Can P2P Help the Cloud Go Green?

Christopher Jarabek, Mea Wang
Department of Computer Science
University of Calgary
{*cjjarabe, meawang*}*@ucalgary.ca*

Abstract

The demand for cloud services is growing at a phenomenal rate, and so is the energy cost of the data centres powering those services. This is pressing cloud service providers to look for ways of reducing energy consumption. One approach is to utilize energy-efficient hardware and/or software in the data centres, the other approach is to relocate some services, e.g., *personal files and data rendering, to end-host computers, a.k.a.* peers. *In the later approach, peers contribute their communication and computation resources to exchange data and provide services, while the data centre performs central administration and authentication, as well as backend processing. In this paper, we model the energy consumption for both approaches and then perform analytical studies. Our analysis shows that (1) making the data centre energy efficient can reduce the energy cost significantly; (2) the number of hops from the data centre to the peers and among peers directly influences the energy saving; (3) it is preferred to utilize peers that are already online for other purposes; (4) introducing content delivery network (CDN) servers and enabling proxy service on home modems are the keys to make a hybrid P2P-cloud network go green. We further verified these findings by simulation.*

1 Introduction

In recent years, we have witnessed the enormous increase in cloud computing, exemplified by Amazon Elastic Compute Cloud (EC2), Google AppEngine, and Microsofts Azure. This is largely due to the demand for reducing IT expenditures by moving from self-owned IT resources to the computation-as-a-service model offered by cloud computing providers. While this shift can often result in cost savings for businesses, the costs for powering the data centres can be extreme. Currently, data centres consume 1.5% of all the electricity generated in the United States [2]. It is important to note that this measurement does not include the energy consumed by the Internet, the one large entity that connects and delivers cloud computing services to customers. A 2008 study found that the energy consumption of the Internet (routers, switches, etc.) in a single year is nearly 74 Terawatt-hours or 2% of the energy consumed by the entire planet [9]. In addition, home and office computers that connect to the data centres and the Internet consume roughly 16 Terawatt-hours of electricity in 2007 [12]. The Uptime Institute, a global data centre authority, recently surveyed 525 data centre owners and operators, with 71% based in North America, and reported that 36% worried their data centre facility would run out of power, cooling and/or space in 2011-2012. The authority predicted that by the year 2020, the carbon footprint of data centres alone will eclipse that of the airline industry [4].

The increases in energy cost and the demand for cloud computing urges the cloud service providers to look for ways to reduce energy usage in the years to come. Currently, cloud services still primarily resort to the client/server (C/S) model. In contrast to the traditional C/S model, the data centre is empowered by a large collection of servers instead of a single server or a cluster. Moreover, the service provider may setup data centres at different geographical locations for communication efficiency and service reliability. Such a C/S setup is attractive to business and system administrators due to easy procurement, maintenance, and administration. Nonetheless, this setup still inheres the two main drawbacks of C/S model: single-point-of-failure and lack of scalability. Though large and distributed data centres reduce the severeness of these drawbacks, their constraint on cloud business can be profound, *e.g.*, Amazon EC2 experienced an outage lasted about four days in April 2011, Gmail suffered a massive outage in late February 2011, and Microsoft's Windows Live Hotmail cloud-based e-mail had an out-

978-1-4673-0012-4/11/$26.00 ©2011 IEEE

age for more than four days from December 2010 to January 2011. Though, these outages do not prevent the emergence of the cloud services, they raise concerns regarding energy efficiency and scalability.

Although reducing energy consumed by data centres can effectively reduced the overall cost, this will ultimately limit the coverage and scalability of the service provisioning. Alternatively, the data centres can be revamped by relocating some services to end-host computers, a.k.a. *peers*. Peers contribute their communication, storage and computation resources to exchange data and provide services, while the data centre performs central administration and authentication, as well as backend processing. The Peer-to-Peer (P2P) network, formed by the peers, offers greater flexibility and scalability in service delivery. However, distributed ownership, hardware heterogeneity, lack of centralized management, and loss of privacy are just some of the issues with which cloud providers do not want to contend. Therefore, P2P services can assist and enhance cloud services, rather than a replace current cloud services.

In this paper, we will study the energy efficiency of the two aforementioned energy-saving approaches. We will first present the energy consumption model. Given the model, it is obvious that utilizing energy-efficient hardware and/or software can significantly reduce the energy demand in the data centres. However, we argue that the rapid growth of cloud services and demand will still outrun these savings. This leads us to focus on the second approach, utilizing a hybrid P2P-Cloud approach. Our goal is to answer the question, "Can P2P help the cloud go green?", *i.e.*, to determine if the total energy consumption of the clients, routers and servers used to power a P2P-cloud, is less than a conventional, data centre driven cloud offering. Our analysis shows that the P2P infrastructure will actually lead to substantial increase in network energy consumption. However, it can help the cloud go green under certain circumstances. Our simulation results also agree with the analytical results in principle.

The rest of this paper is organized as follows. Sec. 2 provides an overview of existing literature on hybrid P2P-Cloud network architectures and energy efficiency analysis. In Sec. 3, we will present our energy consumption model, followed by the analysis of this model in Sec. 4. Sec. 5 verifies the energy model in a simulated hybrid P2P-cloud network. Sec. 6 will conclude this paper.

