weekend. Such optimizations can achieve a reduction of energy usage up to 60% [61].

Utilisation The proper utilization of the host is fundamental to the reduction of energy usage. As the idle power consumption is a significant part of under-utilized servers. This can be improved by having multiple energy states in computers [62, 63] and staggering periodic scheduled activity [63].

[2] This is one of the main reasons public cloud uses less energy

The effectiveness of placement techniques, such as virtual machine scaling, to decrease energy consumption highly depends on the utilization level of the host [57]. The placement of virtual machines has similarities with the bin-packing problem, where items of different sizes need to be fitted in the least amount of bins. Task consolidation is one approach that can increase resource utilization and therefore reduce the energy consumption by using two parameters to determine the best fit. Lee et al. [53] proposed two methods using this technique of which ECTC performed best with a reduction of 18%. This method incorporated static and dynamic power usage to minimize the energy usage of one task. Moreover, a comparison of four different bin-packing approaches for VM placement was done in 2021. Of these four EUBFD[3] algorithm worked best. EUBFD lists the VMs in ascending order of computing and servers in ascending of power consumption. This is then matched in this order.

Another approach to reduce energy consumption is to use live migration based on the actual utilization to place the different VMs on servers. However, doing this incorrectly can lead to performance degradation. To address this the Server Level Agreements (SLA) have to be incorporated. SLA is an agreement between the provider and the client about the conditions of the server. Violations of SLA should be kept at a minimum. A model can be developed based on this to migrate virtual machines to keep hosts from under- and over-utilization [52]. Two different heuristics can be used to minimize energy consumption. In Max-Utilisation as many servers as possible are switched off by migrating VM's away from underutilized hosts. Min-Utilization, on the other hand, tries to minimize the utilization of individual hosts. With the use of both techniques, the emissions can be reduced by up to 94%. [64].

A specific method for optimizing utilization is by reusing servers. Reusing servers is slightly different than just optimizing utilization since they try to use the 'leftovers'. Heracles was developed based on this principle and is used for scheduling batch jobs during periods when services, such as web search, are under-utilized. This way server utilization could be increased to an average of 90% [54].

2.4.2 Computing with green energy

There are multiple definitions of computing with green energy. In the context of this thesis, it refers to the use of green energy to power data centers and servers. With this definition in mind, we examined major contributions to the field and summarized them.

There are two different ways to utilize green energy. One is to use the stranded power, or in other words the leftover energy. The other is to compute in places with the best carbon intensity.

Follow the renewable 'Follow the renewable' is a concept created for data centers to build closer toward renewable sources [65]. Jobs will be scheduled in such a way that maximizes the use of green energy while minimizing the use of brown energy, referring to polluting energy sources. This reduces emissions while it can still have an average response time [65].

Khosravi et al. [55] proposed a method for VM placement that aimed to minimize the extra power consumption. This way a job would emit the lowest amount of carbon emissions. This method was later improved to include green energy. They introduced a cost function for different energy sources. On-site green energy sources had a cost of zero and brown energy had a higher cost. This made jobs follow the green energy resulting in an emission reduction of 60% [56].

Stranded power Chien et al. [66] proposed Zero Carbon Cloud (ZCCloud) which uses stranded power, that is power generated by green sources which cannot be used since they are generating more power than is currently requested. ZCCLoud proposes the establishment of data centers at the source of the green power, resulting in zero carbon emissions from energy sources. While volatile resources may pose challenges for batch jobs this could be an interesting use case. A further study in ZCCloud indicated it reduces cost in comparison with traditional servers [67]. ZCCloud was improved upon to also use the 'follow the renewable' principle, which was called CAISCO. At the moment a data center is run totally on green energy jobs are migrated to that data center. If there is leftover green energy a ZZCloud site is started and jobs are migrated to this site [68].

[3] Energy utilised Best Fit Decreasing

2.4.3 Heat reuse

Another way to reduce emissions is to reuse the heat of data centers. Even though it is not standard practice, multiple data centers are reusing heat [69]. A borefield can be placed between the data center and the houses in areas around the data center. Here, the heat from data centers is supplemented to become quality heat [69]. In optimal conditions, such as when houses are heated with this heat, it can reduce up to 84% of the emissions. However, the best way from an economic standpoint may differ from the one that reduces the greatest amount of emissions.

There are several challenges when reusing heat [70, 71]. The heat is low quality and has to be supplemented, moreover it has a high investment of cost. Additionally, reaching an agreement that benefits both the provider and the distributor of the heat is crucial. When this can be done it is a good investment for the data center. The environmental impact depends on multiple factors, such as utilization and which fuel it replaces. A case study in London investigated the reuse of heat and estimated that it can save up to 4000 tonnes of CO2e and nearly 1 million pounds per year [72].

2.4.4 Green certificates

It is common for data centers to buy Renewable Energy Certificates (REC) so that their annual input of green energy surpasses their energy needs. While this does not decrease the amount of emissions made by the service or data center, this supports generating green energy. This results that even if a data center state they are a 'green data center', this might not be true. On an hourly basis, their need for energy can surpass the amount of green energy that can be delivered. Due to this, a real net-zero emission data center is hard to make [73].

2.4.5 Other approaches

There are also other approaches to reducing power consumption. One of them is to choose hardware for a specific job. With this, a 'Cornucopia Corollary' can be added to Amdahl's law, namely "Low utilization of a huge, cheap resource can still deliver high, cost-effective performance" [74]. A more generalized approach to this principle can be done with the use of LegoOS. LegoOS is a distributed OS based on disaggregated, network-attached hardware components instead of traditional monolithic design. This can be used as building blocks to only use the parts for the given task. Since the hardware is now more catered for the task it will reduce the overhead of unused hardware parts, reducing the energy consumption [75].

Another approach is edge computing, that is computing closer to the source of the data. This reduces the data sent through the network. This could lead to a reduction of 50% of emissions. Apart from this, other techniques such as power capping, more efficient network protocol and hotspot and coldspot migration can be used to reduce emissions [76].

2.5 Energy and Estimation Approaches

This section will focus on various ways of estimating carbon emissions in data centers. In order to estimate the emissions, energy usage has to be estimated first. The energy usage can then be multiplied by the carbon intensity, that is the amount of carbon emitted per unit of power usage during a given hour.

2.5.1 Energy usage

Energy usage is one of the main expenses of a data center and the main source of emissions during the operational phase.

The energy usage estimations can be categorized in different ways. One distinction is between models based on OS-provided input, such as utilization, and hardware-provided metrics [77]. A way to estimate the energy consumption based on OS-provided input is to fit the energy data for a specific computer linearly on different utilization metrics [78, 79]. Vasan et al. [63] used a model based on CPU utilization on a wide range of data servers and achieved a mean error of less than 6% in the predicated power. While at first the estimations based on CPU dominated the literature, memory and I/O is also an important factor in energy consumption [80]. A model was proposed based on I/O-intensiveness. They define it as I/O- volume in MB divided by the time the task spends performing solely computation. The model coefficients are specific for a machine, requiring benchmarking on a machine before using the model.

Hardware-provided metrics use either performance monitoring events (PME) or performance monitoring counts (PMC), such as accessing cache. One challenge with PMC is the architectural knowledge needed [77]. This was partially solved by relative PMC, which reduced the architectural knowledge by simplifying the number of parameters needed. Despite simplifying the error was still below 5% [81]. For example, PME can be used to determine the power consumption for a specific CPU [82]. There are also some models which use a combination of PME and PMC, which are called hybrid models.

Another categorization is by dividing the power consumption into the following categories. The first category is additive models, where each component contributes to total power usage. The second is baseline power + active power (BA) models, which assumes the idle state of a server already consumes a significant amount. Then the active power is added based on the utilization of certain resources. Lastly, there are also other models, which are all those that do not fall into the above categories.

BA models can be divided into four different models, linear regression, power function, non-linear and polynomial. Among these models, the polynomial model was the most accurate with an error of 1.6%. It is important to note that these models only perform well if the server has a workload that is CPU-heavy [83].

BA models can also be used in a distributed server. Lin et al. proposed a distributed energy measurement (DEM) tool that determines the energy usage of a distributed server by using a master-slave configuration, where each slave uses a BA model [84].

For additive models, the separate parts first need to be calculated. One model to use this method is Cloud Jewels. They roughly attribute power to different components; compute, memory storage and networking. These components were all multiplied by calculated constants to estimate the total KWh [85].

Network

A network consists of various different nodes which communicate with each other. When sending data from A to B, the network can select different paths. There is a significant difference in the energy usage of different paths. The shortest, which is assumed to be the fastest, is not always also the most energy efficient [86].

Apart from the path also the kind of transportation makes a difference. Networks can consist of different kinds of technologies between nodes. The most energy-efficient transportation is the use of optical networks [87]. Despite the power consumption of network cables, the biggest part of the consumption is switches and routers, whereas the cables only account for 1% of the total power usage [86, 87].

Multiple models have been proposed for estimating the power consumption of network devices. The model by Mhadevan et al. proposed that the power consumption of a network depends on many different parameters, such as the number of active ports, the line speed and various other factors [88]. Another model estimates the power consumption based on only three parameters, the idle energy, the energy used for packet processing and the energy used for storing and forwarding each byte of the payload [89]. While the first model stated that the power consumption depends on port utilization, the second stated that the power consumption is linear with the utilization [88, 89]. A technical document of the network devices used by BT, however, showed that at 10% utilization it already consumes most of the max power consumption.

In contrast to previous models that estimate the power consumption of a single device, Aslan et al. estimated the consumption of the entire network. They identified four methods to derive the energy consumption of a network: modeling, annual electricity consumption, direct measurement and extrapolation. Moreover, they also identified that previous studies used different scopes and time to estimate it. This contrast between studies leads to no major variation between estimates. Also taken into account are two different kinds of energy. The first is the energy used to transmit data, which according to Aslan et al. is the full responsibility of the one using it. The second is the energy needed to sustain underutilized networks, which needs to be divided into those requiring peak data capacity. To provide an up-to-date estimate they looked at the years the estimates were made, fitted it with a linear model and observed the fact that every two years the amount of energy was halved. Combining all these factors gives an average of 0.06 Kilowatt-hours per Gigabyte in 2018. This does not include the routes and switches inside the client's homes [90].

Virtual Machines

Servers are often shared between multiple virtual machines, which are all from different tenants. This makes it important to calculate the energy consumption for one virtual machine. Bohra and Chaudhary [91] proposed Vmeter, which is a model to estimate the power consumption of virtual machines. They took into account CPU, cache, DRAM and disk usage. Here, CPU and cache were dependent on each other, as well as DRAM and disk. These measurements are estimated using a linear BA approach. These workloads have to be calculated manually for each individual application. Later it was improved to calculate these weights automatically [92].

Kansal et al proposed Joulemeter, a similar approach [93]. A combination of CPU, memory and disk is used to determine the energy usage. They use an additive model, where each of these components uses PMC to a certain extent to give an estimation. The model has to be trained before giving an accurate estimation. In their study, they used Joulmeter to estimate the server run in a Windows server hyper-V environment, but it will work for other hypervisors as well.

Both of these models have to be trained beforehand with specific parameters. However, in cases where limited or no data is available for training, a black box approach has been proposed. Such approaches estimate the power consumption by estimating a lower and upper bound on low and high utilization of CPU and memory [94].

The energy consumption can also be estimated by first defining static and dynamic power consumption. Static power is the idle power of the machine divided by the number of VMs. The dynamic power is based on a Look Up Table (LUT) of CPU utilization and cache misses. When a VM first enters a host it records the data needed for the LUT [95].

Cooling

The energy consumption by cooling devices is a significant component of the overall energy usage. Cooling techniques can be categorized into three main types: air-cooling, liquid-cooling and free-cooling. The power consumption of cooling system differs widely per data center. A combination of these categories is used in a data center and should be incorporated in a model. Two approaches used to estimate power consumption are mechanism-based and data-driven methods. A mechanism-based method, which estimates the power consumption explicitly based on a model. The data-driven method creates a black-box or grey-box model instead. The latter is more suitable for complex processes. Both models consider the idle part as an important factor, but with the data-driven method, this factor is less easy to predict [96].

2.5.2 Carbon Estimates

The carbon emissions can be calculated for cloud solutions at different levels of abstraction with various online tools. One such tool is Climatiq, which developed a model to estimate the emissions emitted by different cloud solutions. For the public cloud, they used AWS services, but it can also incorporate private emissions and be used for LCA calculations [97]. SDIAlliance created a model called carbon footprint SSA [98]. This model calculates the footprint of Virtual Machines based on power consumption, resource utilization and VM schedule. This data is then used to calculate the amount of energy one VM used. This is then used to calculate the footprint.

A tool from ThoughtWorks shows the energy usage and carbon emissions of the different public cloud providers; AWS, Google Cloud and Microsoft Azure. This tool both includes the operational and embodied emissions. Their operational emissions are estimated based on the Etsy cloud jewel [99].

Jetraw has developed a tool to estimate emissions of public & private cloud [100]. While they take a lot of different parameters into consideration, they are not transparent about how the estimates are calculated. Moreover, whatever option is chosen, their own servers are always around 10 times better. This does make the reliability of this model lower. Kainos also created a tool to estimate emissions [101], but like Jetraw, they do not show how they calculate the emissions exactly.

Carbon Intensity

The carbon intensity is the amount of carbon emitted due to the used power consumption. The carbon intensity of the grid can be calculated by first acquiring the generated energy per source and then multiplying it by the carbon intensity of that source. Multiple parties, such as NGOs, governmental organizations or companies who sell this data keep track of the carbon intensity of the power usage of the grid.

A collaboration between national grid ESO, EDF, University of Oxford and WWF led to the calculation of the carbon intensity of the UK [102]. Since 2017 they recorded the hourly carbon intensity. Apart from this they also estimate a forecast based on the weather forecast [103]. Using a bus network they take into account how the energy flows through the system. They measure the amount of energy from each source and multiply it by the carbon intensity for that source.

Another source of information is Electricity Maps, which is a paid open-source program. They used a flow-based approach, where there is a producer and a consumer. This especially works well when there is an electricity exchange between borders. For example, as a source for the Netherlands, they use `https://energieopwek.nl/` which gives the amount of KwH generated per source [104].

2.5.3 Embodied Emissions

While not the primary focus of this research, a large part of the total emission during the lifespan of a server is due to embodied emissions. Embodied emissions are emitted during the manufacturing phase of the product. Depending on the study, embodied emissions are either most of the total emissions [105] or when considering the entire data center, range between 8% and 20% of the total emissions [106].

2.5.4 Ranking and scoring services

Instead of estimating energy consumption or carbon emissions, another approach is to assign a score to the performance of data centers and services. Garg et al. [107] proposed SMICloud a framework for comparing and ranking cloud services, which combines various performance indicators into a single metric. The green grid proposed the use of SERT which measures the work per unit of power [108]. They state that this works better than using idle power as a metric. And that the use of idle power as a metric can lead to a negative effect of which up to 35% of energy could be saved.

Steenhof et al. [109] developed a model to quantify the carbon reductions of different services, which is based on the comparison between the energy consumption for the baseline and an improved scenario. By subtracting these numbers, the reduction is shown. They used it for a relocation of the GeoChronos project. A later study goes into more detail about this [110]. Here they quantified the different emission sources, sinks and reservoirs to determine the reductions.

Apart from the services, software design can also be ranked. Even though often overlooked, it can have a huge impact on power consumption [79]. Katal et al. note that "If software developed is not as efficient as the hardware technology advancements and consumes a large number of resources then overall energy consumption still remains high defying the whole purpose of developing green data centers." [111] The energy consumption of software can be improved using multiple different techniques such as using containers instead of virtual machines and load balancing.

The software design can be improved by looking at different scores, such as the Tool to Estimate Energy Consumption [79]. Another metric was the Resource Utilization Score (RUS), which is a star-shaped metric to rank different services which takes into account the memory, power, disk and CPU. RUS cannot be used across classes of systems, but can be used to improve the software [112]. RUS was later improved upon and the Software Energy Consumption (SEC) metric was proposed. This is defined as in Equation 2.3 [113].

These studies show that it is important to see energy as another resource to include in calculations.

$$SEC = ECwhileoperating - idleEC \tag{2.3}$$

Chapter 3

Case Study

This research is exploratory research to give an in-depth study of how to include accountability of cloud services with different stakeholders. The use case is defined based on the guidelines presented by Runeson and Höst [114]. In Section 3.1 the case is described in detail. Then the case problem will be discussed in Section 3.2. The methodology employed to approach this case is described in Section 3.3. The requirements for the model are identified in Section 3.4. The collection procedure of the data is in Section 3.5. In Section 3.6 the procedure to be followed for data analysis is described and the data validation is reported in Section 3.7.

3.1 Case Description

Tenants of BT are interested in the carbon footprint of the services they utilized and how energy efficient they are in comparison with other tenants. As explained in the introduction the focus of this work is on the CCC call service. BT's objective is to calculate the total carbon footprint and present it to the tenants. To achieve this, it is crucial to rely on precise calculations based on trustworthy sources. Additionally, the research of Westerhof showed that the tenants were interested in assessing their performance in comparison with each other as well as identifying ways they could reduce their carbon emissions.

The CCC service is hosted in third-party data centers, marking the combined entity of BT and the data center provider responsible for its operation. BT provides its own clients, referred to from now on as tenants, the CCC system for answering calls in their call centers. These call centers are staffed by employees who use the CCC software, which are referred to from now on as users. The users handle the incoming phone calls and each time they answer the phone it is called a contact. These contacts are out of the scope of this model as will be explained in Section 4.2.3. For the users, the log-in and log-out times are recorded and after some time it will be converted to the maximum number of users per half hour, called the concurrent users.

CCC is developed by Cisco. Apart from this Cisco also provides the hardware that is needed for this software. The hardware is owned by BT. Currently, there is a movement to transition the entire service to Cisco. This entails that Cisco will be the owner of the hardware. BT can then rent the software for their clients. This service is hosted in the private cloud of Cisco. Cisco can provide us with some data in order to make a comparison.

