Reducing Electricity Usage in Internet using Transactional Data

Bhushan Ahire[1], Meet Shah[2], Ketan Prabhulkar[3], Nilima Nikam[4]

[1,2,3]*Student, Dept. of Computer Science and Engineering, YTIET College, Karjat, India*
[4]*Professer, Dept. of Computer Science and Engineering, YTIET College, Karjat, India*

---***---

Abstract - *Along with the increasing Internet services and cloud computing in recent years, the power usage associated with data center operations has been surging significantly. The demand for data centers is expected to grow owing to the increasing requirement for data processing related to Internet of Things (IoT) applications. This huge energy consumption leads to not only heavy operation cost but also negative environmental impact. Consequently, reducing the power consumption of data centers has become an urgent and major issue. In this paper, we investigate how to utilize the temporal diversities of geographical and electricity price to data center within the specified cost budget. To this end, we model the long-term cost budget & dynamic data constraint and formulate the energy utilization maximization problem as a stochastic programming problem. While all future knowledge of electricity price, workload and energy availability are unknown, we apply Lyapunov optimization framework and design an efficient online algorithm. We redevelop the Dynamic Request Mapping algorithm to get desired results for this problem. Extensive performance evaluations show that the proposed algorithm can significantly outperforms the baseline strategies with respect to system cost, queue backlogs and delay.*

***Key Words*: Power usage, dynamic data, stochastic, Lyapunov optimization framework, request mapping**

1. INTRODUCTION

The significant continuous increase in the power consumption of data centers, brought about by increased usage of cloud computing, online storage services, and social networking, is a major social problem. Moreover, the demand for data centers is expected to grow due to the increasing requirement for data processing related to Internet of Things (IoT) applications. Recently with the increasing electricity cost of cloud data centre to the point that it can often override the cost of IT equipment over a period of time, power consumption has become an important problem for infrastructure providers. In 2013, the electricity power consumption of data centers in the US was estimated as 91 billion kWh and it is predicted to reach 140 billion kWh in 2020.On the global scale, data centres are poised to be the largest global energy users by 2025 at 4.5%, an increase from just 0.9% in 2015, according to Andrae's report.

Recently with the increasing electricity cost of cloud data centre to the point that it can often override the cost of IT equipment over a period of time, power consumption has become an important problem for infrastructure providers. As a result, millions of dollars must be spent on electricity by the infrastructure holders or service. However, a reduction in the power consumption of certain equipment does not necessarily reduce the total power consumption of the data center. For example, in general the heat generated by a server and its CPU utilization follows a linear relationship. Therefore when the tasks are consolidated to a portion of servers, the temperature of those servers rises. Following that, cooling with air conditioners is required to cope with the rise of temperature. Consequently, the power consumption of air conditioners increases and the total power consumption of the data center cannot be reduced as expected. With the rapid growth of the Internet-scale systems in size, how to reduce the corresponding energy consumption or cost has been attracting a growing number of research interests.

1.1. Related Work

A considerable amount of investigation and research have been conducted to improve the efficiency of Internet-scale systems by exploiting the geographical diversity. In the following, we conduct a literature review of existing efforts relevant to our study. Generally speaking, the existing schemes can be categorized into predictable model-based schemes and stochastic optimization-based schemes.

1.1.1. Stochastic Optimization-Based Schemes

Considering that most system states are stochastic and potentially unpredictable in nature, several recent studies have attempted to exploit geographical diversity and improve system performance via a stochastic optimization framework [9],[10], [11], [12]–[14], such as Lyapunov optimization [15],[16], [17]. For example, Liu et al. in [9] introduced a control framework to optimize the power-performance tradeoff by making online decisions on request admission control, routing, and scheduling of virtual machines. This study was launched in response to dynamic and unpredictable user requests from heterogeneous applications served by a SaaS cloud datacenter platform. For cloud-based video applications with unpredicted request arrival, Xiao et al. in [10] introduced an online algorithm to minimize the renting cost of virtual machines while maintaining the QoE (Quality of Experience) requirements of users. The algorithm was designed based on the Lyapunov optimization framework for dynamic request direction and resource provision. Considering the temporal and spatial variations in the workload arrival and electricity price, Yao etal. in [11] extended the Lyapunov optimization framework to two different time scales, and presented a stochastic power reduction algorithm to control the job distribution and the number of active servers within data center. Yu et al. in [12], [13] attempted to minimize the energy cost by considering uncertainties in electricity price, workload, renewable energy generation, and power outages in microgrids of the smart grid. The problem was formulated as a stochastic program for job request distribution, server provisioning, energy storage management, generator scheduling, and power transmission. The operation algorithm, based on the Lyapunov optimization technique, was designed to balance the tradeoff among various considerations.