2 Related Work

The P2P infrastructure was originally adopted to overcome the scalability issue in the traditional C/S setup. Since its introduction, P2P has demonstrated its power though popular file sharing applications like BitTorrent and Emule as well as multimedia streaming applications like PPLive and Joost. The emergence of P2P infrastructure in cloud computing drew research attention only recently, although a successful implementation has already been deployed by [5] in 2008. Wuala provides an online storage system that allows users to backup, share, and access files from anywhere. Every new user is provided with 1GB of storage, and can either buy more storage or trade their hard disk space for more online storage. Wuala organizes nodes into three groups: super nodes, storage nodes, and client nodes. The super nodes keep track of the location of data. The storage nodes provide the initial storage and additional storage for purchase. The super nodes and the storage nodes are hosted by the Wuala data centre. The client nodes are the participating user computers (peers) that can host, publish, and retrieve files. Each group of nodes are organized into a ring structure governed by Chord, a well-known P2P distributed hash table (DHT) based protocol for data and node indexing. To date, Wuala is hosting over 566 million files and provides 99.9% guarantee of file preservation. In contrast to cloud offerings provided by Google, Microsoft, and Amazon, Wuala achieves good data redundancy, while reducing the demand for storage capacity at the data centres, a successful demonstration of hybrid P2P-Cloud systems.

Similarity, Clostera [10] provides storage utilizing both a monolithic data centre and a distributed peer network. Here, desktop computers are used to offset the storage requirements of the data centre. Cooperative Peer assists and Multicast [11] is another example of the hybrid architecture. The key difference being the focus on providing video-on-demand as opposed to cloud storage.

It is obvious such a hybrid design could alleviate the computation and communication demand at the data centre, which will in turn result in energy savings. To this end, Valancius et al. [14] focused on the energy cost implications of hybrid architectures. The authors show that incorporating storage into near-edge gateways such as home modems can result in 20% to 30% of energy saving over the C/S data centre architecture. Nedevschi et al. [13] presented one of the first comparisons on energy efficiency between C/S and P2P infrastructures. The authors considered a network with three types of entities: servers, router, and peers. In a C/S cloud system, the servers at the data centres will communicate directly with the peers via a set of routers. The energy cost encompasses only the energy consumed by the servers and the routers, not the peers. In a P2P system, peers are communicating with each other via a set of routers. The energy cost encompasses only the energy consumed by the peers and the routers, not the servers. The authors concluded that P2P systems could be more than five times as efficient as C/S systems when the cost of the baseline energy consumption of the network routers was ignored. However, when this cost was included, P2P systems were almost half as efficient as a comparable C/S architecture. We noted that peers and routers are not under the direct control of the cloud service provider. It would be interesting to study their impacts on the total energy cost. Moreover, since it is clear that P2P cannot be a sole replacement for a C/S-based cloud service, what is the energy cost for a P2P-Cloud system?

After reviewing existing energy analyses, we still cannot find answers to the following questions, which all boil down to the single question "Can P2P help the cloud go green?"

- Peers and routers are not under the direct control of the cloud service provider. What are their impacts on the total energy cost?
- Since it is clear that P2P cannot be a replacement for the C/S Cloud service, what is the energy cost for a hybrid P2P-Cloud system?
- If a hybrid P2P-Cloud system can save energy, how feasible is it?

3 Energy Model

In this paper, our goal is to provide a general view and a fair comparison of the energy consumed by a P2P-Cloud system, and the energy consumption of each of the network entities. To do so, we designed a series of models and perform an analysis from the network perspective, which is more objective and complete than the point of view of a cloud service provider, an ISP, or a user. Since the network implications of a storage cloud are more significant than for a distributed computing cloud like SETI@Home [3], we resort to a discussion of a cloud storage service such as Amazon's Simple Storage Service (S3) [1].

We model the P2P-Cloud system with three entity groups, as shown in Fig. 1: servers (S_i), routers (R_i), and peers (P_i). The *servers* may be located at the same data centre or different data centres. The physical location of the servers does not matter in our analysis since we will be using the average Internet hop count for distances between two network nodes. A *router* can be part of the cloud, *i.e.*, owned by the cloud provider, or outside the cloud. We assume that the router will behave similarly regardless of the owner. A *peer* accesses the Internet via a router. In this P2P-Cloud, any two nodes can communicate so long as there exists an intermediate path of routers between them. For example, the *path* from P_2 to P_3 is $P_2 \to R_5 \to R_1 \to R_3 \to R_6 \to P_3$, and the path has 5 *hops*, which equals to the number of links along this path.

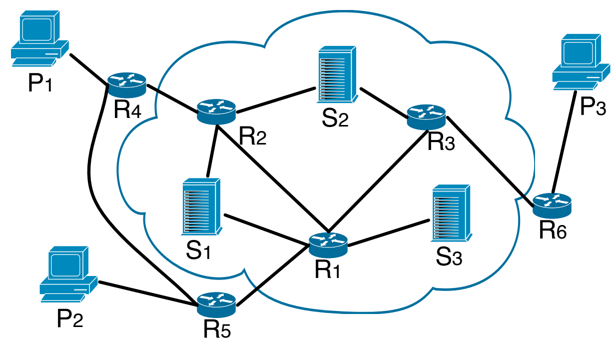


Figure 1. A Hybrid P2P-Cloud System.