3.2 Problem Definition

It is important for BT's tenants that the carbon footprint is calculated as accurately as possible. Since the tenants want to report their scopes as accurately as possible. However, since there are multiple tenants involved it is not fair to fully make one tenant fully responsible. The distribution to different stakeholders has to be done in such a way that their reported emissions for the GHG protocol are correct.

Providing a framework to address this issue will be the primary focus of this thesis. As a significant portion of the equipment is shared to a certain extent, it is necessary to differentiate the emissions attributed to each tenant. In order to provide a more accurate perspective and by making stakeholders responsible for the entire service, while not placing the full responsibility on their shoulders.

Another challenge lies in obtaining the data. Cisco is the owner of the private cloud and holds the required data. Since this is a third party, receiving the data can be a little bit more troublesome. Cisco only host it fully for one tenant. BT owns the hardware for the other tenants, which makes more data available. However, as this is a pilot study a lot of the infrastructure needed to collect the necessary data is not in place in the Cisco private cloud model. The main example would be the network equipment as the power consumption is not recorded at this point. Lastly, due to the way the concurrent users were recorded, only the maximum per half hour was available.

Another challenge arises from the comparison of different tenants. The infrastructure for the tenants can differ widely as they have different requirements. In particular, there are three different requirements each tenant can have differently. First, they have a different maximum amount of concurrent users and the service should be able to handle this. Second, the infrastructure is adapted to different degrees to adapt to the tenant's wishes. The third requirement is the different usage profiles. While some tenants open their call centers for a specific country or region, others serve the entire world. To ensure a fair comparison, it is necessary to normalize the data.

3.3 Method

To study this case, we will begin by reviewing the model proposed by Westerhof. Then based on the literature this model is to be improved in close cooperation with BT. The model will have to include a better way to share responsibility between the stakeholders. On top of this, a metric will have to be designed to compare different tenants. Once this is completed, the model will be applied to different tenants to verify the results. Furthermore, a survey will be distributed among different stakeholders to assess how intuitive the metrics are.

3.4 Requirements

One of the major deliverables at the end of the project will be the model to estimate the carbon footprint. The following requirements were made based on studying the case, talking to different stakeholders, and taking into account the literature, including the research done by Westerhof.

Functional requirements

- The model must compute the carbon footprint per concurrent user.
- The model must include a way to calculate Scope 1, 2 and 3 emissions.
- Include a way to compare different tenants

Non-functional requirements

- The model must be fair.
- The estimates from the model should be grounded in the literature.
- The model has to be transparent.
- A guidance for policy implementation for allocating the emissions among stakeholders should be provided.

3.5 Data collection procedures

The data will be collected by BT and Cisco. As per the definition of Runeson and Höst, the data collected indirectly without direct involvement is referred to as second-degree data. For data from BT Cloud, we will receive the data directly from BT in the form of comma-delimited CSV files and logs. The data will be collected for one tenant, which is denoted by the letter M, which will be used to develop the model. Moreover, the data for six more tenants, denoted by the letters A-F, on the basis of their different sizes and usage footprint will be collected for comparison and validation purpose. The tenants are identified by single letters to ensure confidentiality.

3.6 Analysis procedure

The results are analyzed quantitatively to determine if the collected data can be interpreted and estimated based on a correlation. The model from Westherof is adapted and used. To find correlations and other relations the data was plotted in various ways. After this, a metric was developed in order to compare different tenants. Two ways were used to analyze it. The regression was checked with the root mean squared error. This classified how good the regression was. To determine how well it is understood the metric was explained in a document and sent together with a survey.

There was the possibility for abnormalities in the data. BT stated that there is a linear relationship between the concurrent users and the server's energy consumption with two possible exceptions. One event is the clear cache command, which is performed once a year, requiring every user to reload the entire cache, which causes a massive spike. There could also be a random peak in calls. Which would happen only during rare events. It was checked whether these events happened during the time the data was collected, none of these events happened during this time period.

3.7 Validity procedure

Two things are developed in this thesis. A way to estimate the TCFP and a method of comparing different tenants. Validating the model improvements is out of the scope of this work. Currently, there is no realistic way of doing so, due to the lack of data and other models. We, therefore, rely on correctness by construction. In order to validate if the method to compare different tenants is useful a survey is sent to the employees of BT. They will receive an email with the survey along with an explanation of the different types of scores. In this survey, three different ways of comparing tenants will be discussed. The survey consists of four groups of questions: personal, control, usefulness and overall. The personal group of questions has four questions, their name, email in case the author needed additional information, the company they represent and their role in the company. The second group is the control questions. For each of the three metrics, a hypothetical situation is created where they score differently based on one of the three metrics. Then the respondents will be asked which one is better. This way it can be checked if the respondents fully understood each of the metrics. The third group asks which of the three is most useful. This consists of asking which metric they feel is most appropriate for comparing and why they have chosen this metric. And two Likert scales are added to see how clear and useful it is. The last group is the overall questions. Here they get all the scores and are asked with a Likert scale whether it makes it easier to compare different tenants. There is also one field added for any comments or feedback not explicitly asked for.

Chapter 4

Model

After reviewing different models for estimating the carbon footprint of services, it was decided to use Westerhof as a starting point.

In the following chapter, we will first look at the model proposed by Westerhof [23]. Then we will address the shortcomings of this model, and we will propose an extension and improvement.

4.1 Westerhof model

Westerhof's model is based on the GHG protocol and calculates the Total Carbon Footprint (TCFP) as defined in Equation 4.1 [23]. In his work, Westerhof calculated the TCFP per tenant. This also extends to the different scopes. This means all the scopes are already calculated for the specific tenant or can be converted using Equation 4.3. All energy measurements are done in Watt-hours and emissions are done in grams.

$$TCFP = Scope1 + Scope2 + Scope3 \tag{4.1}$$

4.1.1 Responsibility

Westerhof used the method proposed by Lenzen and Murray [14] to share responsibility and avoid double counting. This is important for calculating global carbon emissions. This method uses a percentage share, where a certain percentage x is assigned to the cloud service provider and $1 - x$ to the tenant. In the case of multiple tenants, it will be split accordingly as seen in Equation 4.2. The multitenancy share is based on the estimated Scope 2 energy consumption, expressed as the percentage of energy consumption that the tenant uses of the total energy consumption of the data center. The value is then used in Equation 4.3. Where L_{share} is the share attributed to the tenant. Westerhof attributed everything to the tenant, thus L_{share} is 100%.

$$multitenancyshare = \frac{Tenant_{scope2}}{DC_{scope2}} \tag{4.2}$$

$$r = multitenancyshare * L_{share} \tag{4.3}$$

4.1.2 Scopes

Each scope used in Equation 4.1 can be broken down.

Scope 1 emissions are all devices that directly cause emissions, such as backup generators. Most data centers have these for when power is out. It is defined in Equation 4.4. Where F_{device} is the amount of fuel consumed and c_{device} is the carbon intensity of the device.

$$Scope1 = \sum_{device \in devices} (F_{device} * c_{device} * r) \tag{4.4}$$

In the case of BT, they do not have any of these devices except for the backup generators, which are negligible. Therefore the Scope 1 emissions are zero. Moreover, since this also produces energy, which is counted for Scope 2 emissions this can lead to double counting. When this model is used for a data center for which there are Scope 1 emissions this is something to look out for. Scope 2 includes all emissions caused by energy use and is defined in Equation 4.5.

$$Scope2 = \sum_{DC \in DCs} ((E_{DC_{server}} + E_{DC_{network}} + E_{DC_{cooling}} + E_{DC_{misc}}) * c_{DC} * L_{share}) \tag{4.5}$$

Where DC is each data center the tenant is hosted on. $E_{DC_{server}}$ is the energy consumed by the server. $E_{DC_{network}}$ denotes the energy consumed by the network. $E_{DC_{cooling}}$ represents the energy consumed by the cooling system. $E_{DC_{misc}}$ is the miscellaneous energy consumed. Additionally, c_{DC} is the carbon intensity of the data center.

Scope 3 takes into account the amount of energy the data centers Scope 3 is generating and multiplying it with the responsibility share. This can be found in Equation 4.6.

$$Scope3 = Scope3_{DC} * r \tag{4.6}$$

Where $Scope3_{DC}$ is the Scope 3 emissions for the entire data center.

4.1.3 Scope 2

The energy consumption of the server is estimated using the Bohra and Chaudhary model. And is defined as in Equation 4.7 [91].

$$E_{server} = c_0 + c_1 P_{CPU} + c_2 P_{cache} + c_3 P_{DRAM} + c_4 P_{disk} \tag{4.7}$$

Where c_0 gives the energy consumption when the computer is in idle mode. The parameters c_1, c_2, c_3, c_4 are the weights for P_{CPU}, P_{cache}, P_{DRAM}, P_{disk} respectively. CPU_CLK_UNHALTED is used for P_{CPU}. The INSTRUCTION_CACHE_FETCHES and DATA_CACHE_FETCHES are used for P_{cache}, P_{DRAM} uses DRAM_ACCESSES. Lastly, P_{disk} uses the number of bytes read and written to the disk.

To get these parameters a linear regression model is used in the same format as Equation 4.7.

$E_{DC_{network}}$ is the average amount of energy used per byte times the number of bytes sent over the network. The amount of energy used is 0.06 kWh per GB [90]. From this follows the Equation 4.8.

$$E_{DC_{network}} = 6 * 10^{-8} * (bytes\_send + bytes\_received) \tag{4.8}$$

Westerhof defined $E_{DC_{cooling}}$ as Equation 4.9.

$$E_{DC_{cooling}} = \sum_{dev \in cooling\_devs} dev * multitenancy\_share \tag{4.9}$$

And similarly, $E_{DC_{misc}}$ is defined as in Equation 4.10. The change here is that there might be a change in that a device is directly affecting only the hardware used by the tenant.

$$E_{DC_{misc}} = \sum_{dev \in misc\_devs} \begin{cases} dev.power * dev.carbon\_intensity, & \text{if dev is a directly attributable emission source} \\ dev.power * dev.carbon\_intensity * multitenancy\_share, & \text{otherwise} \end{cases} \tag{4.10}$$

4.1.4 Net emissions

We now have the total amount of carbon emitted. However, data centers use practices to offset the carbon footprint. The TCFP can be calculated by subtracting the offset methods from the gross emissions. The calculation can be seen in Equation 4.11.

$$TCFP_{net} = \sum_{DC \in DCs} (TCFP_{DC} - E_{DC_{green}} * c_{DC} - REC_{DC} * r) \tag{4.11}$$

4.2 The improved model

Westerhof's model has multiple shortcomings. The model uses data that is not generally available, such as the information for E_{misc}. Moreover, there is also a circular dependency in his calculations. These will be addressed in the proposed model.

In addition, it is hard to divide accountability among different stakeholders, and a solution for this will be proposed.

4.2.1 Updating the model

Firstly, the model had to be updated in order to work with available data and address its other shortcomings.

One of the shortcomings is in the original model the multitenancy share is calculated in such a way that there is a circular dependency. The multitenancy share is calculated with the use of the Scope 2 emissions (Equation 4.2). Scope 2 emissions are based on the energy consumption for cooling (Equation 4.5), which in turn uses the multitenancy share again (Equation 4.9).

This model will work in the absence of data for the cooling and the miscellaneous devices, but when the PUE is available. This is most often available and is also our use case. The model assumes that the combined power of the server and network is equal to the total energy consumption by IT equipment. This gives Equation 4.12[1].

$$E_{DC_{cooling\_total}} = (PUE_{DC} - 1) * (E_{DC_{server}} + E_{DC_{network}}) \tag{4.12}$$

Now $E_{DC_{cooling\_total}}$ and $E_{DC_{misc}}$ are combined into one parameter. Equation 4.12 can then also be used in order to update Equation 4.2 to Equation 4.13.

$$multitenancyshare = \frac{E_{DC_{server}} + E_{DC_{network}} + E_{DC_{cooling\_total}}}{DC_{scope2}} \tag{4.13}$$

Substituting $E_{DC_{cooling\_total}}$ in Equation 4.13 gives Equation 4.14.

$$\begin{aligned} multitenancyshare &= \frac{E_{DC_{server}} + E_{DC_{network}} + (PUE_{DC} - 1) * (E_{DC_{server}} + E_{DC_{network}})}{DC_{scope2}} \\ &= \frac{PUE_{DC} * (E_{DC_{server}} + E_{DC_{network}})}{DC_{scope2}} \end{aligned} \tag{4.14}$$

4.2.2 Sharing responsibility between the tenant and data center provider

In order to separate the policy from the model calculation the equations will become more parameterised. This enables to more accurately hold stakeholders responsible for the part they are responsible for. Resulting in a more fair distribution between the different stakeholders. We divide Equation 4.1 into two different sections. The static power consumption E_{static} and the dynamic power consumption $E_{dynamic}$. In order to assign responsibility, we define the parameter γ for the percentage of energy that a tenant is responsible for E_{static} and ζ for $E_{dynamic}$. These parameters are on a scale from 0-1. A value of zero defines no responsibility and one defines full responsibility.

These parameters now replace L_{share} in Scope 2. The attribution is now split into static and dynamic power. This changes Equation 4.5 to Equation 4.15.

$$Scope2 = \sum_{DC \in DCs} ((E_{DC_{server}} + E_{DC_{network}} + E_{DC_{cooling}}) * c_{DC}) \tag{4.15}$$

L_{share} cannot be expressed solely in terms of γ and ζ in a straightforward manner. It also depends on other parameters later explained in this section and these parameters sometimes change on an hourly basis. More about the combination of the energy consumption of $E_{DC_{cooling}}$ and $E_{DC_{misc}}$ below. However, we still need L_{share} for Scope 1 and Scope 3 emissions. For that reason, this parameter is still important.

[1] Technically the cooling devices now also include lighting and other stuff like this. But the major part will be the cooling

Server The determination of the static part of $E_{DC_{server}}$ is straightforward. As the estimation already has an idle (E_{static}) and $E_{dynamic}$ part. This is calculated using the same model as Westerhof as in Equation 4.7. For clarity's sake we rename c_0 to $E_{DC_{server\_static}}$. However, since in this use case, $E_{DC_{server}}$ is already known since it is tracked by the server itself, this does not need to be estimated. Nevertheless, the model is still used to determine the static and dynamic parts. In order to estimate $E_{DC_{server\_static}}$, the model is fitted with the data of the server. Then the $E_{DC_{server\_static}}$ is subtracted from the hourly power consumption to get the dynamic power consumption of the servers. This leads to Equation 4.16.

$$E_{DC_{server}} = \gamma * E_{DC_{server\_static}} + \zeta * (E_{DC_{server\_total}} - E_{DC_{server\_static}}) \tag{4.16}$$

Network It would improve the model if all energy consumption parts could be split into dynamic and static parts. In the model used by Westerhof $E_{DC_{network}}$ is not divided into a static and dynamic part of the power consumption. For that reason, we redefine the network component. The formula can be seen in Equation 4.17.

$$E_{DC_{network}} = \gamma * E_{DC_{network\_static}} + \zeta * E_{DC_{network\_dynamic}} \tag{4.17}$$

In not all use cases the static and dynamic parts can be estimated. Then the following methodology is proposed. First, the total energy consumption is calculated using a constant amount of kWh per GB. This is done in Equation 4.8 and from now on this value is called $E_{DC_{network\_total}}$. This value is then used to calculate Equation 4.18.

$$E_{DC_{network}} = \gamma * \mu * E_{DC_{network\_total}} + \zeta * (1 - \mu) * E_{DC_{network\_total}} \tag{4.18}$$

μ is the percentage of static power consumption. To calculate this value we look at the study from Mahadevan et al. in which they rate network equipment [88]. This is a study from 2009, so the values might be changed a bit, but despite this, it does improve accuracy, since at least some division can be made. They look at different network devices and look at their power consumption. The idle power is determined by plugging in no cables. The maximum measured power is the power consumption when all network cables are connected. The rated max power is the max power consumption reported by the device manufacturer. We choose to use the idle power and measured max power to determine μ, as this is the same way as was done in [88]. We take the average over all these types of network devices except for wireless access points (WAP). Based on the assumption that data centers will not use the WAPs. In this study, the percentage of idle energy against the max measured energy lies, depending on the network device, between 0.75 and 0.91 with an average of 0.81.

This gives an μ of 0.81. To validate this number, we look at the catalyst 9300 series from Cisco. The specification in Table 22 in their documentation[2] reports similar values. Thus we conclude that even though μ is an estimate on average for all network equipment in 2013, this value is still applicable today.

However, this is just for the network devices themselves and not the power used for the network cables. Baker et al. [86] shows the components used to send the data. In their paper, they show the power consumption per component for the shortest path. Network cables only account for 1% of the total energy consumption and thus can be removed from the calculation.

Cooling & Miscellaneous devices Like the calculation for the network, we now also adapt $E_{DC_{cooling\_total}}$ to include a static and dynamic part. This results in the Equation 4.19

$$E_{DC_{cooling}} = \gamma * \nu * E_{DC_{cooling\_total}} + \zeta * (1 - \nu) * E_{DC_{cooling\_total}} \tag{4.19}$$

We now have to determine ν. If for a data center, there is more information ν can be changed to reflect this. For now, we will give an average. ν is determined by two different parts. Firstly, the cooling devices have an idle consumption which is, due to the differences in cooling devices, difficult to determine exactly. Secondly, idle servers and network devices have to be cooled. This cost a certain amount of power. Due to

[2] https://www.cisco.com/c/en/us/products/collateral/switches/catalyst-9300-series-switches/nb-06-cat9300-ser-data-sheet-cte-en.html#Specifications

calculation Equation 4.17 and Equation 4.16 we know for any given hour the division between static and dynamic power. Therefore, ν can be calculated as stated in Equation 4.20

$$\nu = \frac{E_{DC_{server\_static}} + \mu * E_{DC_{network\_total}}}{E_{DC_{server}} + E_{DC_{network\_total}}} \tag{4.20}$$

4.2.3 Concurrent users

Currently, the model described above gives a good indication of a single data center. However, we also want to compare it to different data centers with different structures. Thus, it needs to be normalized since the usage of the system and the workload can vary widely.