1.1.2. Predictable Model-Based Schemes

This type of scheme requires the optimization strategy to be precomputed based on the full knowledge of the random system model, such as job arrival, random service rate, etc. The corresponding control decisions, or schedule policies, at front-end servers and back-end data centers are generally made by solving a deterministic optimization problem. For example, Qureshi et al. in [4] argued that significant economic gains can be achieved by exploiting the variation of fluctuating electricity prices. Rao et al. in [5] proposed a scheme to minimize the total electricity cost by controlling the request allocation and the number of active servers. They formulated the problem as mixed-integer programming with delay constraints, and solved the approximate formulation via a proposed polynomial-time algorithm. The job request mapping in [6] considers not only the electricity cost of a particular workload assignment, but also the transition cost associated with reallocating workload between data centers. Liu et al. in [7] modeled a more complicated energy optimization system, which incorporates renewable generation, storage, an optimized cooling micro-grid, and batch job scheduling. To this end, an integrated workload management system was proposed for data centers to take advantage of the possible gains by shifting workload in a way that exploits time variations in electricity price, the availability of renewable energy, and the efficiency of cooling. In addition, Gao et al. in [8] introduced a Flow Optimization-based framework for Request-routing and Traffic Engineering (FORTE) to balance the weighted sum of electricity cost, carbon footprint, and access latency. The problem is formulated as a linear program under the constraint of data placement, and is solved by a fast heuristic algorithm.

1.1.3.System Model and Problem Formulation

In this section, we first describe the functional architecture of Internet-scale service systems with dynamic data. Next, the system models are introduced, including job request arrival and dispatching model, service rate model, dynamic queue model, and system cost model. Using these models, we formulate the problem of job request mapping for energy efficient system as a constrained optimization problem.

1.1.4. Abbreviations and Acronyms

DREM= Dynamic Request Mapping Control Algorithm

FORTE=Flow Optimization-based framework Request-routing and Traffic Engineering

2. System Architecture

A large Internet-scale service system is considered in this paper. It consists of multiple geographically distributed infrastructures. A service provider runs servers over the geo-distributed systems. The general architecture of each system involves the following two main components: front-end servers, and back-end processing infrastructures. Generally speaking, the basic process of the system works as follows. In the system each server having ‘n’ number of data centers if the users send the job request for the front end server, the Front-end servers are responsible for collecting the job requests from users and dispatching the job requests to the appropriate back-end system for processing. In a practical system, front-end servers can be DNS (Domain Name System) servers used by Akamai and most CDNs (Content Delivery Networks), or HTTP proxy servers used by Google and Yahoo. The front end server sends the user queries to the back end system using dispatcher. Back-end system is responsible for handling the job requests, and consists of multiple geo-distributed data centers. Each back-end infrastructure consists of two subcomponents: processing servers and database servers. A database server is used to store the dynamic data, and handle the data-related operations (e.g., query, transfer) from the database of other back-end systems.

2.1 ADVANTAGES

- Timing consumption is less
- Energy consumption is reduced.
- User gets immediate response.

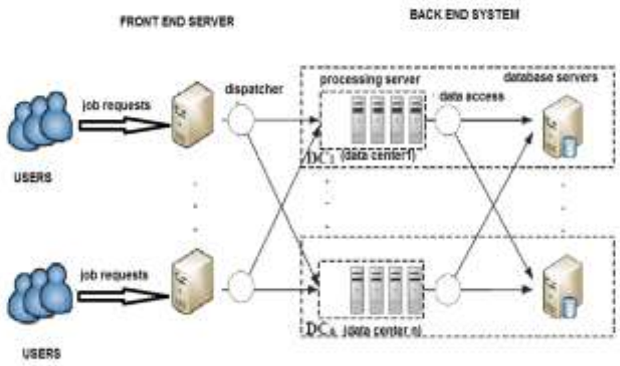

Fig - 1: System Architecture

2.2 Module Description

1. Request processing
2. Data dispatching
3. Processor invoke

2.2.1 Request processing

The multiple user were send request to the server the server is connected to the dispatcher, the dispatcher check the user request that which domain they want to access for. After checking the user queries the dispatcher will dispatch the user process to the available data center. The data server having n-number of data centers, the dispatcher check which data center is available, after checking the data center the dispatchers send the queries to the processing servers. The processing server will processing the data and sends it the database server through data access.