Our energy model is inspired by the model in [13]. For each network node, we consider two energy measurements: baseline energy consumption γ and energy consumed per work unit δ. The baseline energy is the energy consumed to keep the device on, *i.e.*, keeping the system running with all necessary applications and services for normal operation in the absence of any requests from clients. On the servers, a multiplicative overhead c is considered for energy consumed by non-computing hardware, most notably, cooling. On the routers and the peers, a multiplicative overhead ω is considered for overhead introduced by P2P communication protocols. Without loss of generality, we assume that devices within the same entity group have similar hardware configurations. Let d_p and d_s denote the average number of hops between peers and from client to server, respectively, and B be the number of bits transmitted by a network interface. The symbols are summarized in Table 3, and their initial values are adapted from [13].

A P2P-Cloud system involves both a data centre-based network and a P2P network. As already characterized in [13], the energy consumed by a cloud network is $c(\delta_s + \gamma_s)B + d_s\delta_r B$, and the energy consumed by a P2P network is $\omega_p\delta_p B + \omega_r d_p \delta_r B$. For a P2P-Cloud system, we introduce a weight variable $n \in [0, 1]$ specifying the proportion of data that is flowing through

	Meaning	Initial Value
c	Server overhead	2
δ_s	Server energy used per unit	$5.2 \times 10^{-8}(J/b)$
γ_s	Server baseline energy	$6.7 \times 10^{-7}(J/b)$
δ_r	Router energy used per unit	$8.0 \times 10^{-9}(J/b)$
γ_r	Router baseline energy	$1.5 \times 10^{-7}(J/b)$
δ_p	Peer energy used per unit	$1.6 \times 10^{-7}(J/b)$
γ_p	Peer baseline energy	$5.8 \times 10^{-6}(J/b)$
ω_p	Peer overhead for P2P	2
ω_r	Router overhead for P2P	2
d_p	Avg. # hops between peers	15
d_s	Avg. # hops (client to server)	13
B	Number of bits transmitted	N/A (b)

Table 1. Description and starting values for model parameters.

the P2P network. Hence, the total energy consumed by a P2P-Cloud can be characterized by Eqn. 1.

$$E_{h1} = ((1-n)(c(\delta_s + \gamma_s)B + d_s\delta_r B) + n(\omega_p\delta_p B + \omega_r d_p \delta_r B)) \quad (1)$$

When $n = 0$, Eqn. 1 reduces to the energy consumed by a strictly cloud network. When $n = 1$, Eqn. 1 reduces to the energy consumed by a strictly P2P network. Since these two extreme cases have been extensively studied in [13], we will focus on $0 \leq n \leq 1$ and use these two cases as reference points. Note that Eqn. 1 ignores the baseline energy of the routers and the peers, which means that the routers and peers are already powered on for other purposes. However, since this is not always the case, we introduce Eqn. 2 to include router baseline energy for each hop along the network path for both the cloud network and the P2P network.

$$E_{h2} = ((1-n)(c(\delta_s + \gamma_s)B + d_s(\delta_r + \gamma_r)B) + n(\omega_p\delta_p B + \omega_r d_p(\delta_r + \gamma_r)B)) \quad (2)$$

One of the consistent assumptions surrounding hybrid network architectures is that P2P systems can provide an energy savings because the peer devices are already powered on for other purposes. This means that the baseline power consumption of these devices is not factored into the measurements. This assumption can be found in [13, 7], and [14]. However, according to [6], many BitTorrent users actually leave their computers on for the express purpose of participating in the file sharing network. Therefore, it behooves us to include the baseline energy consumption of these peers, as in Eqn. 3. In this paper, we refer to peers who stay in the network solely for participating in the cloud service as *dedicated peers*.

$$E_{h3} = ((1-n)(c(\delta_s + \gamma_s)B + d_s(\delta_r + \gamma_r)B) + n(\omega_p(\delta_p + \gamma_p)B + \omega_r d_p(\delta_r + \gamma_r)B)) \quad (3)$$

4 Analysis

This section provides a detailed analysis of the energy model presented in Sec. 3. We have also simulated the model, and the simulation results agree with the analytical results in principle, which justifies the correctness of the model.

4.1 Server Cost vs. Router Cost

We will begin the analysis with an examination of the energy consumed by each network entity, namely, servers, routers, and peers, as well as their impact on the total energy consumption. To do this, we first create two reference points: the energy consumed by the data centre network (E_{cloud}), by setting $n = 0$ in Eqn. 1, and the energy consumed by the P2P network (E_{P2P}), by setting $n = 1$ in Eqn. 1. By applying the initial values from Table 3, we have $E_{cloud} = 1.56 \times 10^{-6} J/b$ and $E_{P2P} = 5.65 \times 10^{-7} J/b$. In this case, the cloud network consumes 2.75 times more energy than a pure P2P network does. If we consider a more optimal server overhead $c = 1.2$, $E_{cloud} = 9.75 \times 10^{-7} J/b$, which is more comparable to that of the P2P network. Hence, in a pure data centre-based cloud network, the server energy consumption has significant influence on the overall energy efficiency.

When we include the router baseline energy using Eqn. 2, $E_{cloud} = 3.51 \times 10^{-6} J/b$ and $E_{P2P} = 5.07 \times 10^{-6} J/b$. This implies that once the network costs are added to the model, the P2P network consumes about 45% more energy than the strictly-data centre network, a very dramatic reversal. It is clear that the server baseline energy is the factor that weighs down the efficiency of the data centre, whereas the energy consumed by the routers hinders the efficiency of the whole network

We will focus on $0 < n < 1$, the hybrid case, and use the extreme cases of E_{cloud} and E_{P2P} as reference points. Fig. 2 compares energy cost with and without the router baseline energy consumption, for two different server overhead (c) values, over different values of n. We observed that on one hand, when the baseline router energy consumption is not considered, the energy consumption drops as the hybrid infrastructure moves towards a pure P2P network. On the other hand, when the baseline router energy is included, the energy consumption increases as the hybrid infrastructure moves towards a pure P2P network.