There are multiple options on how it can be normalized. It can be done on a per-tenant basis, which is what Westerhof does [23]. The workload from different tenants can, however, vary widely. Which makes it difficult to compare different tenants. For that reason, we need a more fine-grained method.

The most fine-grained method is using the number of contacts. However, for different use cases, the workload can also vary a lot as the time needed for contacts varies a lot.

Then we have the number of concurrent users. We know that the stress they put on the service is linear with the number of concurrent users. For that reason, it is divided by the number of concurrent users per hour.

However, there are multiple ways of measuring the number of concurrent users. A standard way of measuring is with minimum and maximum values. These give certain problems. For the minimum, if in a given hour 10 users log in, but in the middle of this hour, there may be one minute where only one person is logged in, resulting in a warped view. For the maximum, it is the other way around. If 10 people log in 5 minutes before the end of the hour, the maximum is higher than the actual use and gives a higher usage than desirable. This leaves the average. This does need a good method of averaging. Just using the minimum and maximum values and diving it by two does not give the true average. We got the timestamps when people log in and out, this means we can get the total number of minutes people are logged in during a given hour. Dividing this by 60 gives the Concurrent User Equivalent (CUE) of concurrent users during that hour. This is formulated for a given hour h in Equation 4.21.

$$CUE = \frac{1}{60} * \sum_{cu \in concurrent\_users} (\#minutes\ in\ h\ for\ cu) \tag{4.21}$$

This value can then be used to normalize the TCFP for a given hour. So the TCFP per concurrent user for a given hour is Equation 4.22. The $TCFP_{CU}$ can then be used to compare how efficient the different tenants are. This will be further explained in Chapter 7.

$$TCFP_{CU} = \frac{TCFP}{CUE} \tag{4.22}$$

4.2.4 Carbon intensity

Westerhof's model used a constant as carbon intensity in order to translate power consumption to their carbon footprint. In order to improve this we use hourly carbon intensity data. For this research was chosen to use the data freely available at https://carbonintensity.org.uk/, which provides the carbon intensity for the London data centers. For the purposes of this pilot study, only one source was chosen. More data can be found at https://app.electricitymaps.com/map, but this is outside the scope of this thesis.

4.3 Policy

Determining the division of emissions between stakeholders can be challenging, as there are different perspectives on sharing responsibility. One extreme attributes everything to the tenant, while the other extreme attributes everything to the provider. The question of which party is responsible and for which percentage is not a straightforward task. To address this, we create a range of responsibility by defining a lower and upper bound. Where the lower bound is the minimum part the stakeholder is responsible for, and the upper bound is the maximum part for which the stakeholder is responsible for. This can make multiple stakeholders responsible without double counting. Due to the flexibility of the formulas, we can now differentiate between the E_{static} and $E_{dynamic}$ power consumption. We can also assign a different share for Scope 3 emissions. These can also have a lower and upper bound, but for clarity's sake, it will have its own paragraph.

Lower bound To calculate the lower bound, we have to define ζ, γ and LSHARE. One possibility is to set both ζ and γ to 0. Here, the tenant takes no responsibility and here the full responsibility lies with the data center. This gives a free card to the tenant, which means they will care less about reducing carbon emissions. This will ensure that the data centers want to reduce carbon emissions since they are fully responsible.

However, when part of the emissions is the responsibility of the tenant, the data centers would also want to reduce this part since it could be a selling point.

The emissions caused by E_{static} are not under the direct control of the tenant, which should be reflected in the value of γ.

The emissions caused by $E_{dynamic}$ are something the tenant has control over. They have direct control over the resources used. This should be incorporated into the decision about ζ.

Upper bound For the upper bound, it depends on the structure of the hardware used. ζ, γ and LShare must be at least the same value as used in the lower bound.

For $E_{dynamic}$ it follows the same line of thought as for the lower bound.

E_{static} is harder to determine. There are multiple schools of thought. The first is that the tenant is not responsible for the hardware used by the data center. Since this is already the lower bound. The second is that they are both equally responsible for it. The tenant will try to get the data center with the lowest footprint and due to this split the static part is still taken into consideration. The data center also takes responsibility. This will ensure that they will purchase hardware and schedule while taking the emissions into account. The last one is that the tenant is fully responsible for the emissions. This could lead to a tenant choosing another data center. The data center however is not responsible in his scenario, and for their sustainability report, they do not have to report any emissions. This would give a warped viewpoint.

Multitenancy Share & LShare For the multitenancy share, there are also multiple schools of thought. This section will focus on using this for the Scope 3 emissions. Here, multiple philosophies can be used as a foundation.

There are two philosophies for LShare. One is they are not responsible. So LShare is 0.

Another is by dividing the share between the tenant and the data center. Where like the static part of the upper bound multiple values can be used. These follow a similar pattern.

In Equation 4.3 r is calculated by multiplying the multitenancy share by LShare. Since LShare is now non-zero, the multitenancy share also has to be defined.

For the multitenancy share, there are also two different options. One takes the share of the actual use. For example, you have 5 concurrent users during a given interval, and in total 15 concurrent users are giving a contribution to the Scope 3 emissions. Your multitenancy share would be 0.33.

The other takes the share of the maximum use. If you have a system that at a maximum could handle 100 concurrent users, and in total the system could handle 150. You have a multitenancy share of 0.66.

The last one is more accurate and is highly desirable. However, this data is also harder to come by. The maximum capacity needed for the tenant's service is known by the tenant. The maximum capacity of the other tenants is unknown.

For this reason, the first policy is used. We take the average consumption per month for the data center and the total monthly consumption of energy consumption. Dividing these as per Equation 4.13 will give this result.

Now with this multitenancy share, the total amount of emissions caused by using their service is calculated. Then with LShare, the amount responsible has to be determined.

The energy consumption for which the tenant is responsible is somewhere between the given upper and lower bound. This depends on the agreements between the tenant and the data center. For the data center, it is the inverse. However in the case of BT and Cisco: Cisco is the data center provider and BT is a middleman. In the formulas above, these together are seen as the data center provider. The two should come to an agreement to fairly distribute the emissions between them. This is outside the scope of this thesis.

4.3.1 Converging to Westerhof model

This model is also backward compatible with Westerhof's model. Westerhof's model only has one value instead of a lower and upper bound. So the upper and lower bounds have to be equal to also get one value.

To show how to converge the parameters, we will use the cooling part in Equation 4.19. However, this is equal for all individual parts. We do assume that, for example, the value of $E_{DC_{cooling\_total}}$ is the same as in Westerhof's model. γ and ζ are used instead of LShare. If we take $\gamma = \zeta = LShare$ we can rewrite Equation 4.19 as in Equation 4.23.

$$\begin{aligned} E_{DC_{cooling}} &= LShare * \nu * E_{DC_{cooling\_total}} + LShare * (1-\nu) * E_{DC_{cooling\_total}} \\ &= LShare * (\nu * E_{DC_{cooling\_total}} + (1-\nu) * E_{DC_{cooling\_total}}) \\ &= LShare * ((\nu + 1 - \nu) * E_{DC_{cooling\_total}}) \\ &= LShare * E_{DC_{cooling\_total}} \end{aligned} \tag{4.23}$$

Equation 4.23 shows that the value of ν, and with the same logic, also μ does not matter. As long as $\gamma = \zeta = LShare$.

This shows that the Westerhof model is equal to the upper bound of this model. So in theory this model will allocate less to the tenant and will also hold the data center accountable.

4.3.2 BT Use Case

Lower bound For the lower bound, for the rest of this work we presume that the $E_{dynamic}$ emissions are fully the responsibility of the tenant. If they did not do anything, that energy would not be used. For that reason, we propose that ζ should be 1. The tenant can argue they can do nothing about how inefficient a data center is allocating its servers. So γ should then be 0.

Upper bound For the upper bound it depends on the structure of the used hardware. $E_{dynamic}$ still has a ζ of 1 for the reason described above. For E_{static} we choose a γ of 0.5. Since the tenant has no full control over the allocation of resources. However, if we would remove this from our equation comparing different data centers would not go well. This makes sure that for the tenant a fair comparison can be made while still laying part of the responsibility by the data center.

Chapter 5

Implementation

As discussed in Chapter 3, BT has specific requirements and access to limited data. Once the model was designed it was then fitted for the BT case. Section 5.1 will outline the changes to adapt the model for BT. Subsequently, Section 5.2 will provide a brief explanation for the implementation. The data used in the model first had to be pre-processed. The steps to do this are addressed in Section 5.3. Lastly in Section 5.4 we will examine the output of the model for six tenants.

5.1 Model implementation

In addition to the policy explained in the previous chapter the model itself was adjusted to the available data. At its core, it remains the same model, but specific modifications were made, particularly at the network part, to better suit the data.

5.1.1 Carbon Intensity

The carbon intensity from the UK is available for free download from `carbonintensity.org.uk/`. This is categorized by different locations and one of them is London. Since multiple tenants use the data centers located here, these values were used. Figure 5.1 illustrates the carbon intensity fluctuations observed over a two-month period, indicating no clear visible pattern.

Although not 100% accurate, the data from London was also used for the Netherlands to serve as a proof of concept. To highlight the inaccuracy, maps provided by Electricity Maps [1] show the daily usage for the past month for both the Netherlands and Great Britain. These graphs [2] are shown in Figure 5.2. The comparison shows a minimal correlation between the two, with the Netherlands having a higher carbon intensity in general. This must be acknowledged as a flaw in the data rather than the model itself.

5.1.2 Concurrent Users

BT records the login and logout times for each user. This is just saved for a couple of days. After a couple of days, this is summarized in the maximum amount of users per half hour. Since this model works on an hourly basis the maximum per hour is taken and then used for this algorithm. The concurrent users on an hourly basis can be seen in Figure 5.3.

The figure illustrates very clearly the different days as well as the weekends. Every day there is a peak at around 12 o clock, while there is a valley at midnight. This pattern thus follows a 24-hour cycle. On top of this, there is also a 7-day cycle where the weekends have consistently fewer users logged in.

[1] `https://app.electricitymaps.com/map`

[2] These graphs are a screenshot from the website

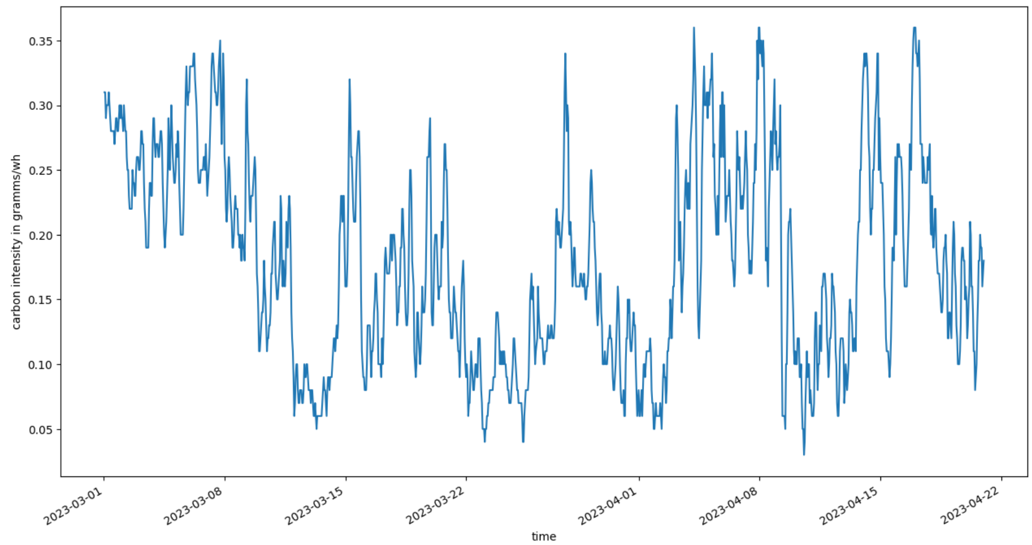
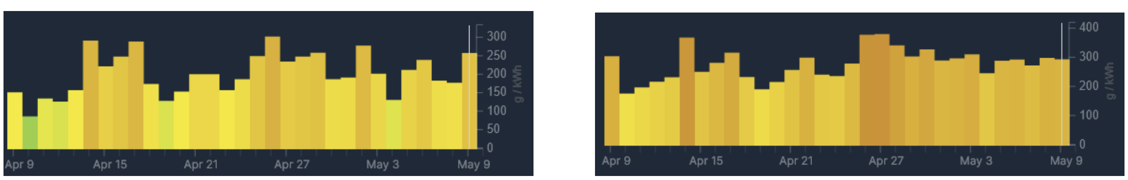


Figure 5.1: Carbon Intensity of London during a 2 month period



(a) Carbon Intensity of Great Britain

(b) Carbon Intensity of the Netherlands

Figure 5.2: Carbon Intensity during between 9 April and 9 May

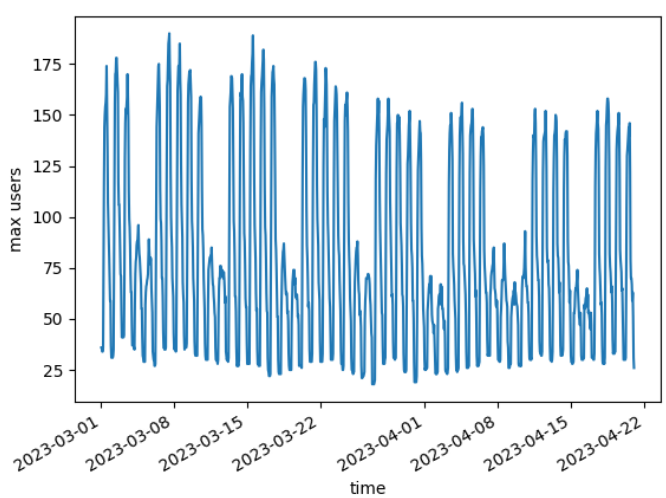


Figure 5.3: Max users per hour

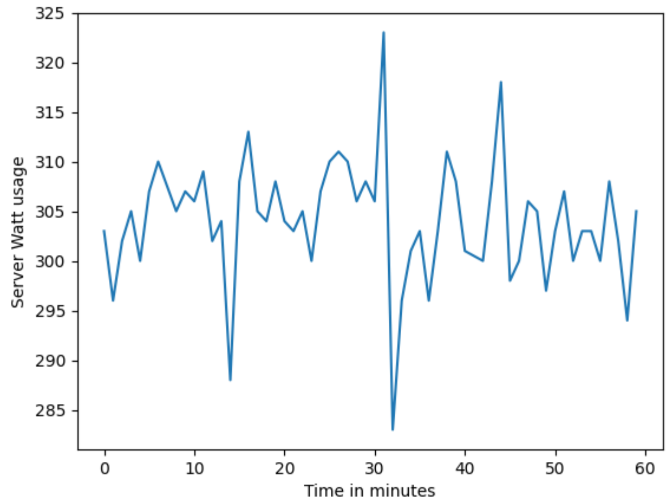


Figure 5.4: Watt usage of a server recorded once every minute

5.1.3 eServer - power consumption

Each server can provide power consumption data per minute through its BIOS. The BIOS only saves this for a week, making it impractical to retrieve the data manually. Instead, we used the data from an administrative dashboard which gets a value once every hour. This does introduce some inaccuracies. The dashboard retrieves the watt usage at a moment during the hour and considers it representative of the entire hour. An example of the data for a random hour can be seen in Figure 5.4.

For this particular hour, the following metrics apply: Avg: 304 watts, min: 283 watts, max:323 watts. Calculating the error for this hour gives an inaccuracy of 20 or around 7%.

The model was developed based on the monthly data for the M case. To compare the six other tenants only a week's worth of data was used. This could be provided via the BIOS, therefore leaving this particular aspect to be only relevant to the M case.

5.1.4 eServer - static part

In theory, we also aim to determine the static energy consumption of the server using multiple parameters such as RAM, disk usage, CPU and memory. However, the hard disk always rotates due to the selected configuration for it, resulting in a constant value. Apart from this the docker containers always reserve the same amount of memory, resulting in a constant value as well. This leaves only CPU usage as a variable factor, which was then used to determine the static server energy consumption.

Due to the consistent nature of the CPU usage for the server, the data was very hard to interpolate. For that reason, another server, from tenant F, was chosen which had a bit more variant data. This server is of the same type, thus will provide accurate results. Based on this server the static part was determined to be 217 watts.

After we did this we cross-verified these values with the data specified in the specifications. BT uses two kinds of servers, M4 and M5. According to the specifications, the idle power for M4 is 238.4 watts and for M5 it is 205.6 watts.

Given the significant difference between the two different types, the TCFP is calculated with distinct values for the idle power. As per Equation 4.16 this only changes the amount of static power and not the total power usage.

5.1.5 eNetwork

In the model's core, the static and dynamic part of the power is split up. This stays the same, only the way the static and dynamic part is calculated changes. We determine the 'true maximum' by taking the median of the maximum power usage per day, denoted as $E_{network\_max}$. Similarly, the minimum power usage, denoted as $E_{network\_min}$, is calculated. The static part of the network was obtained by multiplying $E_{network\_max}$ times μ. To determine the dynamic part, we calculated an hourly μ_t. Where t represents the specific time for which it is calculated. $E_{network\_t}$ denotes the network power usage at time t. However, since we took the mean there were times when the actual maximum power exceeded the 'true maximum'. To address this, μ was capped at 0.81, ensuring it does not go below 0.81. Any value lower than 0.81, defaults to 0.81. This can be seen in Equation 5.1.

$$\mu_t = max(1 - (1 - \mu) * \frac{E_{network\_t} - E_{network\_min}}{E_{network\_max}}, \mu) \tag{5.1}$$

BT's infrastructure includes a backup router, which remains idle until a fault occurs in the original router. The original router is replaced as soon as possible, once a fault is detected. The idle router is included in the static part of the electricity consumption.

Furthermore, BT's infrastructure has the tenants share a single set of routers, meaning they share responsibility. For simplicity's sake, we assume that all tenants have an equal utilization of the router. Therefore the power usage is spread evenly along these servers.

5.1.6 Scope 1 and Scope 3 emissions

Scope 1 emissions only entail the backup generators. These are only turned on when there is a power outage. This has not occurred during the time that the data was collected. For that reason, Scope 1 emissions were set to 0.