2.2.2 Data Dispatching

In data dispatching module the user queries are stored in the data server; the dispatchers get the user data and check the available data centers. If the data center is available for the user queries, then the dispatcher will dispatch the queries to the data center. The data center having processing servers, the processing servers will process the data which is send by the users. After processing the data the data center will get the data for the user queries from database server through data access. Every group having one dispatcher for processing, they will dispatch the data without any delay.

2.2.3 Processing invokes

In processing invoke module many processing were running in the data center, the data center having the 'n' number of processing servers. The processing server will process the queries and get the data from the database server through data access. This module calculates the processing of user queries and stored the result. It shows which data center is running for which server. It shows which user queries is run by which data center. It shows the processing time of the data center.

3. Algorithm

Dynamic Request Mapping Control (DREM) Algorithm

Our study differs from the existing work in the following aspects: First, distinct from the solution obtained through solving a deterministic optimization problem with the Lyapunov optimization framework, our dynamical control algorithm DREM is more practical and can be performed based on the measurable system states (e.g., current queue backlog) without any prediction of future job arrival and service rate.

Algorithm procedures of DREM:

1: Initialization: U(0) = 0;
2: while the service is running do
3: for each time slot t do
4: // Operations at the front-end servers:
5: for each front-end server Fi 2 F do
6: Observing the job arrival Ji(t);
7: Observing the queue backlogs of all DCs U(t);
8: Observing the current system states for cost computation;
9: Getting the decisions j(t) by solving (19);
10: Executing the mapping decisions j(t) ;
11: end for
12: // Operations at the back-end DCs:
13: for each back-end DC Di 2 D do
14: Observing the new job arrival Aj (t);
15: Updating the queue according to (3);
16: end for
17: end for
18: end while

4. Experimental Results

This following section shows snapshots of our system web application. It shows how our system works and gives us an idea about the system.

Fig - 2: User Registration Page

The fig shows the user registration form where a user can register his./her name and email-id and create an account for himself/herself.

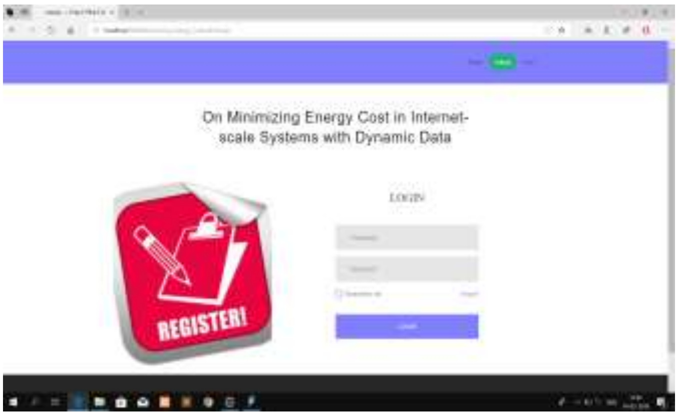

Fig - 3: Admin Login Page

The following screenshot shows Admin login page. The administrator can login and upload data through admin panel.

Fig - 4: User Login Page

The following screenshot shows user login page. The user can login and search places with there location.

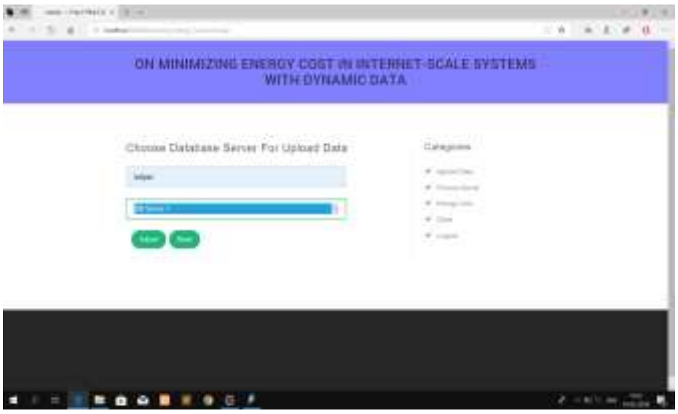

Fig - 5: Admin Panel for Uploading Data

The following screenshot shows the admin panel. The administrator can upload data and information regarding the servers, there location and information inside the database servers.