Compared to the router energy consumption, the server overhead has a relatively small impact on the overall energy consumption. As P2P applications constitute over 60% of traffic on the Internet, many routers are added to the network just to accommodate P2P traffic. Similarly, future hybrid cloud systems will also require more routers. Therefore, Eqn. 2 provides a better view of overall energy consumption than Eqn. 1, and in reality, the value of ω_r might be different for each router.

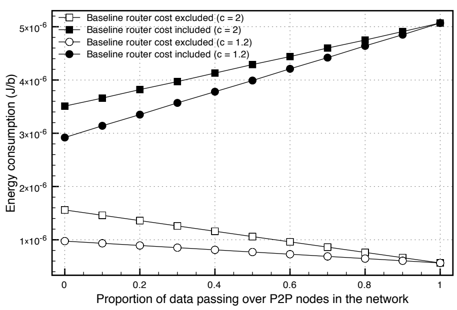


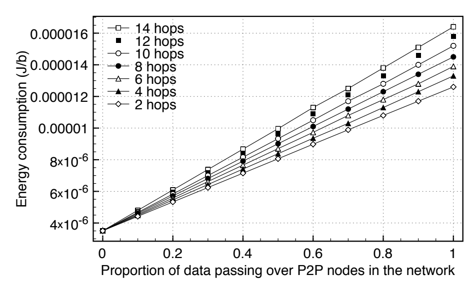
Figure 2. Energy Eqn. 2 vs. n, for $c = 1.2$ and 2

It is clear that the energy consumption increases as more data is portioned over the P2P nodes in the hybrid system. The edge values of n are included here for completeness, despite the fact that in reality it is unlikely that a cloud provider would ever choose to use a hybrid architecture that only puts 5-10% of the data on the P2P portion of the network. This would result in incurring the administrative hardships involved in overseeing and integrating two types of network architectures, while failing to realize the full extent of energy saving benefits a hybrid architecture might offer. Similarly, a hybrid system where the vast majority (85-95%) of data was stored in the P2P network might raise issues of data persistence in the face of node churn. In reality a value between 0.25 and 0.75 seems to be optimal for n. However, we noted that if all peers in the network are dedicated peers, then Eqn. 3 leads to increasing energy consumption regardless the c value and whether the baseline router cost is included. So far, we have learned that P2P cannot help the cloud go green, which leads to further investigation on the source of the extra energy utilization.

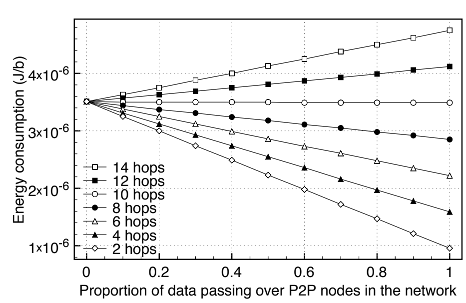
4.2 Network Distance

Given that the router cost has such a high impact on the overall energy consumption, intuitively, we would want to make the routers more energy efficient or reduce the number of routers involved in a data flow, *i.e.*, reducing network distance d_p between two peers. The number of hops incurred by P2P networks was originally set at 15. However, given that in a P2P network it is possible to download data from a variety of sources, a downloading peer can be selective in the peers from which they choose to download.

Fig. 3(a) and 3(b) compares the overall energy consumption, with and without the peer baseline energy cost, for different average inter-peer hop counts d_p, over different values of n. We observed that regardless the values of n and d_p, the energy consumption is monotonically increasing when the peer baseline energy is included. In other words, there is no power savings over P2P networks if all peers are dedicated peers. Without the peer baseline energy cost, Fig. 3(b) shows that if d_p drops below 10, the hybrid cloud architecture will be more energy efficient as n increases. In fact, researchers in the P2P community have already been studying ways to reduce the distances between peers. For example, modified BitTorrent clients choose closer peers to reduce cross-ISP traffic in [8] As a result, 50% of the peers are able to download from peers within 6 hops, and 20% of the peers are able to download from peers that are only 1 hop away. It is also worth noting that the plug-in in question only used biased selection for a small portion of the client's peers. This means the presented hop counts might potentially be very conservative, and even greater savings might be possible.



(a) With peer baseline energy cost (Eqn. 3)



(b) Without peer baseline energy cost (Eqn. 2)

Figure 3. Energy vs. n, for different d_p

The reduction in hops that can be achieved through biased peer selection, only looks at the downloading half of the transaction. The notion is that a peer would search for other peers close to them that had the content they wanted. This immediately constrains the pool of available peers to only those that have the particular data the downloader is interested in obtaining. There exists a compelling opportunity to even further reduce the number of network hops in a hybrid storage cloud through the process of selective uploading. Individuals interested in storing files in the cloud could locate peers within the storage cloud that are close to them, and then choose to upload to these peers over peers which may be further away. Because a hybrid storage cloud would involve some amount of administrative oversight by the service provider, it would be sensible for peers to occasionally report their uptime. This way uploading peers could not only bias their selection towards close peers, but also those that have been online the longest. Obviously, some degree of redundancy would need to be built into the upload process such that if a peer went offline, owners of the files stored on that peer's computer would have another way of gaining access to their content. In addition to reducing energy consumption, a pleasant side- effect of reducing IP hops is that it can result in reduced latency and higher throughput communications between peers.