Determining the Scope 3 emissions is a challenging task and the data center could give a definitive answer. For that reason, it was determined to set LSHARE to 0. The methodology was kept in the thesis, despite not including it when comparing different tenants, as it shows where the Scope 3 emissions should be put once it has been calculated.

5.2 Technologies

The implementation was done in Python using Pandas. The CSV files were directly read into the pandas and the calculations were performed on those dataframes.

The software is split into different scopes. The main chunk of calculations is the Scope 2 emissions, which is further divided into multiple sections. First, the static and dynamic parts for the server, network, and cooling parts are calculated. Then, these values are multiplied by the carbon intensity during that time.

The software can be found on GitHub [3].

5.3 Pre-Processing

Before using the data for the model some pre-processing steps were carried out.

In the case of certain servers, the energy reported was in watt usage per minute. To standardize the data, it was converted to kilowatt-hours. This algorithm had a couple of simple steps:

- The time difference in seconds between the timestamp of the reported value and the next value was calculated.
- This time difference was converted to hours instead of seconds.
- The watt usage was multiplied by the duration.

[3] https://github.com/RUGAlbert/PrivateAndPublicCloudEstimator

- Lastly this was re-sampled to get the hourly KWh value.

This was not the only data that required pre-processing.

The calculation of the number of concurrent users in a given hour was necessary. Although, this model incorporates a way to give a 'true average', this was not needed due to limitations in the received data. It is still left in the thesis so it can be used in future work.

The carbon intensity also needed some preprocessing. The data made available by www.carbonintensity.org.uk can only be downloaded per 14-day period. This was done for a 2-month period and combined in a single file. It is important to note that the carbon intensity is measured in grams of CO2 equivalent per kilowatt-hour, while the application uses watt-hours. Thus a conversion has to be required. An API could be used in the future in order to automate it and create a live dashboard.

The last preprocessing step is necessary because the tenants have data coming from different time zones. The network and computer data are in UTC, while the concurrent users are in their respective timezone. To ensure consistency, all the data is normalized to UTC.

5.4 Output

As output, the model implementation generates a different CSV file for each server and one CSV file for the total consumption. In total, there are 17 different columns that can be grouped into three different categories. One of these columns is the time, which for clarity's sake is included in each of the groups. The three groups are: energy consumption, emissions and emissions per concurrent user. For clarity, we split this into a table for each group. In each CSV file the following columns are included:

- **time**: The timestamp of that moment
- **scope2E**: The total scope 2 energy consumption
- **eServerStatic**: the static part of the server energy consumption
- **eServerDynamic**: the dynamic part of the server energy consumption
- **eNetworkStatic**: the static part of the network energy consumption
- **eNetworkDynamic**: the dynamic part of the network energy consumption
- **eCoolingStatic**: the static part of the cooling energy consumption
- **eCoolingDynamic**: the dynamic part of the cooling energy consumption
- **scope1**: the Scope 1 emissions
- **scope2Lower**: the lower bound of Scope 2 emissions based on the policy
- **scope2Upper**: the upper bound of Scope 2 emissions based on the policy
- **scope3**: the Scope 3 emissions.
- **TCFPLower**: the lower bound of the total carbon footprint
- **TCFPUpper**: the upper bound of the total carbon footprint
- **ci**: the carbon intensity during that hour
- **maxUsers**: the amount of max users
- **TCFPLowerPerUser**: the lower bound of the total carbon footprint per user
- **TCFPUpperPerUser**: the upper bound of the total carbon footprint per user
- **scope2EPerUser**: the scope 2 energy consumption per user

time	scope2E	eServer Static	eServer Dynamic	eNetwork Static	eNetwork Dynamic	eCooling Static	eCooling Dynamic
01/03/2023 01:00	2065.73	953.6	222.4	561.78	0.0	286.56	41.39
01/03/2023 02:00	2028.34	953.6	190.4	561.78	0.0	286.56	36.0
01/03/2023 03:00	2047.03	953.6	206.4	561.78	0.0	286.56	38.69
01/03/2023 04:00	2018.74	953.6	182.4	561.78	0.0	286.56	34.4
01/03/2023 05:00	2037.6	953.6	198.4	561.78	0.0	286.56	37.26
01/03/2023 06:00	2018.57	953.6	182.4	561.78	0.0	286.56	34.23
01/03/2023 07:00	2027.83	953.6	190.4	561.78	0.0	286.56	35.49
01/03/2023 08:00	2008.97	953.6	174.4	561.78	0.0	286.56	32.63
01/03/2023 09:00	2033.32	953.6	190.4	561.78	4.42	286.56	36.54

Table 5.1: Total energy consumption for CCC during a 9-hour time period

time	scope1	scope2Lower	scope2Upper	scope3	TCFPLower	TCFPUpper	ci
01/03/2023 01:00	0.0	1801.94	2065.73	0.0	1801.94	2065.73	0.31
01/03/2023 02:00	0.0	1801.94	2028.34	0.0	1801.94	2028.34	0.31
01/03/2023 03:00	0.0	1801.94	2047.03	0.0	1801.94	2047.03	0.29
01/03/2023 04:00	0.0	1801.94	2018.74	0.0	1801.94	2018.74	0.3
01/03/2023 05:00	0.0	1801.94	2037.6	0.0	1801.94	2037.6	0.3
01/03/2023 06:00	0.0	1801.94	2018.57	0.0	1801.94	2018.57	0.3
01/03/2023 07:00	0.0	1801.94	2027.83	0.0	1801.94	2027.83	0.31
01/03/2023 08:00	0.0	1801.94	2008.97	0.0	1801.94	2008.97	0.3
01/03/2023 09:00	0.0	1801.94	2033.32	0.0	1801.94	2033.32	0.29

Table 5.2: Total emissions for CCC during a 9-hour time period

The energy consumption group consists of scope2E, eServerStatic, eServerDynamic, eNetworkStatic, eNetworkDynamic, eCoolingStatic and eCoolingDynamic. This group focuses on Scope 2 energy consumption. An example of this output is shown in Table 5.1.

The emissions group consists of scope1, scope2Lower, scope2Upper, scope3, TCFPLower, TCFPUpper and ci. This summarizes the total carbon footprint of the current stakeholder. Scope 1 and Scope 3 do not have an upper and lower bound. Only a single value. Therefore only Scope 2 reports the upper and lower bound.

A table with an example of this part of the output can be seen in Table 5.2.

Lastly, the carbon footprint is normalized, this is summarized in the last group, emissions per user. This group has the following columns: maxUsers, TCFPLowerPerUser, TCFPUpperPerUser and scope2EPerUser.

It also prints out the statistics which will be explained in Section 7.1. An example is below:

time	maxUsers	TCFPLowerPerUser	TCFPUpperPerUser	scope2EPerUser
01/03/2023 01:00	36	50.06	57.38	57.38
01/03/2023 02:00	36	50.06	56.34	56.34
01/03/2023 03:00	34	53.0	60.2	60.2
01/03/2023 04:00	34	53.0	59.37	59.37
01/03/2023 05:00	36	50.06	56.61	56.61
01/03/2023 06:00	77	23.4	26.21	26.21
01/03/2023 07:00	116	15.54	17.48	17.48
01/03/2023 08:00	144	12.52	13.95	13.95
01/03/2023 09:00	150	12.02	13.55	13.55

Table 5.3: Total emissions and energy consumption per user for CCC during a 9-hour time period

```
mSQR of  0.9988614954920007 for a n-value of  1.05
Area score 27.625588432482054
95 percent is less than 78.89913665383224
50 percent is less than 30.09910669661668
5 percent is less than 12.402728049359856
percentages of different components: Server: 0.58 Network: 0.27 Cooling: 0.16
percentages of static/dynamic 0.85 0.15
```

Listing 5.1: example of output values after running the program

Chapter 6

Applying the model

In this chapter, the model will be applied to different use cases. In Section 6.1 the results of applying the model to tenants from BT will be presented. In particular, one tenant will be used to demonstrate the model. Following that, in Section 6.2 we will examine the impact of hourly carbon intensity on the collected data. The effect of different policies will be explained in Section 6.3. Based on these policies, two distinct use cases for calculating the upper and lower bounds will be shown in Section 6.4. Additionally, one tenant hosts its server by Cisco, the received data is discussed in Section 6.5. Lastly, in Section 6.6 we will reflect on the ease of data collection for the model.

6.1 Power Consumption

The primary focus of this research is to estimate the power consumption. Initially, this was done for one tenant throughout the majority of the study (tenant M, as discussed in the previous chapter), the model was later verified by looking at six other tenants (labeled A to F to protect their confidentiality).

The entire energy consumption model, along with the maximum concurrent users per hour is plotted in Figure 6.1. Two noteworthy observations can be made. First, a distinct relation between power consumption and the maximum concurrent users is evident. For that reason, we split the power consumption into different categories. Second, the different days and weekends are also evident. There is a one-day cycle for which the peaks of both plotted values are during the day and valleys during the night. A seven-day pattern is also visible. During the weekends both values have lower peaks.

Figure 6.2 shows the different components and their correlation with the number of concurrent users. This shows no distinct correlation between server power consumption and concurrent users. However, the number of concurrent users and network power consumption has a direct correlation. This is logical, as the network is used more when more users are online.

The energy consumption of both the server and the network is used to determine the energy consumption of the cooling equipment. This causes a correlation between the number of concurrent users and the power consumption of the cooling equipment. The correlation between the energy consumption and the number of concurrent users is, therefore, mainly due to the correlation between the number of concurrent users and the network components.

The different components do not have an equal share in the total power consumption. In this case, on average, the server equipment uses 58%, the network accounts for 27% and the cooling uses 16% of the total power consumption. For the other tenants, this distribution can vary depending on factors such as the PUE and their network usage. An overview for all six tenants can be seen in Table 6.1. For most of the tenants, the server equipment is the largest portion of power consumption, except for tenant E which uses a lot of energy for the network. The cooling has a similar percentage for most tenants, as they use the same data centers. Therefore, their PUE is the same and thus the percentage of energy that is consumed by cooling devices.

On average, the static part of the total power consumption is 87% and the dynamic part is 13%. This distribution was true for all the tenants, with only slight differences of up to one percentage point. Given

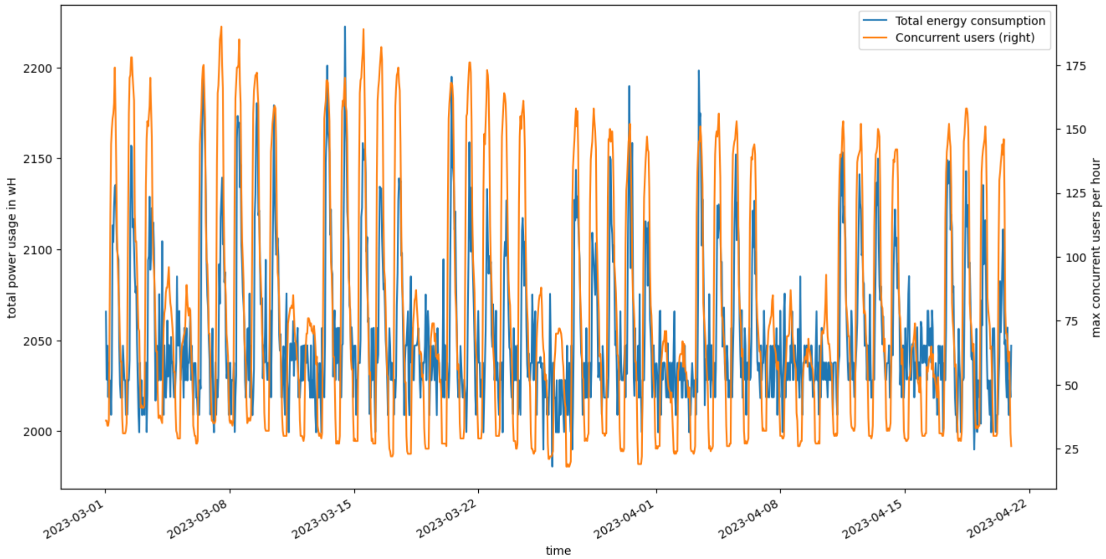
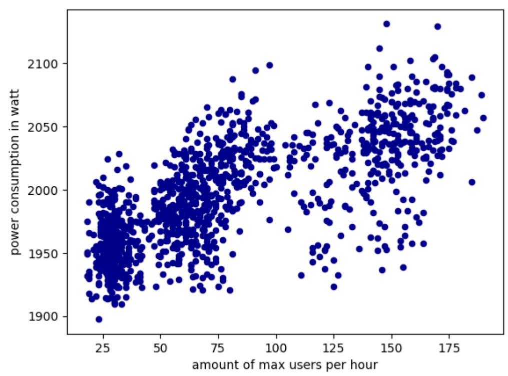


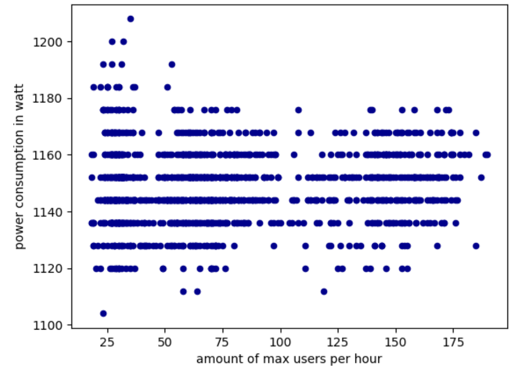
Figure 6.1: The energy consumption and max users over time

Tenant	server	network	cooling
A	0.7	0.15	0.16
B	0.57	0.27	0.16
C	0.78	0.07	0.16
D	0.66	0.18	0.16
E	0.39	0.45	0.16
F	0.45	0.21	0.35

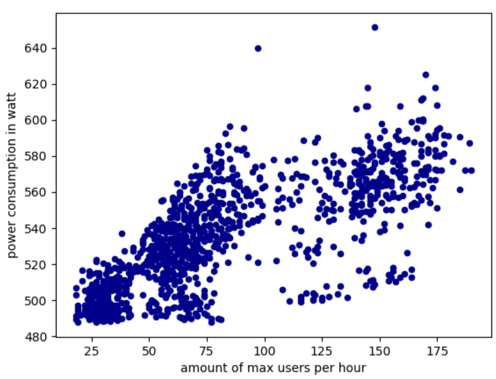
Table 6.1: The share of the components of the tenants



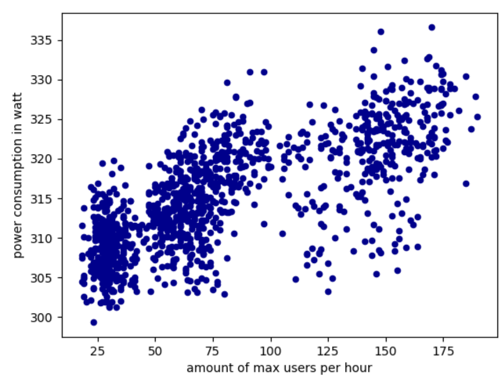
(a) Scatterplot for the total power consumption vs the max users



(b) Scatterplot for the server power consumption vs the max users

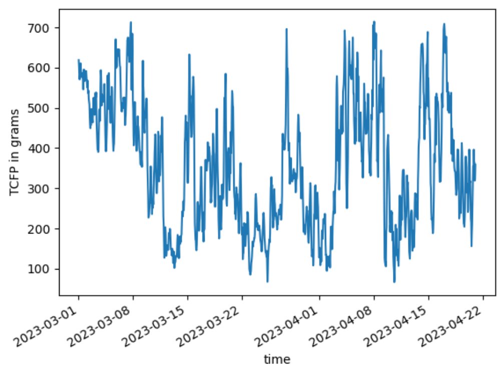


(c) Scatterplot for the network power consumption vs the max users

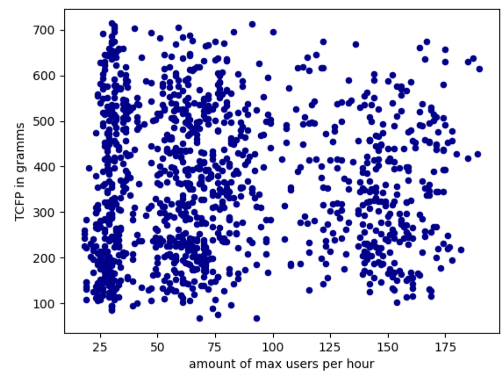


(d) Scatterplot for the cooling power consumption vs the max users

Figure 6.2: Plots of all power consumption vs the max users for the same tenant



(a) Carbon footprint of Scope 2 over a two month period



(b) Carbon footprint of Scope 2 scattered against the max users

Figure 6.3: Carbon footprint of Scope 2 using hourly carbon intensity data

that the static part is most of the energy consumption, the service will become more energy efficient when used by more users.

6.2 Carbon intensity

The carbon intensity, as can be seen in Figure 5.1, fluctuates a lot and has no obvious pattern. Consequently, there is no clear correlation between the number of concurrent users and the TCFP.

Multiplying the energy consumption by the carbon intensity adds in essence an element of randomness. This can be seen in Figure 6.3. Unlike the energy consumption, which had a one-day and seven-day pattern, the TCFP has no visible patterns as can be seen in Figure 6.3a. The TCFP is plotted against the number of concurrent users in Figure 6.3b. This does not reveal any patterns.

Using a constant value, such as the average carbon intensity over a month to remove this element of randomness, would give a warped result. A high carbon intensity during the day, when a lot of energy is used, is worse than a high carbon intensity during the night.

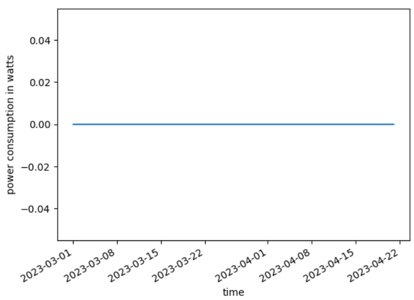
6.3 Policy

The policy changes for which part of the emissions the tenant is responsible. This is achieved by a combination of parameters, namely γ, ζ and LSHARE. The impact of LSHARE will be examined alone, while γ and ζ will be shown together. All power consumption data in the following is for the same tenant. We use power consumption in place of carbon footprint in order to remove the effect of carbon intensity as discussed in the previous section.