Fig - 6: Energy Consumption results for each location

The following screenshot shows the energy cost. The administrator can review data and information regarding the servers, there electricity consumption, power cost, total cost.

Fig - 7: User Panel for Searching Places

The following screenshot shows the user panel. The user can search places through there location and the categories of their choice.

5. CONCLUSION

In this paper, we have addressed the issue of how to optimally map job request to back-end data centers for energy efficient Internet-scale systems with dynamic data. We have developed a stochastic optimization based Dynamic Request Mapping control algorithm (DREM), which is capable of minimizing the system cost caused by power consumption, while achieving the system stability. DREM not only can operate devoid of the knowledge of job arrival, service rate, and data distribution, but also can direct the total system cost to be arbitrarily close to the minimum average cost achieved by the optimal policy with complete knowledge of future events. Our evaluation results demonstrate that our approach significantly outperforms baseline strategies with respect to system cost, memory overhead (e.g., queue backlogs), and adaptively to system dynamics.

REFERENCES

[1] F. Chen, T. Xiang, X. Fu, and W. Yu, "User differentiated verifiable file search on the cloud," IEEE Transactions on Services Computing, vol. PP, no. 99, pp. 1–1, 2017
[2] J. Koomey, "Growth in data center electricity use 2005 to 2010," A report by Analytical Press, completed at the request of The New York Times, p. 9, 2011.
[3] S. Vakilinia, B. Heidarpour, and M. Cheriet, "Energy efficient resource allocation in cloud computing environments," IEEE Access, vol. 4, pp.8544–8557, 2016
[4] Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for internet-scale systems," SIGCOMM Comput.Commun. Rev., vol. 39, no. 4, pp. 123–134, Aug. 2009.
[5] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: optimization of distributed internet data centers in a multi-electricity market environment," in 2010 Proceedings IEEE INFOCOM.IEEE,2010, pp. 1–9.
[6] M. S. Ilyas, S. Raza, C.-C. Chen, Z. A. Uzmi, and C.-N. Chuah, "Red bl: energy solution for loading data centers," in 2012 Proceedings IEEE INFOCOM.IEEE, 2012, pp. 2866–2870.
[7] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah,and C. Hyser, "Renewable and cooling aware workload management for sustainable data centers," in ACM SIGMETRICS Performance Evaluation Review, vol. 40, no. 1.ACM, 2012, pp. 175–186.
[8] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, "It's not easy being green," ACM SIGCOMM Computer Communication Review, vol. 42,no. 4, pp. 211–222, 2012.
[9] F. Liu, Z. Zhou, H. Jin, B. Li, B. Li, and H. Jiang, "On arbitrating the power-performance tradeoff in saas clouds," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 10, pp. 2648–2658, 2014.
[10] W. Xiao, W. Bao, X. Zhu, C. Wang, L. Chen, and L. T. Yang, "Dynamic request redirection and resource provisioning for cloud-based video services under heterogeneous environment," IEEE Transactions on Parallel and Distributed Systems, vol. PP, no. 99, pp. 1–1, 2015.
[11] Y. Yao, L. Huang, A. B. Sharma, L. Golubchik, and M. J. Neely ,"Power cost reduction in distributed data centers: A two-time-scale approach for delay tolerant workloads," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 200–211, 2014.
[12] L. Yu, T. Jiang, and Y. Zou, "Real-time energy management for cloud data centers in smart microgrids," IEEE Access, vol. 4, pp. 941–950,2016.
[13] L. Yu, T. Jiang, and Y. Cao, "Energy cost minimization for distributed internet data centers in smart microgrids considering power outages," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 1,pp. 120–130, 2015.
[14] Y. Guo, M. Pan, Y. Gong, and Y. Fang, "Dynamic multi-tenant coordination for sustainable colocation data centers," IEEE Transactions on Cloud Computing, 2017.
[15] L. Yu, T. Jiang, and Y. Zou, "Distributed real-time energy management in data center microgrids," IEEE Transactions on Smart Grid, 2016.
[16] M. J. Neely, "Energy optimal control for time-varying wireless networks," IEEE Transactions on Information Theory, vol. 52, no. 7, pp.2915–2934, July 2006.
[17] M. J. NNeely, "Dynamic power allocation and routing for satellite and wireless networks with time varying channels," Ph.D. dissertation, Massachusetts Institute of Technology, 2003.