Given the potential to reduce the value of d_p to less than 6. We use r, defined in Eqn. 4, to quantify the power saved over a strictly data centre-based cloud service.

$$r = 1 - \frac{E_{h2}}{E_{cloud}} \tag{4}$$

Where $E_{cloud} = E_{h2}$ when $n = 0$. Table 2 presents the power saving ratio r for different values of n and d_p. We can save anywhere from 9.2% to 72.7% of energy, depending on the network distance and the amount of traffic flowing through the P2P network.

n	6 hops	5 hops	4 hops	3 hops	2 hops
0.25	9.2%	11.4%	13.7%	15.9%	18.2%
0.50	18.3%	22.8%	27.3%	31.8%	36.3%
0.75	27.5%	34.2%	41.0%	47.8%	54.5%
1.00	34.6%	45.7%	54.7%	63.7%	72.7%

Table 2. Power saving in a P2P-Cloud system for different values of n and d_p, and $c = 2$.

Furthermore, we can fix $n = 0$ to study the impact of the network distance on energy consumption in a strictly C/S architecture. The power saving ratio r is then redefined as:

$$r = 1 - \frac{E_{h2}}{E_{basic}} \tag{5}$$

where E_{basic} is the energy cost of the data centre network based on E_{h2} when $n = 0$ and $d_s = 13$. As we reduce d_s from 6 to 2, we observe 31.5% to 49.6% energy savings.

4.3 Can P2P Help the Cloud Go Green?

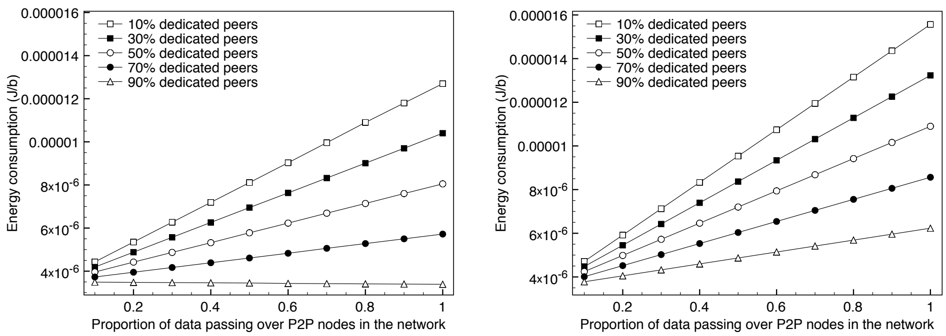
At last, we will examine the impact of the peers on the overall energy consumption in a hybrid cloud system. Since we have concluded that it is very likely that $d_p \leq 6$ can be achieved, we will set d_p to 6 hereafter. We first compute the two reference points $E_{cloud} = 3.51 \times 10^{-6} J/b$ and $E_{P2P} = 1.39 \times 10^{-5} J/b$ by setting n in Eqn. 3 to 0 and 1, respectively. It is clear that the data centre network is 4 times more energy efficient than the P2P network if users leave their computer on just to take part in the P2P service. In order to conserve energy, we must either: make each peer more energy efficient, or turn peers off as much as possible. In reality, peers can be online at any time for arbitrary purposes. To model the general case, we introduce a new weighted formula.

$$\begin{aligned} E_{h4} = &\alpha_1(c(\delta_s + \gamma_s)B + d_s(\delta_r + \gamma_r)B) \\ &+\alpha_2(\omega_p(\delta_p + \gamma_p)B + \omega_r d_p(\delta_r + \gamma_r)B) \\ &+\alpha_3(\omega_p \delta_p B + \omega_r d_p(\delta_r + \gamma_r)B) \end{aligned} \tag{6}$$

where $0 \leq \alpha_1, \alpha_2, \alpha_3 \leq 1$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. The weight factor α_1 is the proportion of data flowing through the data centre portion of the network, α_2 is the proportion of data flowing through peers that are online just for the cloud service, and α_3 is the proportion of data flowing through peers that are online for another purpose ($\gamma_p = 0$).

Fig. 4(a) and Fig. 4(b) compares the overall energy consumption defined in Eqn. 6 for different average number of hops between peers. In each figure, we vary the proportion of data flowing from the peers ($\alpha_2 + \alpha_3$) and the percentage of dedicated peers ($\frac{\alpha_2}{\alpha_2+\alpha_3}$). We first noted that there is no energy saving if the average number of hops between peers is more than 6. Furthermore, as shown in Fig. 4(a), if dedicated peers constitutes more than 10% of the P2P network, then the energy consumption will grow as more data passes through the P2P network.

Nonetheless, it is not possible to require a peer to power off, but it is possible to make the device go idle until further notice. This inspires the use of a proxy, as suggested in [6]. Peers can take turns to stay online and



(a) $d_p = 6$ (b) $d_p = 15$

Figure 4. Energy (Eqn. 6) vs. proportion of data flowing from the peers ($\alpha_2 + \alpha_3$), for different percentage of dedicated peers ($\frac{\alpha_2}{\alpha_2+\alpha_3}$)

act as proxies for nearby idle peers. When a transmission completes or the service requires responses from the user, the idle peer will be brought online and receive data from the proxies. The assumption is that a connection between a peer and a proxy is much faster than the connection between two peers separated by a greater number of network hops.