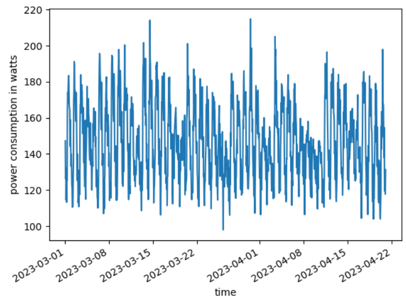
6.3.1 γ and ζ

The combinations of γ and ζ will be shown for the values 0, 0.5 and 1, resulting in different effects. This is shown in Figure 6.4.

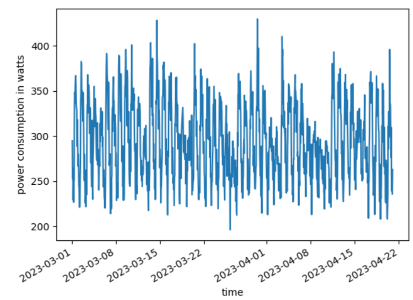
Figures a,d and g only have different values of γ, while Figures a,b and c have different values of ζ. Figures e, f, h and i are a combination of γ and ζ. This shows that the static part is, in fact, static and does not have fluctuations. Therefore, increasing γ only changes the base value. As can be seen in the first column of the figures in Figure 6.4. On the other hand, ζ changes the dynamic value. As ζ increases the valleys and peaks differ a lot more.



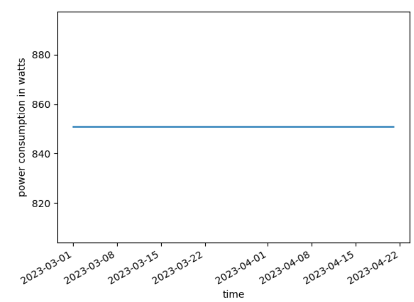
(a) A γ of 0 and a ζ of 0



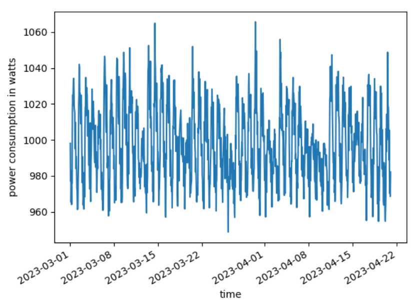
(b) A γ of 0 and a ζ of 0.5



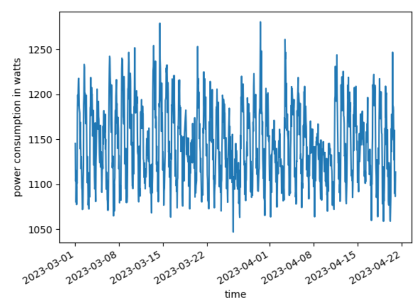
(c) A γ of 0 and a ζ of 1



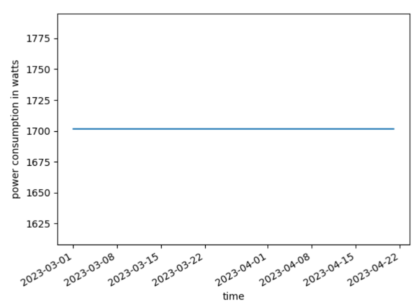
(d) A γ of 0.5 and a ζ of 0



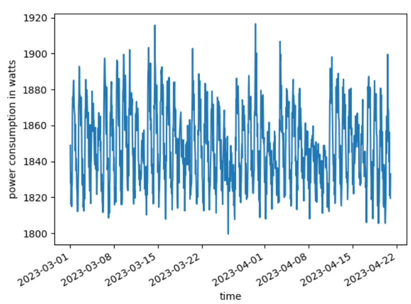
(e) A γ of 0.5 and a ζ of 0.5



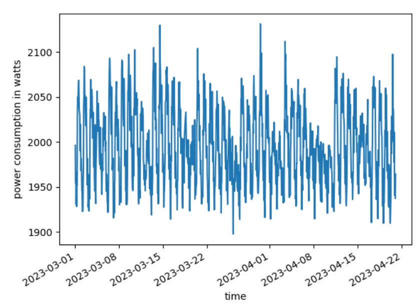
(f) A γ of 0.5 and a ζ of 1



(g) A γ of 1 and a ζ of 0



(h) A γ of 1 and a ζ of 0.5



(i) A γ of 1 and a ζ of 1

Figure 6.4: Different values for γ and ζ

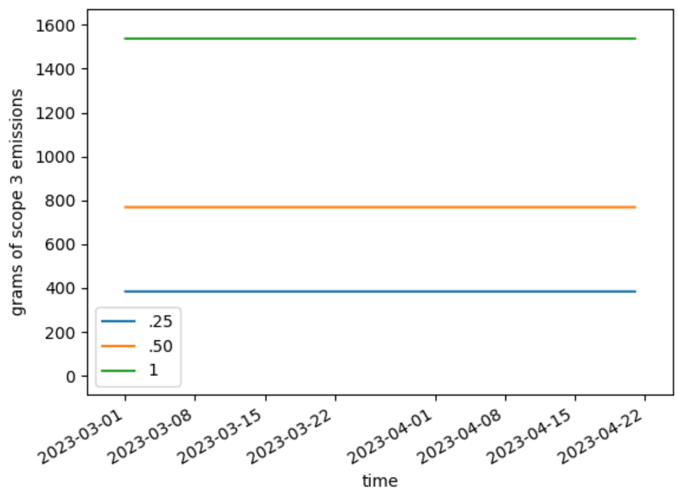


Figure 6.5: Different TCFP for different values of LSHARE

6.3.2 LSHARE

LSHARE determines the percentage of Scope 3 emissions for the tenant. Scope 3 emissions are difficult to determine, which is why it has been left out of the results so far. However, the effect of different values of LSHARE on Scope 3 emissions is still interesting. Calculating Scope 3 emissions for each tenant requires data on the energy consumption of the tenant, the entire data center and the Scope 3 emissions associated with the data center. The Scope 3 emissions per tenant are estimated by using the energy consumption for the tenant, the entire data center, and the Scope 3 emissions for the data center. The last two components are currently unknown, as the necessary data is currently unavailable. For that reason, we decided to calculate the Scope 3 emissions of the tenant via an alternative approach. Downie and Stubbs [12] state 81% of the total amount of the TCFP is attributed to Scope 3 emissions. We used this value to estimate a realistic estimation of the Scope 3 emissions.

The Scope 3 emissions with different values of LSHARE can be seen in Figure 6.5. Where LSHARE values of 0.25, 0.5 and 1 were taken. Scope 3 emissions are mostly calculated as a constant instead of an hourly differential value, resulting in a straight constant line.

Considering the impact of Scope 3 emissions, when incorporating them into the scatter plot which plots the TCFP per concurrent user vs. the number of concurrent users, we obtain Figure 6.6. This shows an inverse relationship between the TCFP per user and the number of concurrent users.

6.4 Upper and Lower bound

The upper and lower bounds show a region for which parts a tenant is responsible. To illustrate this, two different policies will be shown.

Policy A can be seen in Figure 6.7a. Under this policy, the tenant is fully accountable for the static part of the energy consumption. The lower bound, therefore, has a γ of 1 and a ζ of 0. For the upper bound, the tenant is also partly responsible for the dynamic part, resulting in a γ of 1, and a ζ of 0.5.

Policy B can be seen in Figure 6.7b. In this policy, the scenario is reversed and the tenant is fully responsible for the dynamic part of the energy consumption. Therefore, the lower bound has a γ of 0 and a ζ of 1. For the upper bound the tenant is also partly responsible for the static part. Therefore, it has γ of 0.5, and a ζ of 1.

Figure 6.7 also clearly shows the difference these policies make.

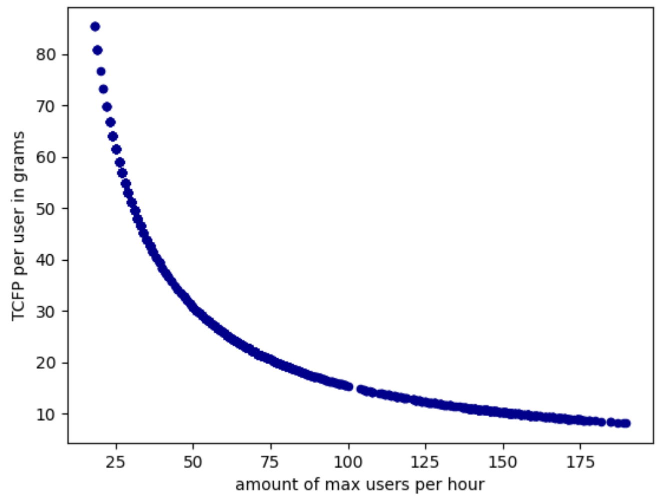
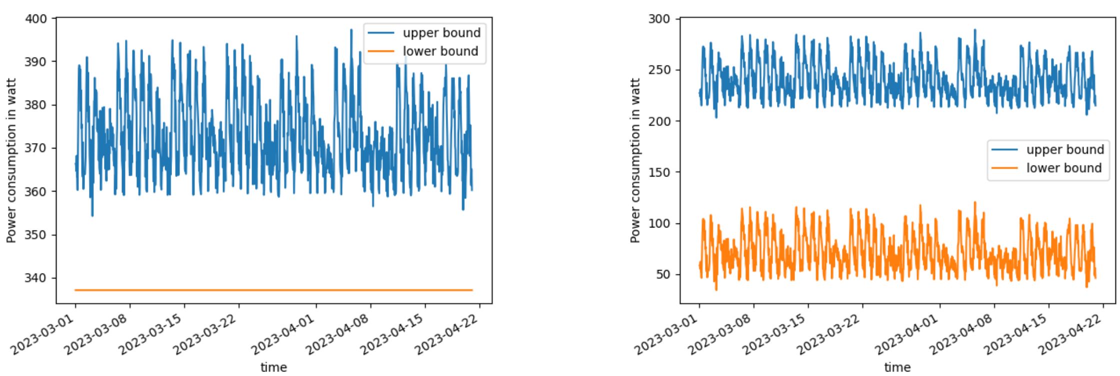


Figure 6.6: Scatterplot for the TCFP per user vs the max users



(a) Lower and upper bound for Policy A

(b) Lower and upper bound for Policy B

Figure 6.7: Difference in having a different policies

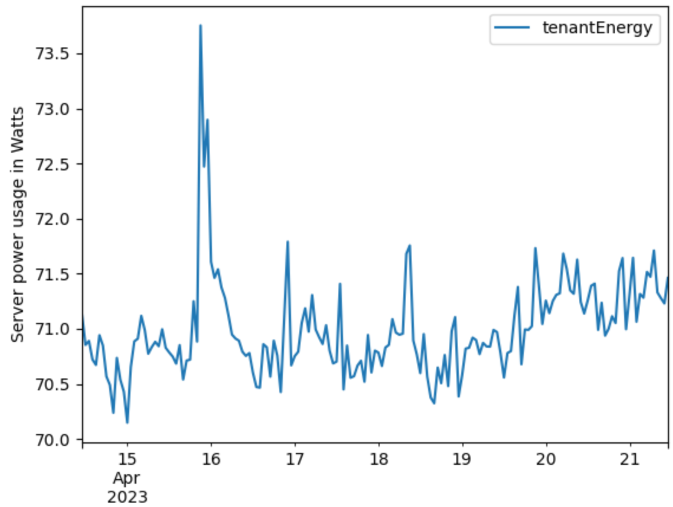


Figure 6.8: The power consumption for one server hosted by Cisco

6.5 The Cisco case

For Cisco, we received the watt usage of their servers, network usage and the PUE. Watt usage was split for each node. A node was used for processing purposes or memory. According to Cisco, the nodes were used by five tenants, and the total watt consumption was therefore divided by five.

The power consumption of one server can be seen in Figure 6.8. Since the power consumption is in all data points divided by five, we assume the total power consumption is the recorded value for an entire rack. The rack is shared by five tenants, of whom they provided us with the average power consumption per tenant. However, this assumption has not been confirmed, but for now, we consider the values to be correct. Comparing the power consumption, an idle BT server consumes at least 205 watts, while Cisco's servers appear to be more energy-efficient based on the recorded values.

For the network usage, we were provided with the average and maximum values for a one-month period. The average network consumption was 19.475 Mbps and the maximum usage was 87.671 Mbps. This is a lot higher in comparison with a similar tenant hosted by BT, where the average was 1 Mbps and the maximum was 24 Mbps. The exact reason for this difference is currently unknown. Multiple plausible theories have been suggested, such as more services using these network devices and increased security measures. However, none of these suggestions have been confirmed at the time of writing this.

BT and Cisco both host their servers from the same or similar data centers. Which leads to similar PUE values. As a result, the cooling power consumption has the same share as the TCFP.

Given the lack of valid and comprehensive data, it is currently challenging to determine how well Cisco performs in comparison to BT. Any concussion should be made with caution. Cisco may consume less energy due to lower server power usage. However, their network usage is higher, and depending on how they have configured their network devices, they either perform better or worse.

6.6 Discussion

In this section, we will discuss the various components that contribute to the input of the model and reflect on the challenges associated with gathering the necessary data. To calculate the TCFP, the three Scopes have to be calculated.

Scope 1 is straightforward, as it is estimated to be zero in the case of data centers. On the other hand, estimating Scope 3 currently poses significant challenges. We hope that further research will improve the Scope 3 estimations.

For Scope 2 emissions, three parts are required: the energy consumption, the carbon intensity and the values of γ and ζ.

The energy consumption can be divided into three parts E_{server}, $E_{network}$ and $E_{cooling}$. Each of these parts can be further split up into static and dynamic power consumption.

In many cases, E_{server} can be easily obtained or estimated. If the hardware keeps track of the power consumption, as in the case of BT, this is straightforward. Otherwise, power consumption can be estimated using a model based on various parameters such as CPU usage. These models have been extensively researched and can provide reasonably accurate estimations. As BA models already base their estimation on the concept of a static and dynamic part, the division between these parts is already done. In the case that the power consumption is known, the static power usage has to be estimated based on various parameters such as the CPU usage. We showed during this project that depending on the usage profile this can be either straightforward or rather challenging.

$E_{network}$ is more difficult to obtain. The models used to estimate power consumption often use parameters that are not readily available. The outdated and averaged estimate for watts per byte used in this thesis highlights the difficulty in obtaining an accurate estimation. Currently, the network power consumption of the devices used in this use case is not recorded. Even if it were recorded, an extra challenge lies in the nature of networks being shared by multiple tenants, making fair allocation of energy consumption difficult. As a result, this makes it difficult to get a better estimate than used in the model. The proposed model to estimate the network energy consumption s can be used for most use cases, as network usage is typically recorded.

Estimating $E_{cooling}$ is relatively easy as the PUE is known for most data centers. Apart from E_{server} and $E_{network}$, this is the only additional part needed.

Apart from energy consumption, the model requires carbon intensity, which can be relatively easily obtained through estimates provided by various companies.

Determining the values for γ and ζ depends on the involved stakeholders. It might prove difficult to determine the lower and upper bounds since no official guidelines have been made for these values yet.

Overall, while there are certain challenges, it is realistic to obtain the necessary data for the model. The two main difficulties lie in estimating the energy consumption of the networks and the Scope 3 emissions. Apart from this, the data can be easily integrated into the model.

Chapter 7

Tenants Comparison

In the following chapter, we will look at the different tenants in the BT case. First the UPES metric will be introduced in order to compare different tenants in Section 7.1. Section 7.2 will focus on which UPES values the different tenants have and how well they score in comparison with each other. Section 7.3 will evaluate the answers from the survey sent to employees of BT. The results will be discussed in Section 7.4. Lastly, Section 7.5 will suggest improvements tenants can make to reduce their TCFP.

7.1 Comparing different tenants

To compare different tenants with each other, it is necessary to normalize the energy consumption. The normalization is done as per Equation 4.22, where the energy consumption replaces the TCFP. This adjustment is made because the carbon intensity is semi-random and not in the control of the tenants. The difference in using the energy consumption instead of the TCFP can be seen in Figure 7.1.

In Figure 7.1a, the total power consumption follows a curve, which can be approximated by an inverse relation. The regression line is shown in red. The inverse relation is due to the high percentage of static energy, which results in relatively constant total energy consumption. An inverse relationship is observed when a constant value is divided by an increasing number.

In contrast, Figure 7.1b shows the correlation between the concurrent users and the TCFP per concurrent users. Here, an inverse relationship can be seen for the upper bound, but there are also data points throughout the entire area below the curve. In other words, the curve can then be used to describe the upper bound. Noted should be that currently only Scope 2 emissions are taken into account. Including the Scope 3 emissions, which are often considered constant, would increase the static part of the TCFP and thus make the relationship more clear.

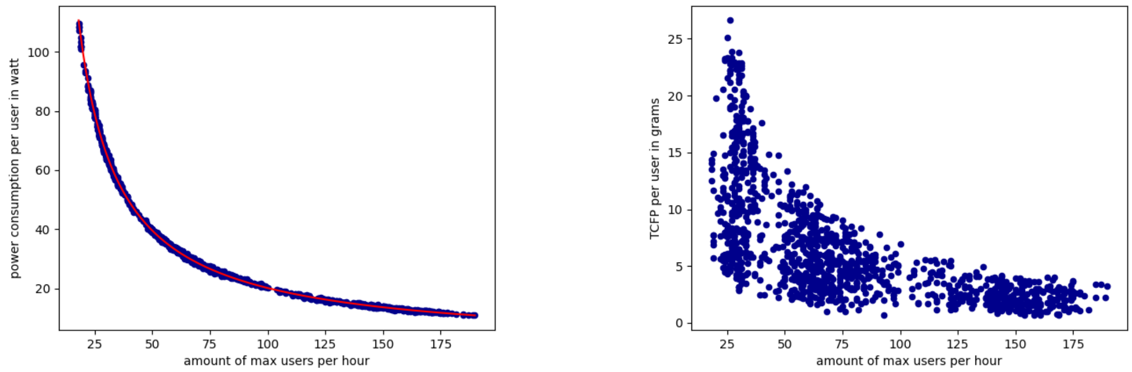
The comparison between different tenants can now be done by comparing different graphs. This is not the most user-friendly approach, and relying on visual inspection to determine which tenant is better can lead to inaccurate results. Furthermore, the difference in the maximum amount of concurrent users makes the comparison more difficult.

In addition, the policy chosen to compare different tenants could favor one tenant over another as a tenant could be only more efficient in either the static or dynamic part. To ensure a fair comparison, we use a γ of 1 and a ζ of 1. Even if the policy distribution is different, this choice provides the fairest comparison.

The linear regression plotted in Figure 7.1a is of the form of Equation 7.1. Where x is the amount of max users, a and b are constants and n is a constant that determines the steepness of the curve. To account for when no users are logged in, it was interpreted as if there was one user logged in.

$$\frac{1}{(a+bx)^n} \tag{7.1}$$

Based on Equation 7.1 three different scores can be calculated. These scores are based on a per-user basis and are therefore called the User Power Efficiency Score (UPES).



(a) Scatterplot for the total power consumption per user vs the max users

(b) Scatterplot for the TCFP per user vs the max users

Figure 7.1: The difference between using carbon intensity or not

The UPES family consists of UPES-P, UPES-A and UPES-N, each providing a different perspective. The combination of these three gives a comprehensive overview. However, it might be preferable to only present one of these values to maintain clarity and user-friendliness. These three scores are based on Equation 7.1

UPES-P UPES-P is based on percentiles. It finds the number of concurrent users for which 95% of the concurrent users is below. This cutoff value is then placed in Equation 7.1, giving the value for which the watt usage per user is 95% of the time lower. An intercept can be seen in Figure 7.2. For a more detailed view, the 50th and 5th percentile can also be provided, if necesary. The reason for using the 95th percentile lies in that this is also commonly used by internet service providers when billing "burstable" internet bandwidth.

UPES-A UPES-A is determined based on the area below the curve. The curve never intersects with both axes, which makes it important to limit both sides. To establish these limits, we define CU_{min}, which is the minimum amount of concurrent users and CU_{max} which is the maximum amount of concurrent users. The area under the curve between CU_{min} and CU_{max} is then calculated to derive the UPES-A score, as shown in Figure 7.3.

Having a higher CU_{max} would result in a larger area under the curve, even if that service would be more efficient. This value is normalized to account for this. This will advocate in favor of having more concurrent users on a service, which is due to the significant part of static energy being a logical choice. The resulting formula, which represents the area between CU_{min} and CU_{max} can be seen in Equation 7.2.

$$UPES - A = \frac{A}{CU_{max} - CU_{min}} \tag{7.2}$$

UPES-N The last metric is UPES-N. The metric depends on the value of n as defined in Equation 7.1. By varying n, different curves can be defined for UPES-N as it controls the steepness of the curve. In Figure 7.4 the curves for the values 1, 1.5 and 2 are displayed, represented in blue, red and green respectively.

With a lower n-value, the curve predicts a lower power usage per user, as can be seen that the blue line is always lower than the green line.

One noteworthy comment is that a lower n-value gives a lower value. The reason for this is that the base is always lower than 1.

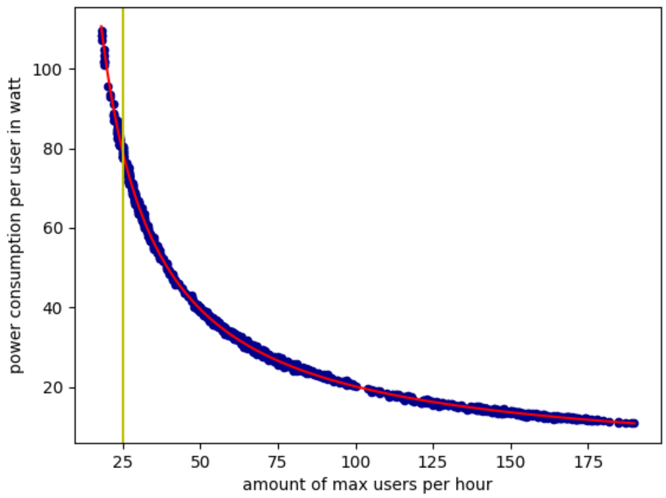


Figure 7.2: In this figure 95% of the time less than 78.90 watts per user, the location is at the intersect

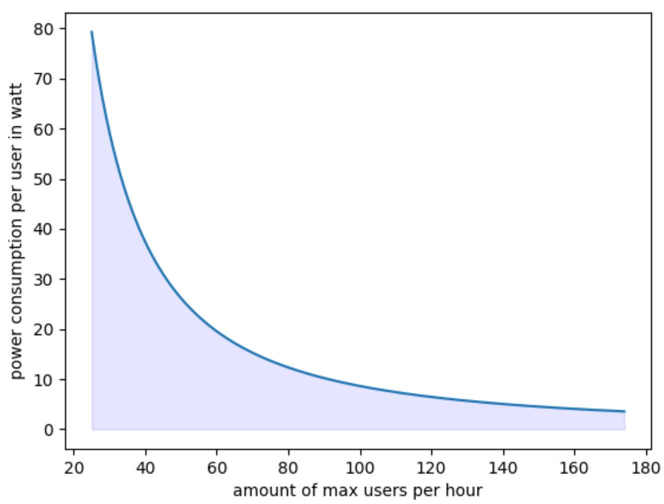


Figure 7.3: Highlighting the area below the regression line

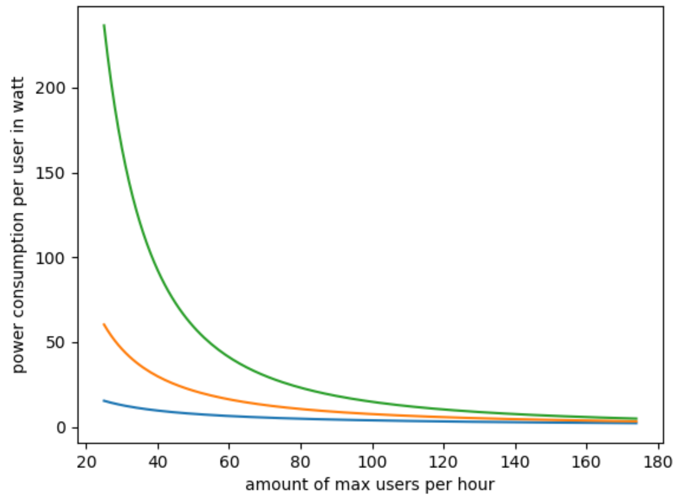


Figure 7.4: Different values of n and their effect on power consumption

Tenant	UPES-P	UPES-A	UPES-N
A	1155.49	26.20	1.05
B	375.43	25.00	1.05
C	974.75	12.49	1
D	18.11	7.20	1.1
E	2619.24	76.84	1.05
F	3536.93	41.64	1

Table 7.1: The UPES scores for the six tenants

7.2 Scoring different tenants

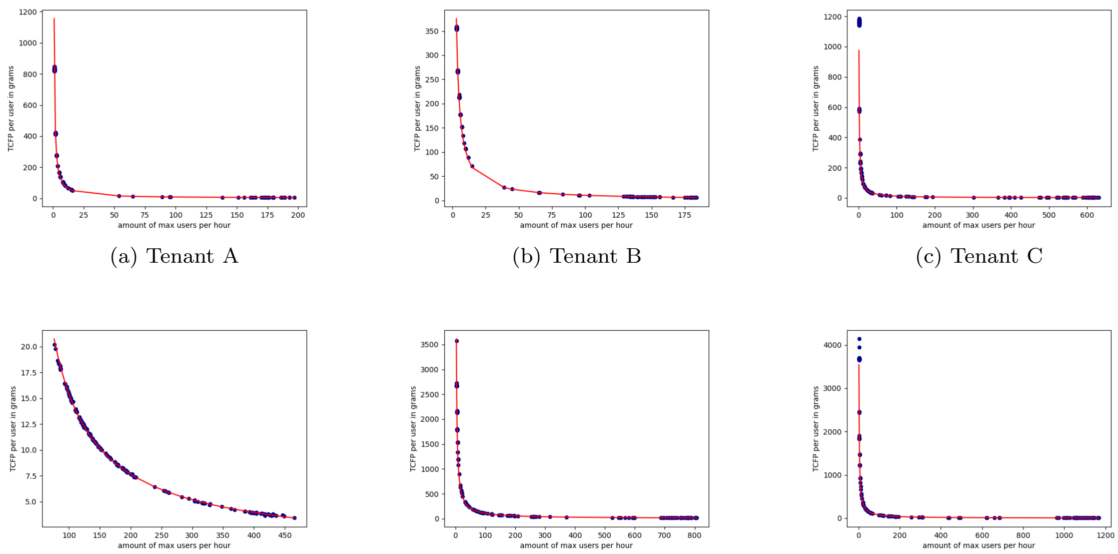
UPES was used to compare different tenants, with the original tenant (M) serving as the basis on which the model was developed and the six tenants functioning as confirmation cases. The six tenants are grouped into three different groups based on the maximum number of concurrent users they have; small, medium and large. The tenants A and B are categorized as small, tenants C and D are categorized as medium and tenants E and F are categorized as large.

The scatterplot of the number of concurrent users versus the power usage per user for all six tenants can be seen in Figure 7.5. They all follow a curve similarly as defined in Equation 7.1.

The UPES scores for the six tenants are presented in Table 7.1. In the next section, an evaluation of the different score types will be given based on the feedback received. However, from these values, some evaluations can already be made.

The n-value of the different tenants is similar, due to the high percentage of static power consumption. Furthermore, it is noteworthy that UPES-A and UPES-P are both not increasing and decreasing at the same time. This can be explained by the different ways the scores are calculated. UPES-A assumes equal distribution of concurrent users. UPES-P does not make that assumption and looks at the 95th percentile. UPES-P also has its limitations. UPES-P does not take into account what happens 5% of the time. Moreover, if exactly 5% of the time the system is inefficient, while the other time it is efficient it will paint the wrong picture. This can be remedied by also providing the 50th and 5th percentile.

This also shows that the size of a customer does not necessarily correlate with energy efficiency. By combining the information provided by BT about the different tenants, certain correlations can be observed. BT can give each tenant a bespoke score, which can be seen in Table 7.2. The bespoke score indicates the extent to which the service was customized to meet their specific requirements.



(a) Tenant A (b) Tenant B (c) Tenant C

(d) Tenant D (e) Tenant E (f) Tenant F

Figure 7.5: The amount of max users versus the power usage per max user for all six tenants

Tenant	bespoke score
A	10%
B	7.5%
C	5%
D	20%
E	50%
F	45%

Table 7.2: The bespoke scores for the six tenants

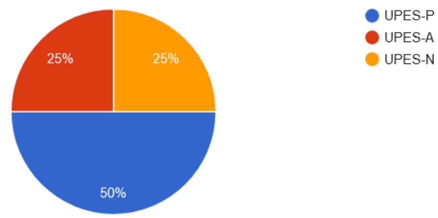


Figure 7.6: Result of which UPES metric the employees liked the most

A and C have a configuration close to the standard configuration of the CCC service. None of them is not the most energy-efficient tenant, but they are around the median.

B also uses a standard configuration and is apart from D the most energy-efficient tenant. Different from A and C, they are energy-efficient as they have a high number of concurrent users during most times.

D has 24/7 service and uses its own voice-enabled routers per country. This eases the workload for CCC. This is directly visible in their UPES score, making it an unfair comparison with the other tenants if this is not taken into account.

E only services one country. They use self-service options in the menu so the call may not have to be handled by an agent. They also rely heavily on speech recognition in comparison with other tenants hosted by BT. Both factors negatively impact their UPES scores.

F only serves the Asia Pacific region and employs multiple custom databases managed by BT. Every call will accrue at least 30 database dips (queries) more than other tenants. Both directly affect their UPES score.

These examples demonstrate that the configuration and usage change their score significantly. A tenant which has the equipment to serve a high amount of concurrent users is only more efficient when used by a lot of concurrent users. However, when the usage is low, the additional equipment required causes the service to become inefficient. This can cause it to be less energy efficient than the smaller tenants. Moreover, large companies also have data centers in other places where the PUE is higher than the smaller companies. Where the data center in London has a PUE of 1.2, for example, Hong Kong has a PUE of 1.85.

7.3 Survey

The survey was sent to different BT employees of which the technical skill widely differed. For example, both engineers and directors answered the survey. In total 12 people answered this survey. Due to the low number, the results should be taken as preliminary results, requiring further in-depth research in the future.

As part of the survey, respondents were tested on their understanding of the metric. Out of the 12, two filled in the wrong answer on a hypothetical scenario for UPES-N. This was the only metric in which people had the incorrect answer. Interestingly, one of these two individuals still expressed a preference for UPES-N, despite selecting the wrong answer.

Apart from these two outliers, the remaining respondents performed well on the questions. They were then asked about their preferred metric. The result, shown in Figure 7.6, showed that UPES-P was seen as the most useful one. This is underlined as respondents stated that: *"Watts per user is easy to understand and compare"*, *"It makes the most sense to me as it is a real value"* and more similar comments.

UPES-N has 25% of the votes. However, one respondent stated that they had chosen this because the lower value would be the least shocking for their client. Moreover, one person who voted for UPES-P stated they did not understand UPES-N. This provides additional evidence that UPES-N is not perceived as the most useful metric.

UPES-A, on the other hand, was received positively by 25%. The people who have chosen this were pleased with the normalization aspect of this metric.

The respondents were then asked whether the chosen metric was clear. The result of this can be seen in Figure 7.7, with a rating scale from 1, perfectly clear, to 5, unclear. On average people gave scored it a 2.5, indicating a moderate level of understanding. Both UPES-P and UPES-A received an average score of 2.33 and UPES-N an average score of 3.

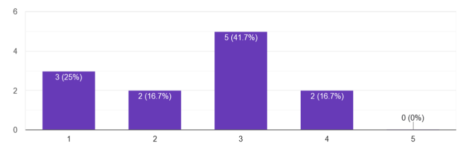


Figure 7.7: Result of how clear the chosen metric was

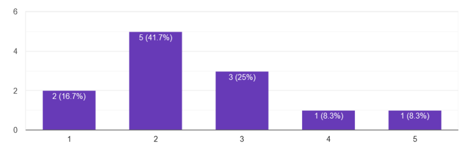


Figure 7.8: Result whether the employees found the metric useful

The respondents were also asked to evaluate the usefulness of the chosen metric. The results are presented in Figure 7.8. On average, they scored the usefulness at 2.5. UPES-A was seen as the most useful with an average of 2 and UPES-N was the least useful with a score of 3. UPES-P scored the same as the overall average, a 2.5.

Lastly, the respondents were asked whether they believed UPES made it easier to compare different tenants. The results can be seen in Figure 7.9. On average it scored 2.42. UPES-P scored best with an average score of 2, UPES-A scored slightly worse with a score of 2.33 and UPES-N scored the lowest with a value of 3.33.

A summary of the averages can be found in Table 7.3.

The survey included the option to leave a final remark. Several of these remarks were questions to clarify specific use cases. Other questions about the impact of different settings and seeking general clarifications to better understand the metric and the applications were also provided.

7.4 Discussion

The survey indicated that UPES is considered useful with room for improvement. It specifically showed that UPES-N performed on all questions worse than the others. Between UPES-P and UPES-A it seems to be mainly a question of preference. However, more people seem to choose UPES-P and UPES-P considers the distribution of concurrent users. For that reason, we propose to use UPES-P to compare different tenants. When a more detailed comparison is needed one can opt for showing all three of the metrics.

Moreover, this metric should not be used in isolation, but in combination with the TCFP. This is best illustrated with a case scenario. If a tenant wishes to increase its UPES score without considering the TCFP, they may log on more users who do not effectively use the system. This will improve the UPES score. However, their TCFP does increase due to the increase of logged-in users. UPES is best suitable to compare different tenants with each other.

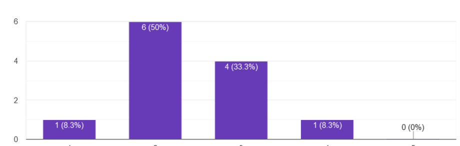


Figure 7.9: Result whether they find UPES easier to compare

	UPES-P	UPES-A	UPES-N	Total
Clarity	2.33	2.33	3	2.5
Usefulness	2.5	2	3	2.5
Ease of use	2	2.33	3.33	2.42

Table 7.3: Average scores for different questions, split into the different metrics

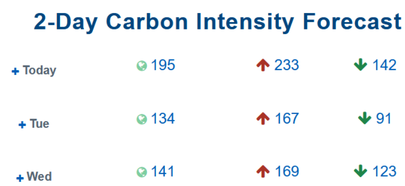


Figure 7.10: 2-day forecast for Great Britain

7.5 Possible improvements for the tenants

This section will give some recommendations for the tenants on how to improve their footprint. These improvements are based on the model and UPES score.

7.5.1 Carbon Intensity

The semi-random carbon intensity values can be used for a good purpose. The main surges of energy such as making a backup of a database can be done during carbon intensity valleys. This has to be done based on a prediction such as shown in Figure 7.10.

Taking the average of these days gives 157 gCo2/KwH. While if the minimum is taken over these two days 91 gCo2/kwh. Which decreases the footprint for this action by over 40%.

7.5.2 Max users

The amount of max users influences the UPES score a lot. Flattening the peaks to fill the valleys will make the biggest difference. This will however not change the TCFP, which might give a skewed result.

However, using the same infrastructure with other tenants will, with this model as its basis, lead to a decrease in the TCFP. This does make the following assumption. That one tenant will have its peaks during the valleys of the other. This way their UPES-P score will improve quite a bit.

7.5.3 Powering down VMs

Another option is to power down VMs when the amount of concurrent users is below a certain threshold, which will result in a lower power consumption. The VMs which are powered down or placed into hibernation mode should be the unused VMs. An example would be when there are zero concurrent users to power down almost all VMs. Despite the decrease in power consumption there is also a risk involved as the availability of the services could be impacted by this. This is similar to the method proposed by Beloglazov and Buyya [52], which should be used as a starting point.