	Meaning	Initial Value
c'	CDN Server overhead	1.2
δ_s	CDN Server energy used per unit	$5.2\times10^{-8}(J/b)$
γ_s	CDN Server baseline energy	$6.7\times10^{-7}(J/b)$
d_c	Avg. # hops (CDN to peers)	6
d_p^{intra}	Avg. # hops within an ISP	6
d_p^{inter}	Avg. # hops across ISPs	15

Table 3. Description and starting values for updated model parameters.

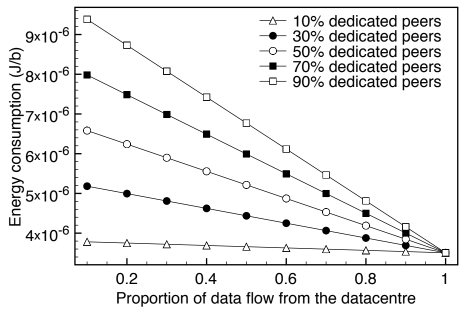
The notion of deploying a collection of highly distributed, modestly resourced nodes to act as proxies within the cloud might actually be appealing to content providers. In other words, content providers could rent servers on content delivery networks (CDNs) or build small data centres closer to users. Either approach allow providers to extend their ability to manage their services. However, this begins to blur the line between a true P2P network, and the more traditional concept of distributed computing. To model this setup, we first introduce three measurements on a CDN server: CDN server overhead c', CDN server energy used per unit δ_c, and CDN server baseline energy γ_c. The average number of hops from the CDN servers to peers is denoted by d_c. Furthermore, we assume that the peers are modified to choose closer peers to reduce cross-ISP traffic, as in [8]. This then required us to treat inter-ISP traffic and intra-ISP traffic differently. Let d_p^{intra} and d_p^{inter} denote the average number of hops between peers within the same ISP and peers across different ISPs, respectively. The new symbols and their initial values are summarized in Table 3.

We argue that the CDN servers are deployed at a small scale, requiring less overhead. As concluded in [8], a peer running ISP-friendly applications tends to find peers within 6 hops. We set the initial value of d_c to 6 as CDN servers are placed close to peers, typically within the same ISP. We now model CDN-based P2P-cloud setup in Eqn. 7.

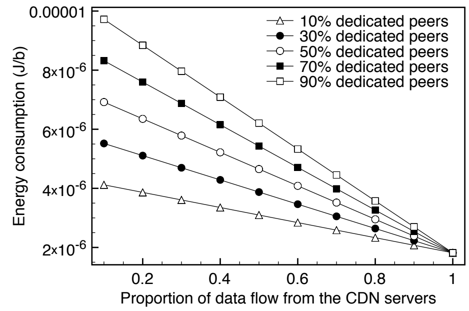
$$\begin{aligned} E_{h5} = {} & \alpha_1(c(\delta_s+\gamma_s)B + d_s(\delta_r+\gamma_r)B) \\ & +\alpha_2(c'(\delta_c+\gamma_c)B + d_c(\delta_r+\gamma_r)B) \\ & +(\alpha_3+\alpha_4)(\beta_1(\omega_p(\delta_p+\gamma_p)B + \beta_2(\omega_p\delta_pB) \\ & +\alpha_3(\omega_r d_p^{inter}(\delta_r+\gamma_r)B) \\ & +\alpha_4(\omega_r d_p^{intra}(\delta_r+\gamma_r)B) \end{aligned} \quad (7)$$

where $0 \le \alpha_1, \alpha_2, \alpha_3, \alpha_4 \le 1$ and $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$. The weight factor α_1 is the proportion of data flowing through the data centre portion of the network, α_2 is the proportion of data flowing from the CDN servers, α_3 is the proportion of inter-ISP data flowing among peers, and α_4 is the proportion of intra-ISP data flowing among peers. In addition, the weight factors β_1 and β_2 quantify the proportion of peers that are online just for the cloud service and the proportion of peers that are online for another purpose, respectively.

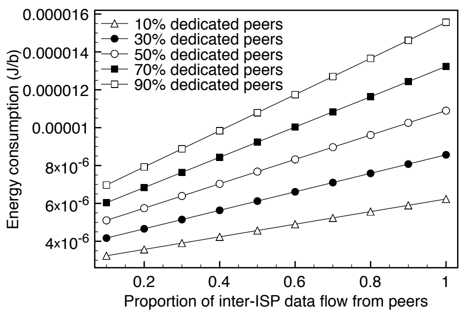
Fig. 5 presents the impact of different network components in a CDN-based P2P-cloud network. We vary one α_i value from 0.1 to 1, and let the remaining weight factors equally share the remaining values, *i.e.*, $\frac{1-\alpha_i}{3}$. For each figure, we also adjust the percentage of dedicated peers from 10% to 90%. Several observations