Chapter 8

Conclusion

In this chapter, we will conclude the thesis. A small summary will be provided in Section 8.1. The fulfillment of requirements will be evaluated in Section 8.2 and the research questions will be answered in Section 8.3. The impact of this work will be discussed in Section 8.4. Lastly, future work will be discussed in Section 8.5.

8.1 Summary

In this project, the author worked together with BT to develop a method to allocate emissions more fairly among stakeholders. For this, we focused on one service, CCC. We improved the model from Westerhof to include static and dynamic electricity consumption where static electricity consumption is the electricity consumption when the machine is idle. Dynamic electricity consumption is the electricity consumption for the consumption which is based on the workload. Both parts are given weights which enable us to create policies. These policies have an upper and lower bound to fairly distribute emissions among stakeholders.

Furthermore, we also designed a way to compare different tenants with each other using a metric called UPES scores. This measures how good the service is on a per-user basis. This is split into three different parts UPES-P, UPES-A, and UPES-N each covering a different aspect. UPES-P is based on percentiles and states the value for which 95% of the energy consumption is below. UPES-A is based on the area beneath the curve. UPES-N is the constant used in the formula used to describe the regression line. We then send a survey to BT employees, which showed that UPES-P was the best understood and perceived metric. Moreover, six different tenants were compared with each other showing the strengths and weaknesses of UPES.

8.2 Requirements

For this model and comparison method, a couple of requirements were identified, as discussed in Section 3.4. The functional requirements were all met to various degrees. The TCFP can be calculated with the model satisfying the first functional requirement. TCFP is calculated with the use of different Scopes. We have mainly focused on Scope 2 emissions. Even though this research did not focus on Scope 3 emissions, the model still incorporated this. This functional requirement was therefore also met. The last functional requirement stated that we needed a way to compare different tenants. This was met through the use of UPES.

Moreover, there were also some non-functional requirements. In Chapter 4 we introduced γ and ζ in combination with lower and upper bounds to make accountability more fair for different stakeholders. This satisfied the first non-functional requirement. The only estimate this model uses for Scope 2 emissions is the network consumption. Which is grounded in the literature. The literature stated an average power consumption per byte as well that there should be static and dynamic power usage. This was combined to give more accurate results, thus satisfying the second non-functional requirement. The model is also transparent. The formulas used are clearly explained, which satisfies the third non-functional requirement. Lastly, we described the effects different policies have, this enables us to satisfy the last non-functional requirement.

8.3 Research questions

Before answering the main research question the other research questions will first be answered.

To answer research question 1: As shown in Section 4.2, to give a more fair assessment, the energy consumption calculated for Scope 2 was divided into two parts: static and dynamic. The static part is the energy consumed when there is no workload on a device, while the dynamic part is the energy consumed based on the workload. These parts were allocated among different stakeholders according to the policy outlined in Section 4.3. The policy consists of an upper and lower bound. Where the lower bound has a γ, ζ and LSHARE lower than the upper bound. The stakeholder's responsibility for the emissions falls between the upper and lower bound. Where γ reflects the static part and ζ the dynamic part. In Section 6.3 we demonstrated the influence of different values of γ, ζ and LSHARE. γ adds a constant to the final power usage, while ζ adds a fluctuating component. LSHARE was mostly left for future work, however, we did show that with the current method for calculating Scope 3 emissions, it behaves similarly to static energy usage by adding a constant value.

To answer research question 2: In Figure 7.1a the maximum users were plotted against the power usage. This shows that when more users are logged in, on a per-user basis less energy is consumed. We then looked at it for six other tenants in Figure 7.5, which showed similar results. Section 6.1 showed that this was due to the high amount of static power consumption.

The situation is however somewhat more complex since Section 7.2 showed that for certain peaks more equipment is needed. This raises the total power consumption needed. This means that the most efficient way to use the equipment is the following. The service should have the minimum amount of equipment needed to facilitate the peaks of max users and the service should have the maximum amount of concurrent users as long as possible.

To answer research question 3: The location of the data centers affects the TCFP a lot. Two important factors are the following. We showed in Section 7.2 that PUE makes a huge difference and in Section 6.2 the same is shown for carbon intensity. The carbon intensity was for the pilot study only available for one location. This makes the difference between different data centers hard to measure. However, for just this location the different times when the data center consumed a lot of power made a big difference. This can be seen in Figure 6.3.

To answer the main research question: The different tenants with different usage profiles can be compared with each other using UPES. In Chapter 7 three different scores were proposed UPES-P, UPES-A, UPES-N. These are based on an inverse regression between the max users and the energy consumption per max user. All three of these were then used in a survey to determine how easy to use they were for the different tenants.

8.4 Impact and implications

The main impact of this model is the addition of static and dynamic parts. By splitting the energy usage into two parts it can be more fairly distributed between the different stakeholders. A fair distribution is especially important as reporting the emissions becomes more important. This was directly applied with the use of an lower and upper bound. The effect of different values can have be seen in Section 6.4. The use of the lower and upper bound solved the issue when both stakeholders are equally responsible for certain emissions. Now both stakeholders can be held responsible without double-counting.

As showed in Section 6.6 when this model was applied to the BT tenants for the server part the energy usage was available. For the cooling part, the PUE was available. For the network part, we had to make an estimation. The network estimation is more inaccurate than the other two parts because of this. Also, not every service might be able to have this data. This results in that this model is dependent on the available data. Despite having not all data, we showed that it is still possible to get an estimate. With the use of μ the static power usage of network devices can be approximated.

Another impact is the way that services can now be compared with each other with the metric proposed in Chapter 7. With UPES tenants can see how well they perform and compare how efficiently their systems run. UPES-P was best received and makes comparing tenants significantly easier. Moreover, UPES was mostly depended on the static part of the total energy consumption and the number of concurrent users using the service. This underlines the need to focus on decreasing the static part of the power consumption as well as making sure the service is not underutilized, something that is done quite frequently by tenants in this study. An example is during the night, when all the servers are fully operational with little to no users logged in. UPES can therefore be used to further optimize the energy consumption and therefore the TCFP.

8.5 Future work

In this section potential future work will be discussed and the improvements that can be made for this model.

There is a drawback to the model that can be improved, however, this will make the model more complex. Here we split γ and ζ into multiple subgroups. Namely, γ_{server}, $\gamma_{network}$, $\gamma_{cooling}$, ζ_{server}, $\zeta_{network}$ and $\zeta_{cooling}$. This gives direct control over the different parts of the energy consumption. The need for this can be explained with a use case.

MyBackup For example there is an imaginary service MyBackup, which makes it possible to backup data. This will require little server power but an extensive amount of network power. Because of them a high amount of static power is needed for the network. Thus, there γ_{server} might be lower than their $\gamma_{network}$.

RenderFarm The imaginary service RenderFarm lets users upload a 3d model and it will create a 3D render. This relies heavily on the server power, but once's uploaded the network is not used anymore. For this case, γ_{server} is higher and $\gamma_{network}$ is lower.

Which effect this will have on the Scope 2 emissions of the tenant and whether this is worth the extra complexity is currently unknown. However, as the different parts can then be assigned to different stakeholders it could make the allocation more fair.

Furthermore, currently, the carbon intensity of London is used throughout the entire pilot study. This makes the results less accurate in other parts of the world. We have shown in Section 5.1 a difference between the Netherlands and London, but could not see how this would affect the carbon emissions of the data centers. In the future, this should be included in the data and show what difference this would make for data centers.

In addition, the network estimate is currently done based on the observed peaks and valleys. This makes the static part less accurate as the true valleys and peaks might be higher and lower depending on the tenant. The dynamic part was made under the assumption that the power usage is linear with the usage of the device. Despite this assumption being common in the literature it is something that was not observed during some preliminary data points. The exact nature of this relationship is therefore unknown and should be researched more.

Lastly, the entire power consumption estimation for network devices is based on a constant from 2018 [90]. This might not be an accurate depiction of the machines currently used. Creating a new constant or improving the way the static and dynamic parts of the network model can be calculated should thus be researched.

Bibliography

[1] A. S. G. Andrae and T. Edler, "On global electricity usage of communication technology: Trends to 2030," *Challenges 2015, Vol. 6, Pages 117-157*, vol. 6, pp. 117–157, 1 Apr. 2015, ISSN: 2078-1547. DOI: 10.3390/CHALLE6010117.

[2] M. Koot and F. Wijnhoven, "Usage impact on data center electricity needs: A system dynamic forecasting model," *Applied Energy*, vol. 291, p. 116 798, Jun. 2021, ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2021.116798.

[3] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, pp. 984–986, 6481 Feb. 2020, ISSN: 10959203. DOI: 10.1126/SCIENCE.ABA3758.

[4] L. Hilty, "Why energy efficiency is not sufficient – some remarks on "green by it"," Jul. 2012. DOI: 10.13140/2.1.4121.6968.

[5] D. R. Williams and Y. Tang, "Methodology to model the energy and greenhouse gas emissions of electronic software distributions," *Environmental Science and Technology*, vol. 46, pp. 1087–1095, 2 Jan. 2012, ISSN: 0013936X. DOI: 10.1021/ES202125J.

[6] B. Whitehead, D. Andrews, A. Shah, and G. Maidment, "Assessing the environmental impact of data centres part 1: Background, energy use and metrics," *Building and Environment*, vol. 82, pp. 151–159, Dec. 2014, ISSN: 0360-1323. DOI: 10.1016/J.BUILDENV.2014.08.021.

[7] C. C. Pertsova, *Ecological Economics Research Trends*. 2007.

[8] K. Fang, R. Heijungs, and G. R. de Snoo, "Theoretical exploration for the combination of the ecological, energy, carbon, and water footprints: Overview of a footprint family," *Ecological Indicators*, vol. 36, pp. 508–518, Jan. 2014, ISSN: 1470160X. DOI: 10.1016/j.ecolind.2013.08.017.

[9] W. WRI, "The greenhouse gas protocol: A corporate accounting and reporting standard," *World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD), Washington, DC and Geneva*, 2004.

[10] ISO. "Iso 14064-1:2018(en), greenhouse gases — part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals." (2018), [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:14064:-1:ed-2:v1:en (visited on 04/01/2023).

[11] M. Sotos, "Ghg protocol scope 2 guidance," *World Resources In-stitute and World Business Council for Sustainable Developmen*, 2015. [Online]. Available: www.ghgprotocol.org..

[12] J. Downie and W. Stubbs, "Evaluation of australian companies' scope 3 greenhouse gas emissions assessments," *Journal of Cleaner Production*, vol. 56, pp. 156–163, Oct. 2013, ISSN: 0959-6526. DOI: 10.1016/J.JCLEPRO.2011.09.010.

[13] P. Bhatia, C. Cummis, A. Brown, D. Rich, L. Draucker, and H. Lahd, "Corporate value chain (scope 3) accounting and reporting standard supplement to the ghg protocol corporate accounting and reporting standard ghg protocol team," *World Resources Institute and World BusinessCouncil for Sustainable Developmen*, 2011.

[14] M. Lenzen and J. Murray, "Conceptualising environmental responsibility," *Ecological Economics*, vol. 70, pp. 261–270, 2 Dec. 2010, ISSN: 0921-8009. DOI: 10.1016/J.ECOLECON.2010.04.005.

[15] D. Biza, "The carbon reduction opportunity of moving to amazon web services," *Black and White paper by 451 research*, 2019.

[16] "The carbon benefits of cloud computing," *Report by Micrisoft in partnership with WSP*, 2020.

[17] "Carbon footprint reporting methodology — carbon footprint documentation — google cloud." (2023), [Online]. Available: `https://cloud.google.com/carbon-footprint/docs/methodology` (visited on 03/20/2023).

[18] Armbrust *et al.*, "Above the clouds: A berkeley view of cloud computing," Feb. 2009.

[19] J. Baliga, R. W. Ayre, K. Hinton, and R. S. Tucker, "Green cloud computing: Balancing energy in processing, storage, and transport," *Proceedings of the IEEE*, vol. 99, pp. 149–167, 1 Jan. 2011, ISSN: 15582256. DOI: `10.1109/JPROC.2010.2060451`.

[20] I. Konstantinou, E. Floros, and N. Koziris, "Public vs private cloud usage costs: The stratuslab case," *Proceedings of the 2nd International Workshop on Cloud Computing Platforms, CloudCP 2012 - Co-located with EuroSys 2012*, 2012. DOI: `10.1145/2168697.2168700`.

[21] J. Dantas, R. Matos, J. Araujo, and P. Maciel, "Eucalyptus-based private clouds: Availability modeling and comparison to the cost of a public cloud," *Computing*, vol. 97, pp. 1121–1140, 11 Nov. 2015, ISSN: 0010485X. DOI: `10.1007/S00607-015-0447-8/FIGURES/10`.

[22] R. Westerhof, "Exploring the factors of sustainability and creating a sustainability dashboard for bt global services," *RuG Internship Report*, Jul. 2021.

[23] R. Westerhof, "A model for the total carbon footprint of a cloud-based software as a service provider and its customers," *RuG Master thesis*, 2022.

[24] "Renewable energy directive." (), [Online]. Available: `https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-directive_en` (visited on 03/23/2023).

[25] C. Belady, A. Rawson, A. M. D. Pfleuger, and D. Cader, "Green grid data center power efficiency metric: Pue and dcie," *The Green Grid*, May 2008.

[26] M. Avgerinou, P. Bertoldi, and L. Castellazzi, "Trends in data centre energy consumption under the european code of conduct for data centre energy efficiency," *Energies*, vol. 10, 10 2017, ISSN: 19961073. DOI: `10.3390/EN10101470`.

[27] G. Brady, N. Kapur, J. Summers, and H. Thompson, "A case study and critical assessment in calculating power usage effectiveness for a data centre," 2013, ISSN: 0196-8904. DOI: `10.1016/j.enconman.2013.07.035`.

[28] P. Mell and T. Grance, "The nist definition of cloud computing recommendations of the national institute of standards and technology," *National Institute of Standards and Technology Special Publication 800-145*, 2011. DOI: `10.6028/NIST.SP.800-145`.

[29] M. Badger, T. Grance, R. Patt-Corner, and J. Voas, "Cloud computing synopsis and recommendations," May 2012. DOI: `10.6028/NIST.SP.800-146`.

[30] G. S. Foundation. "Software carbon intensity (sci) specification." (2022), [Online]. Available: `https://github.com/Green-Software-Foundation/sci/blob/main/Software_Carbon_Intensity/Software_Carbon_Intensity_Specification.md` (visited on 02/14/2023).

[31] financeec. "Corporate sustainability due diligence." (2022), [Online]. Available: `https://commission.europa.eu/business-economy-euro/doing-business-eu/corporate-sustainability-due-diligence_en` (visited on 03/08/2023).

[32] financeec. "Corporate sustainability reporting." (2022), [Online]. Available: `https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en` (visited on 03/01/2023).

[33] Cleer. "Cleer protocol." (2019), [Online]. Available: `https://www.climatelinks.org/resources/cleer-protocol` (visited on 04/21/2023).

[34] sciencebasedtargets, "Sbti financial sector and tcfd reporting guidance," 2023.

[35] D. Mytton, "Hiding greenhouse gas emissions in the cloud," *Nature Climate Change*, vol. 10, p. 701, 8 2020, ISSN: 1758-6798. DOI: 10.1038/s41558-020-0837-6.

[36] H. S. Matthews, C. T. Hendrickson, and C. L. Weber, "The importance of carbon footprint estimation boundaries," *Environmental Science & Technology*, vol. 42, pp. 5839–5842, 16 Aug. 2008, ISSN: 0013-936X. DOI: 10.1021/es703112w.

[37] A. Armstrong and A. Souarez, "Ict sector guidance built on the ghg protocol product life cycle accounting and reporting standard chapter 4:guide for assessing ghg emissions of cloud computing and data center services," 2017. [Online]. Available: http://www.ghgprotocol.org/standards/product-standard.

[38] J. Reap, F. Roman, S. Duncan, and B. Bras, "A survey of unresolved problems in life cycle assessment. part 1: Goal and scope and inventory analysis," *International Journal of Life Cycle Assessment*, vol. 13, pp. 290–300, 4 Jun. 2008, ISSN: 09483349. DOI: 10.1007/S11367-008-0008-X/FIGURES/2.

[39] ISO. "Iso 14044:2006(en), environmental management — life cycle assessment — requirements and guidelines." (2016), [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:14044:ed-1:v1:en (visited on 04/01/2023).

[40] R. H. Crawford, P. A. Bontinck, A. Stephan, T. Wiedmann, and M. Yu, "Hybrid life cycle inventory methods – a review," *Journal of Cleaner Production*, vol. 172, pp. 1273–1288, Jan. 2018, ISSN: 0959-6526. DOI: 10.1016/J.JCLEPRO.2017.10.176.

[41] S. Suh and G. Huppes, "Methods for life cycle inventory of a product," *Journal of Cleaner Production*, vol. 13, pp. 687–697, 7 Jun. 2005, ISSN: 0959-6526. DOI: 10.1016/J.JCLEPRO.2003.04.001.

[42] M. Lenzen, J. Murray, F. Sack, and T. Wiedmann, "Shared producer and consumer responsibility - theory and practice," *Ecological Economics*, vol. 61, pp. 27–42, 1 Feb. 2007, ISSN: 09218009. DOI: 10.1016/J.ECOLECON.2006.05.018.

[43] W. W. Leontief, "Quantitative input and output relations in the economic systems of the united states," *The Review of Economics and Statistics*, vol. 18, p. 105, 3 Aug. 1936, ISSN: 00346535. DOI: 10.2307/1927837.

[44] Y. Yang, "Does hybrid lca with a complete system boundary yield adequate results for product promotion?" *International Journal of Life Cycle Assessment*, vol. 22, pp. 456–460, 3 Mar. 2017, ISSN: 16147502. DOI: 10.1007/S11367-016-1256-9/FIGURES/1.

[45] F. Pomponi and M. Lenzen, "Hybrid life cycle assessment (lca) will likely yield more accurate results than process-based lca," *Journal of Cleaner Production*, vol. 176, pp. 210–215, Mar. 2018, ISSN: 09596526. DOI: 10.1016/J.JCLEPRO.2017.12.119.

[46] M. Lenzen and M. L. ISA, "Double-counting in life cycle calculations," *Journal of Industrial Ecology*, vol. 12, pp. 583–599, 4 Aug. 2008, ISSN: 1530-9290. DOI: 10.1111/J.1530-9290.2008.00067.X.

[47] M. Lenzen, "Dealing with double-counting in tiered hybrid life-cycle inventories: A few comments," *Journal of Cleaner Production*, vol. 17, pp. 1382–1384, 15 Oct. 2009, ISSN: 0959-6526. DOI: 10.1016/J.JCLEPRO.2009.03.005.

[48] F. Caro, C. J. Corbett, T. Tan, and R. Zuidwijk, "Double counting in supply chain carbon footprinting," *Manufacturing & Service Operations Management*, vol. 15, pp. 545–558, 4 Oct. 2013, ISSN: 1523-4614. DOI: 10.1287/msom.2013.0443.

[49] A. Marques, J. Rodrigues, and T. Domingos, "International trade and the geographical separation between income and enabled carbon emissions," *Ecological Economics*, vol. 89, pp. 162–169, May 2013, ISSN: 0921-8009. DOI: 10.1016/J.ECOLECON.2013.02.020.

[50] M. Aggar *et al.*, "Data centre life cycle assessment guidelines," 2012.

[51] B. Whitehead, D. Andrews, and A. Shah, "The life cycle assessment of a uk data centre," *International Journal of Life Cycle Assessment*, vol. 20, pp. 332–349, 3 Mar. 2015, ISSN: 16147502. DOI: 10.1007/S11367-014-0838-7/FIGURES/14.

[52] A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers," *Concurrency and Computation: Practice and Experience*, vol. 24, pp. 1397–1420, 13 2012. DOI: https://doi.org/10.1002/cpe.1867.

[53] Y. C. Lee, A. Y. Zomaya, Y. C. Lee, and ·. A. Y. Zomaya, "Energy efficient utilization of resources in cloud computing systems," *J Supercomput*, vol. 60, pp. 268–280, 2012. DOI: 10.1007/s11227-010-0421-3.

[54] D. Lo, L. Cheng, R. Govindaraju, P. Ranganathan, and C. Kozyrakis, "Heracles: Improving resource efficiency at scale," 2015. DOI: 10.1145/2749469.2749475.

[55] A. Khosravi, S. K. Garg, and R. Buyya, "Energy and carbon-efficient placement of virtual machines in distributed cloud data centers," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8097 LNCS, pp. 317–328, 2013, ISSN: 03029743. DOI: 10.1007/978-3-642-40047-6_33/COVER.

[56] A. Khosravi, L. L. H. Andrew, S. Member, and R. Buyya, "Dynamic vm placement method for minimizing energy and carbon cost in geographically distributed cloud data centers," 2017. DOI: 10.1109/TSUSC.2017.2709980.

[57] J. Krzywda, A. Ali-Eldin, T. E. Carlson, P.-O. Östberg, and E. Elmroth, "Power-performance tradeoffs in data center servers: Dvfs, cpu pinning, horizontal, and vertical scaling," Mar. 2019. DOI: 10.1016/j.future.2017.10.044.

[58] A. Thakkar, K. Chaudhari, and M. Shah, "A comprehensive survey on energy-efficient power management techniques," *Procedia Computer Science*, vol. 167, pp. 1189–1199, Jan. 2020, ISSN: 1877-0509. DOI: 10.1016/J.PROCS.2020.03.432.

[59] S. Bharany *et al.*, "A systematic survey on energy-efficient techniques in sustainable cloud computing," *Sustainability 2022, Vol. 14, Page 6256*, vol. 14, p. 6256, 10 May 2022, ISSN: 2071-1050. DOI: 10.3390/SU14106256.

[60] S. S. Gill and R. Buyya, "A taxonomy and future directions for sustainable cloud computing: 360 degree view," *ACM Computing Surveys*, vol. 51, 5 Aug. 2018, ISSN: 15577341. DOI: 10.1145/3241038.

[61] T. Mukherjee, A. Banerjee, G. Varsamopoulos, S. K. Gupta, and S. Rungta, "Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers," *Computer Networks*, vol. 53, pp. 2888–2904, 17 Dec. 2009, ISSN: 1389-1286. DOI: 10.1016/J.COMNET.2009.06.008.

[62] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, pp. 33–37, 12 2007, ISSN: 00189162. DOI: 10.1109/MC.2007.443.

[63] A. Vasan, A. Sivasubramaniam, V. Shimpi, T. Sivabalan, and R. Subbiah, "Worth their watts? - an empirical study of datacenter servers," pp. 1–10, May 2010. DOI: 10.1109/HPCA.2010.5463056.

[64] U. Wajid *et al.*, "On achieving energy efficiency and reducing co 2 footprint in cloud computing," *IEEE Transactions on Cloud Computing*, vol. 4, pp. 5–6, Feb. 2015. DOI: 10.1109/TCC.2015.2453988.

[65] M. Xu and R. Buyya, "Managing renewable energy and carbon footprint in multi-cloud computing environments," *Journal of Parallel and Distributed Computing*, vol. 135, pp. 191–202, 2020, ISSN: 0743-7315. DOI: https://doi.org/10.1016/j.jpdc.2019.09.015.

[66] A. A. Chien, R. Wolski, and F. Yang, "Zero-carbon cloud: A volatile resource for high-performance computing," *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pp. 1997–2001, 2015. DOI: 10.1109/CIT/IUCC/DASC/PICOM.2015.296.

[67] F. Yang and A. A. Chien, "Large-scale and extreme-scale computing with stranded green power: Opportunities and costs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, pp. 1103–1116, 5 2018. DOI: 10.1109/TPDS.2017.2782677.

[68] J. Zheng, A. A. Chien, and S. Suh, "Mitigating curtailment and carbon emissions through load migration between data centers," *Joule*, vol. 4, pp. 2208–2222, 10 Oct. 2020, ISSN: 25424351. DOI: 10.1016/J.JOULE.2020.08.001.

[69] P. Huang *et al.*, "A review of data centers as prosumers in district energy systems: Renewable energy integration and waste heat reuse for district heating," *Applied Energy*, vol. 258, p. 114 109, Jan. 2020, ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2019.114109.

[70] M. Wahlroos, M. Pärssinen, J. Manner, and S. Syri, "Utilizing data center waste heat in district heating – impacts on energy efficiency and prospects for low-temperature district heating networks," *Energy*, vol. 140, pp. 1228–1238, 2017, ISSN: 03605442. DOI: 10.1016/J.ENERGY.2017.08.078.

[71] M. Wahlroos, M. Pärssinen, S. Rinne, S. Syri, and J. Manner, "Future views on waste heat utilization – case of data centers in northern europe," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1749–1764, Feb. 2018, ISSN: 18790690. DOI: 10.1016/J.RSER.2017.10.058.

[72] G. F. Davies, G. G. Maidment, and R. M. Tozer, "Using data centres for combined heating and cooling: An investigation for london," *Applied Thermal Engineering*, vol. 94, pp. 296–304, Feb. 2016, ISSN: 13594311. DOI: 10.1016/J.APPLTHERMALENG.2015.09.111.

[73] F. Wang *et al.*, "Do 'green' data centres really have zero co2 emissions?" *Sustainable Energy Technologies and Assessments*, vol. 53, p. 102 769, Oct. 2022, ISSN: 2213-1388. DOI: 10.1016/J.SETA.2022.102769.

[74] N. P. Jouppi *et al.*, "In-datacenter performance analysis of a tensor processing unit," 2017. DOI: 10.1145/3079856.3080246.

[75] Y. Shan, Y. Huang, Y. Chen, and Y. Zhang, "Open access to the proceedings of the 13th usenix symposium on operating systems design and implementation is sponsored by usenix. legoos: A disseminated, distributed os for hardware resource disaggregation legoos: A disseminated, distributed os for hardware resource disaggregation," 2018. [Online]. Available: https://www.usenix.org/conference/osdi18/presentation/shan.

[76] P. Patros, J. Spillner, A. V. Papadopoulos, B. Varghese, O. Rana, and S. Dustdar, "Toward sustainable serverless computing," *IEEE Internet Computing*, vol. 25, pp. 42–50, 6 2021, ISSN: 19410131. DOI: 10.1109/MIC.2021.3093105.

[77] C. Möbius, W. Dargie, and A. Schill, "Power consumption estimation models for processors, virtual machines, and servers," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, pp. 1600–1614, 6 2014, ISSN: 10459219. DOI: 10.1109/TPDS.2013.183.

[78] D. Economou, S. Rivoire, C. Kozyrakis, and P. Ranganathan, "Full-system power analysis and modeling for server environments," 2006, ISSN: 0884-7495.

[79] H. Acar, G. I. Alptekin, J.-P. Gelas, and P. Ghodous, "The impact of source code in software on power consumption," *International Journal of Electronic Business Management*, vol. 14, pp. 42–52, 2016.

[80] R. F. da Silva, H. Casanova, A.-C. Orgerie, R. Tanaka, E. Deelman, and F. Suter, "Characterizing, modeling, and accurately simulating power and energy consumption of i/o-intensive scientific workflows," *Journal of Computational Science*, vol. 44, p. 101 157, Jul. 2020, ISSN: 18777503. DOI: 10.1016/j.jocs.2020.101157.

[81] P. Xiao, Z. Hu, D. Liu, G. Yan, and X. Qu, "Virtual machine power measuring technique with bounded error in cloud environments," *Journal of Network and Computer Applications*, vol. 36, pp. 818–828, 2 Mar. 2013, ISSN: 1084-8045. DOI: 10.1016/J.JNCA.2012.12.002.

[82] G. Contreras and M. Martonosi, "Power prediction for intel xscale® processors using performance monitoring unit events," *Proceedings of the 2005 international symposium on Low power electronics and design - ISLPED '05*, p. 221, 2005. DOI: 10.1145/1077603.1077657.

[83] C. Jin, X. Bai, C. Yang, W. Mao, and X. Xu, "A review of power consumption models of servers in data centers," *Applied Energy*, vol. 265, p. 114 806, May 2020, ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2020.114806.

[84] W. Lin, H. Wang, Y. Zhang, D. Qi, J. Z. Wang, and V. Chang, "A cloud server energy consumption measurement system for heterogeneous cloud environments," *Information Sciences*, vol. 468, pp. 47–62, Nov. 2018, ISSN: 0020-0255. DOI: 10.1016/J.INS.2018.08.032.

[85] E. Sommer *et al.* "Cloud jewels: Estimating kwh in the cloud." (Apr. 2020), (visited on 05/16/2023).

[86] T. Baker, Y. Ngoko, R. Tolosana-Calasanz, O. F. Rana, and M. Randles, "Energy efficient cloud computing environment via autonomic meta-director framework," *Proceedings - 2013 6th International Conference on Developments in eSystems Engineering, DeSE 2013*, pp. 198–203, Feb. 2013. DOI: 10.1109/DESE.2013.43.

[87] W. Vereecken *et al.*, "Power consumption in telecommunication networks: Overview and reduction strategies," *IEEE Communications Magazine*, vol. 49, pp. 62–69, 6 Jun. 2011, ISSN: 01636804. DOI: 10.1109/MCOM.2011.5783986.

[88] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, "A power benchmarking framework for network devices.," 2009.

[89] A. Vishwanath, K. Hinton, R. W. Ayre, and R. S. Tucker, "Modeling energy consumption in high-capacity routers and switches," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1524–1532, 8 Aug. 2014, ISSN: 07338716. DOI: 10.1109/JSAC.2014.2335312.

[90] J. Aslan, K. Mayers, J. G. Koomey, and C. France, "Electricity intensity of internet data transmission: Untangling the estimates," *Journal of Industrial Ecology*, vol. 22, pp. 785–798, 4 Aug. 2018, ISSN: 1530-9290. DOI: 10.1111/JIEC.12630.

[91] A. E. Bohra and V. Chaudhary, "Vmeter: Power modelling for virtualized clouds," *Proceedings of the 2010 IEEE International Symposium on Parallel and Distributed Processing, Workshops and Phd Forum, IPDPSW 2010*, 2010. DOI: 10.1109/IPDPSW.2010.5470907.

[92] J. W. Smith, A. Khajeh-Hosseini, J. S. Ward, and I. Sommerville, "Cloudmonitor: Profiling power usage," *Proceedings - 2012 IEEE 5th International Conference on Cloud Computing, CLOUD 2012*, pp. 947–948, 2012. DOI: 10.1109/CLOUD.2012.112.

[93] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," *Proceedings of the 1st ACM Symposium on Cloud Computing, SoCC '10*, pp. 39–50, 2010. DOI: 10.1145/1807128.1807136.

[94] B. Krishnan, H. Amur, A. Gavrilovska, and K. Schwan, "Vm power metering: Feasibility and challenges," *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, pp. 56–60, 3 Jan. 2011, ISSN: 01635999. DOI: 10.1145/1925019.1925031.

[95] Z. Jiang, C. Lu, Y. Cai, Z. Jiang, and C. Ma, "Vpower: Metering power consumption of vm," *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, pp. 483–486, 2013, ISSN: 23270586. DOI: 10.1109/ICSESS.2013.6615354.

[96] Q. Zhang *et al.*, "A survey on data center cooling systems: Technology, power consumption modeling and control strategy optimization," *Journal of Systems Architecture*, vol. 119, p. 102 253, Oct. 2021, ISSN: 1383-7621. DOI: 10.1016/J.SYSARC.2021.102253.

[97] climatiq. "Climatiq compute - automated carbon emissions calculations." (2023), [Online]. Available: https://www.climatiq.io/docs/guides/cloud (visited on 03/01/2023).

[98] SDIAlliance. "Github - sdialliance/carbon-footprint-ssa: A framework for assessing the carbon footprint of server-side software applications (ssa)." (2022), [Online]. Available: https://github.com/SDIAlliance/carbon-footprint-ssa (visited on 03/01/2023).

[99] T. Inc. "Overview - cloud carbon footprint." (2022), [Online]. Available: https://www.cloudcarbonfootprint.org/docs/methodology (visited on 06/20/2023).

[100] Jetraw. "Jetraw calculator." (2023), [Online]. Available: https://www.jetraw.com/data-storage-cost-and-co2-emissions-calculator (visited on 03/01/2023).

[101] Kainos. "Kainos cloud carbon calculator." (2022), [Online]. Available: https://get.kainos.com/Carbon-Calculator.html (visited on 03/01/2023).

[102] C. Intensity. "Carbon intensity." (2023), [Online]. Available: https://carbonintensity.org.uk/ (visited on 05/11/2023).

[103] A. R. W. Bruce, L. Ruffa, J. Kellowaya, F. MacMillana, and A. Rogersb, "Carbon intensity methodology regional carbon intensity," *White paper national grid ESO*, Mar. 2020, ISSN: 03014215. DOI: 10.1016/j.enpol.2016.12.037.

[104] B. Tranberg, O. Corradi, B. Lajoie, T. Gibon, I. Staffell, and G. B. Andresen, "Real-time carbon accounting method for the european electricity markets," Dec. 2018. DOI: 10.1016/j.esr.2019.100367.

[105] U. Gupta *et al.*, "Chasing carbon: The elusive environmental footprint of computing," *CoRR*, vol. abs/2011.02839, 2020.

[106] L. Belkhir and A. Elmeligi, "Assessing ict global emissions footprint: Trends to 2040 & recommendations," *Journal of Cleaner Production*, vol. 177, pp. 448–463, Mar. 2018, ISSN: 0959-6526. DOI: 10.1016/J.JCLEPRO.2017.12.239.

[107] S. K. Garg, S. Versteeg, and R. Buyya, "Smicloud: A framework for comparing and ranking cloud services," *Proceedings - 2011 4th IEEE International Conference on Utility and Cloud Computing, UCC 2011*, pp. 210–218, 2011. DOI: 10.1109/UCC.2011.36.

[108] J. Dietrich, J. Doolittle, G. Verdun, D. Shahid, D. Reiner, and H. Wong, "Server energy efficiency in data centers and offices the case for adopting a new, active efficiency metric over idle power editor(s)," *The Green Grid*, Dec. 2017.

[109] P. Steenhof *et al.*, "A protocol for quantifying the carbon reductions achieved through the provision of low or zero carbon ict services," *Sustainable Computing: Informatics and Systems*, vol. 2, pp. 23–32, 1 Mar. 2012, ISSN: 2210-5379. DOI: 10.1016/J.SUSCOM.2012.01.001.

[110] P. Steenhof, C. Weber, D. Aikema, R. Robison, R. Simmonds, and C. Kiddle, *Advances in cloud computing research Chapter 7 Estimating emission reductions from low carbon information technology: The GEOCHRONOS relocation project*, M. Ramachandran, Ed. Nova Science Publishers, 2014, ISBN: 9781631171925.

[111] A. Katal, S. Dahiya, and T. Choudhury, "Energy efficiency in cloud computing data centers: A survey on software technologies," *Cluster Computing*, pp. 1–31, Aug. 2022, ISSN: 15737543. DOI: 10.1007/S10586-022-03713-0/FIGURES/3.

[112] E. Jagroep, van der Jan Martijn Werf, J. Broekman, L. Blom, van Rob Vliet, and S. Brinkkemper, "A resource utilization score for software energy consumption," *Proceedings of ICT for Sustainability 2016*, pp. 19–28, 2016, ISSN: 2352-538X. DOI: 10.2991/ict4s-16.2016.3.

[113] E. Jagroep, J. M. van der Werf, S. Brinkkemper, L. Blom, and R. van Vliet, "Extending software architecture views with an energy consumption perspective: A case study on resource consumption of enterprise software," *Computing*, vol. 99, pp. 553–573, 6 Jun. 2017, ISSN: 0010485X. DOI: 10.1007/S00607-016-0502-0/FIGURES/7.

[114] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Software Engineering*, vol. 14, pp. 131–164, 2 Apr. 2009, ISSN: 13823256. DOI: 10.1007/S10664-008-9102-8/TABLES/27.