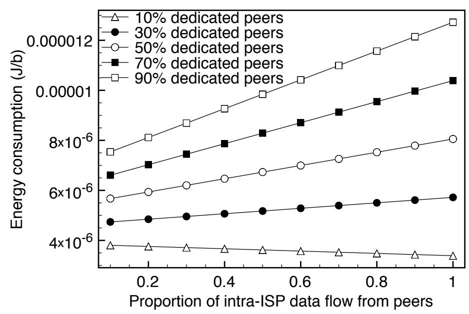
(a) $\alpha_1 = 0.1 \ldots 1$



(b) $\alpha_2 = 0.1 \ldots 1$

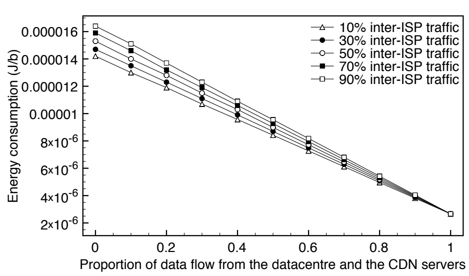


(c) $\alpha_3 = 0.1 \ldots 1$

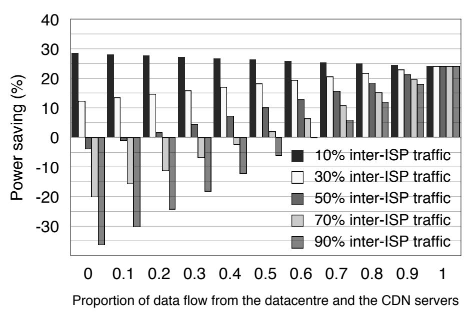


(d) $\alpha_4 = 0.1 \ldots 1$

Figure 5. Energy (Eqn. 7) vs. proportion of traffic over different part of the network, for different percentage of always-on peers ($\frac{\beta_1}{\beta_1+\beta_2}$)



(a) Energy (Eqn. 7) vs. proportion of data flows from the data centre and the CDN servers ($\alpha_1 + \alpha_2$), for different percentage of inter-ISP traffic ($\frac{\alpha_3}{\alpha_3+\alpha_4}$)



(b) Energy saving (Eqn. 4) vs. proportion of data flows from the data centre and the CDN servers ($\alpha_1 + \alpha_2$), for different percentage of inter-ISP traffic ($\frac{\alpha_3}{\alpha_3+\alpha_4}$), with $\beta_1 = 1$.

Figure 6. Energy consumption and savings of the CDN-based P2P-cloud

are drawn from Fig. 5. First, it is nearly impossible to achieve any power saving with more than 10% of dedicated peers. Second, as shown in Fig. 5(a), there are no potential savings when moving data away from the data centre. Third, Fig. 5(b) shows that moving more than 50% of data to CDN servers will lead to energy saving. Forth, Fig. 5(c) and Fig. 5(d) suggests that ISP traffic localization may help to save energy. From these observations, we concluded that P2P infrastructure does not necessary leads to a greener cloud, as it leaves a heavy footprint due to the additional network routers involved.

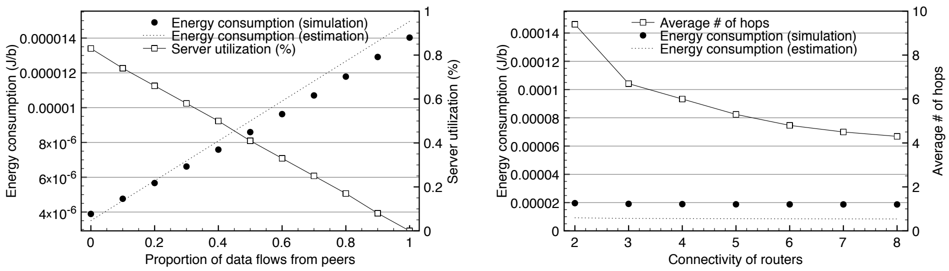
Next, we vary the proportion of data flows from the datacentre and CDN, *i.e.*, $(\alpha_1 + \alpha_2)$. As indicated in Fig. 6(a), there is a sharp decrease in energy demand when increasing the amount of data hosted at the data centre and the CDN servers. Some savings are also observed when reducing inter-ISP traffic. Nonetheless, compared to the energy consumed by a strictly C/S architecture ($\alpha_1 = 1$ for Eqn. 7), the proposed CDN-based P2P-cloud can conserve energy only when hosting more than 90% of the data at the data centre and/or the CDN servers. The actual saving is 24%. So far, we have identified five sources responsible for the high energy costs, namely, the server overhead c, the average number of hops (both among peers d_p and from the data centre to peers d_s), router baseline energy γ_r, percentage of dedicated peers β_1, and proportion of data on the (CDN) servers $(\alpha_1 + \alpha_2)$. Without performing a large-scale infrastructural change, there is very little that we do to change these parameters. However, home modems and routers may be modified at a relatively lower cost. These devices may operate as a proxy for connected peers. Since they are mostly powered on 24/7 to provide Internet access, the baseline energy cost can be ignored, *i.e.*, $\gamma_p = 0$. Fig. 6(b) illustrates the potential energy saving offered by proxy services on home modems and routers. Up to 28% of energy saving can be achieved when there is less than 10% of inter-ISP traffic, even when all data are hosted on the P2P network.

5 Simulation: A Case Study

To verify the our model and analytical results, we implemented a simulator. Instead of creating a synthetic network, we designed the simulator to mimic the Wuala system, the most successful hybrid P2P-cloud storage network. Nodes in this network are organized into three rings. The innermost ring consists of at least 1 server, representing the data centre. Each server connects to one router. The routers are managed in a ring structure, which constitutes the middle ring. Together the server and the routers collectively act as the super nodes and the storage nodes in Wuala. The outermost ring is formed by the peers. Each peer connects to at least one router. Due to space limit, we only present the most representative results in this paper. The network used to generate these results consist of 30 10 Gbps routers and 300 peers with 80 Mbps bandwidth. The server bandwidth is set to 864 Mbps. We started with a single server, whose maximum power utilization is 336 watts. Routers have a maximum power utilization of 830 watts, and peers have a maximum power utilization of 153 watts. These number are adapted from [13]. It is worth to note that our simulator is not driven by the analytical model. Instead, it measures the energy consumed by each network component based on the amount of traffic travelled through the network. In other words, the measurements reflects the actual energy usage.

First, we verify the main conclusion from Sec. 4. We varied the proportion of data flows from peers. To partially prove the correctness of the simulator, we present the server utilization from this experiment in Fig. 7(a) (right y-axis). The server utilization drops to zero when moving data from the data centre to peers. The dots and the dotted line in Fig. 7(a) show the actual measurements from the simulator and the estimation obtained from Eqn. 6. Our model provides a very close estimation of the energy consumption. The simulation confirms that moving data away from the data centre will result in higher energy consumption. Hence, incorporating P2P infrastructure will not necessary make the cloud greener. Next, we split the data 50-50 between the data centre and the peers, and then vary the number edges between the routers from 2 to 8 to make the routers better connected. This results in reduction in average number of hops travelled by data, shown in Fig. 7(b) (right y-axis). The measured energy consumption is slightly higher than the estimation, but have the same decreasing trend as the average number of hops decreases. This confirms our earlier conclusion that reducing distance travelled by data leads to energy saving.

Lastly, we increased the number of servers to simulate a CDN assisted cloud. Since CDNs are servers deployed at different ISPs, they are located closer to peers, which significantly reduces the average number of hops from servers to peers. Moreover, these servers are at a smaller scale compared to data centre servers. For this reason, the baseline power and maximum utilization power are proportionally reduced. The percentage of energy savings is presented in Table 4. With 30 CND servers, we can achieve 40% energy saving, *i.e.*, a greener cloud.



(a) Energy consumption and server utilization v.s. proportion of data flows from the data centre

(b) Energy consumption and average number of hops v.s. connectivity of routers

Figure 7. Energy vs. n, for different d_p

Number of CDN servers	1	10	15	20	30
Average # of hops	4.3	2.9	2	1.4	1
Energy saving	0%	14.5%	29%	38.4%	40%

Table 4. Power saving in a P2P-Cloud system for different number of CDN servers

6 Conclusion

This paper examined the question of whether a cloud storage provider could reduce the energy consumption of their service offering by adding a peer-to-peer network to supplement their data centre, creating a hybrid network. We developed an energy utilization model and used it to examine different network configurations. Both the analytical study and the simulation suggests that energy savings are possible under the right circumstances. We now conclude that the answer to the question "Can P2P help the cloud go green?" is both "yes" and "no". Directly introducing P2P elements into the cloud service will result in higher overall energy consumption. Our study shows that a combined modification at the data centre, ISP, and peers is required. The modifications include reducing server overhead, reducing average number of hops that data travels, localizing network traffic within ISPs, utilizing peers or devices that are already online for other purposes, and distributing data over a CDN network. To the best of our knowledge, this is the first work presents a complete energy analysis on hybrid P2P-cloud network infrastructure.

References

[1] Amazon s3. `http://www.amazon.com/s3`.

[2] Data center report to congress. `http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congess_Final1.pdf`.

[3] Seti@home. `http://setiathome.berkeley.edu`.

[4] The uptime institute. `http://www.uptimeinstitute.org`.

[5] Wuala by lacie. `http://www.wuala.com`.

[6] G. Anastasi, I. Giannetti, and A. Passarella. A bittorrent proxy for green internet file sharing: Design and experimental evaluation. *Computer Communications*, 33(7):794 – 802, 2010.

[7] S. A. Baset, J. Reich, J. Janak, P. Kasparek, V. Misra, D. Rubenstein, and H. Schulzrinnne. How green is ip-telephony? In *Proceedings of the first ACM SIGCOMM workshop on Green networking*, Green Networking '10, pages 77–84, New York, NY, USA, 2010. ACM.

[8] D. R. Choffnes and F. E. Bustamante. Taming the torrent: a practical approach to reducing cross-isp traffic in peer-to-peer systems. *SIGCOMM Comput. Commun. Rev.*, 38:363–374, August 2008.

[9] K. Christensen and A. George. Increasing the energy efficiency of the internet with a focus on edge devices. The Energy Efficient Internet Project, 2005-2008. University of South Florida.

[10] A. Foulonneau, J. Bidet, and P. Raipin. Clostera: Cloud storage enabler for richmedia applications. In *Intelligence in Next Generation Networks (ICIN), 2010 14th International Conference on*, pages 1 –5, oct. 2010.

[11] V. Gopalakrishnan, B. Bhattacharjee, K. Ramakrishnan, R. Jana, and D. Srivastava. Cpm: Adaptive video-on-demand with cooperative peer assists and multicast. In *INFOCOM 2009, IEEE*, pages 91 –99, april 2009.

[12] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen. Reducing the energy consumption of ethernet with adaptive link rate (alr). *Computers, IEEE Transactions on*, 57(4):448 –461, april 2008.

[13] S. Nedevschi, S. Ratnasamy, and J. Padhye. Hot data centers vs. cool peers. In *Proceedings of the 2008 conference on Power aware computing and systems*, HotPower'08, pages 8–8, Berkeley, CA, USA, 2008. USENIX Association.

[14] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez. Greening the internet with nano data centers. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, CoNEXT '09, pages 37–48, New York, NY, USA, 2009. ACM.