Incorporation of Weighted Linear Prediction Technique and M/M/1 Queuing Theory for Improving Energy Efficiency of Cloud Computing Datacenters

Electrical Engineering and Computer Science
Washkewicz College of Engineering, Cleveland State Univeristy

Students: Elham Akbari; Francis Cung; Hardik Patel
Advisor: Dr. Abdul Razaque

ABSTRACT

Cloud computing refers to both the services supplied over the internet and the hardware and software that delivers such services. It has the capability to cover a large part of the IT industry and make software even more appealing as a service. It can also reshape how IT hardware is designed and acquired. Nevertheless, data centers that offer cloud applications consumes a large amount of power which can substantially increase operational costs. On the other hand, as the technology spreads more and participates further into human life there is a larger demand to extend the platforms needed for boosting the development of Cloud computing. As a result, in parallel with this development of infrastructure, there has been also a great deal of attention to energy consumption of such infrastructure. This work reports on the state of the art energy saving technique, weighted linear prediction techniques and M/M/1 Queuing Theory .

INTRODUCTION

- Over the past 20 years the amount of energy utilized in computing data centers has doubled and is expected to double again within the next 20 years. Thus along with forming the infrastructure there is also a great need to update such for becoming more energy efficient in the favor of the environmental sustainability as well as international economy.
- Energy efficiency has been a critical concern even before the presence of the cloud notion. Traditionally there has been a great focus on energy saving of devices such as laptops and mobile devices so as to prolong their battery lifetimes. A number of energy saving methods that were originally proposed for this purpose has been also employed by the cloud servers to lower their power consumption. For instance, Dynamic Voltage and Frequency Scaling (DVFS) is one of the most frequently used power reduction schemes in high performance processors. DVFS can change the frequency and voltage of a microprocessor according to various prediction protocols that estimate CPU utilization level in the future .
- The energy-efficient techniques for managing cloud centers can be divided into three main categories: virtual machine (VM) placement/workload consolidation, resource over-commitment, and workload prediction.
- The resource usage log comprises of very valuable data that may be used to enhance the accuracy of next-time workload predictions. Upon integration of such valuable data with resource prediction approaches the violation rate can be significantly reduced.
- This work aims at exploiting "Weighted Linear Prediction Method" (WLPM) along with the M/M/1 Queuing Theory (MMQMPM) for taking the effect of different workloads and therefore enhancing the response time of the system when several application services are functioning. CloudSim or a similar cloud simulator will be used to empirically compare the results of the proposed versus existing techniques.

SIGNIFICANCE

- Help reduce the carbon footprint by improving energy efficiency in cloud computing datacenters
- Assist in providing a technique to improve energy efficiency within the cloud computing datacenters.
- The research will provide a basis to further look into weighted linear prediction technique to improve energy efficiency in cloud computing datacenters.
- Although, this technique might not be ideal, it will serve as the foundation for others to hopefully consider and enhanced the technique.

EXISTING LIMITATIONS

- The problem of linear prediction method (LPM) is the difference between the power model of old and new generations of servers. As a result during an idle state with no workload, LPM with an old generation server consumes more energy than a new generation server.
- By arbitrarily selecting old and new generation servers from the power benchmark it has been found that the slope of old generation servers do not differ much from new generation servers.
- This means that the energy efficiency of old generation servers is not as good only because they use a lot of energy at idle conditions. Therefore, utilizing a model for the prediction method that takes the effects of such variations and uncertainties into account will be much beneficial.

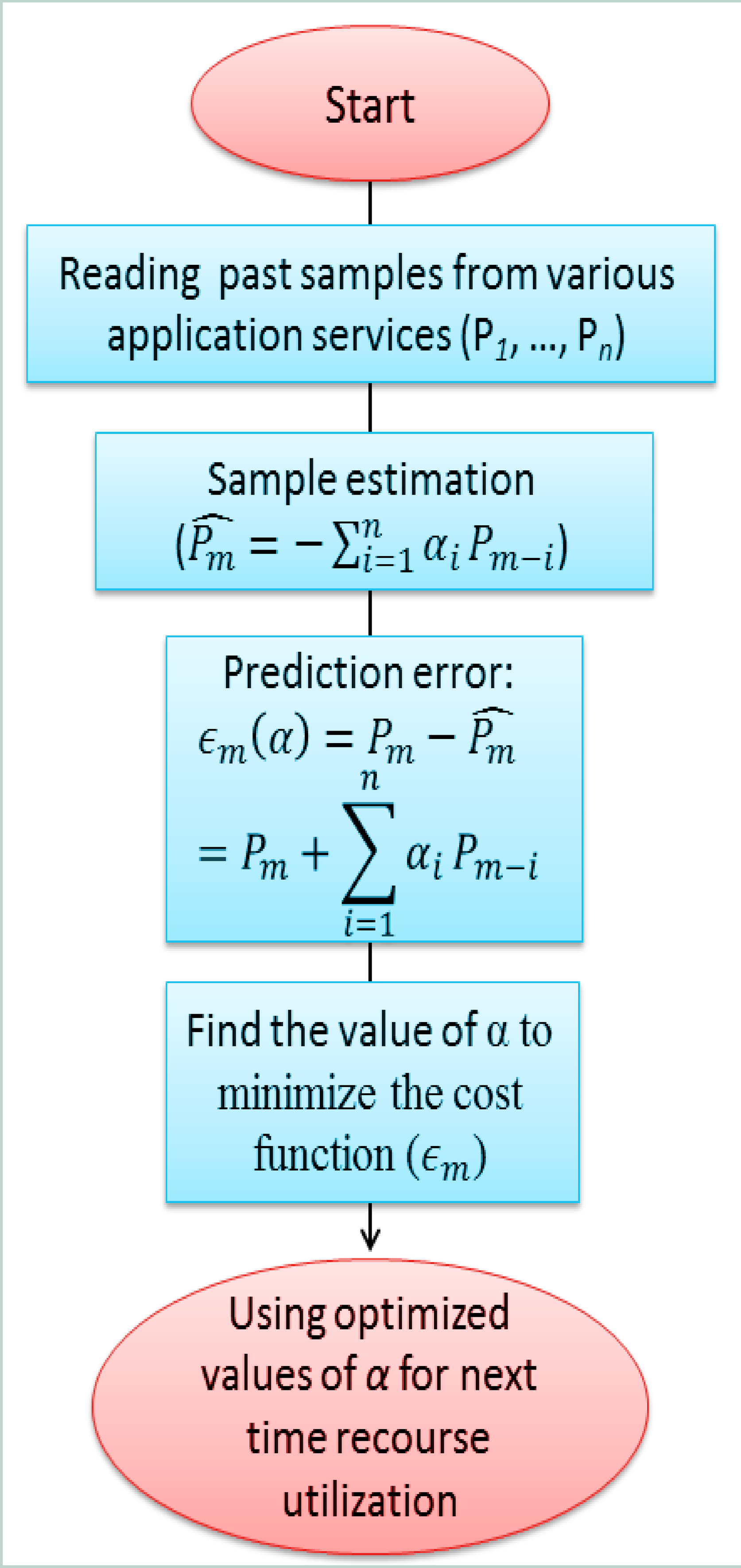
PROPOSED TECHNIQUE

- WLPM always leads to stable all-pole models by introducing a modified short-time energy function.
- Upon application of such technique and its stabilized version to workload energy saving techniques extremely accurate information from the utilization log can be obtained.
- Introducing weights to different application services, the amount of the predicted value can be obtained as follows:

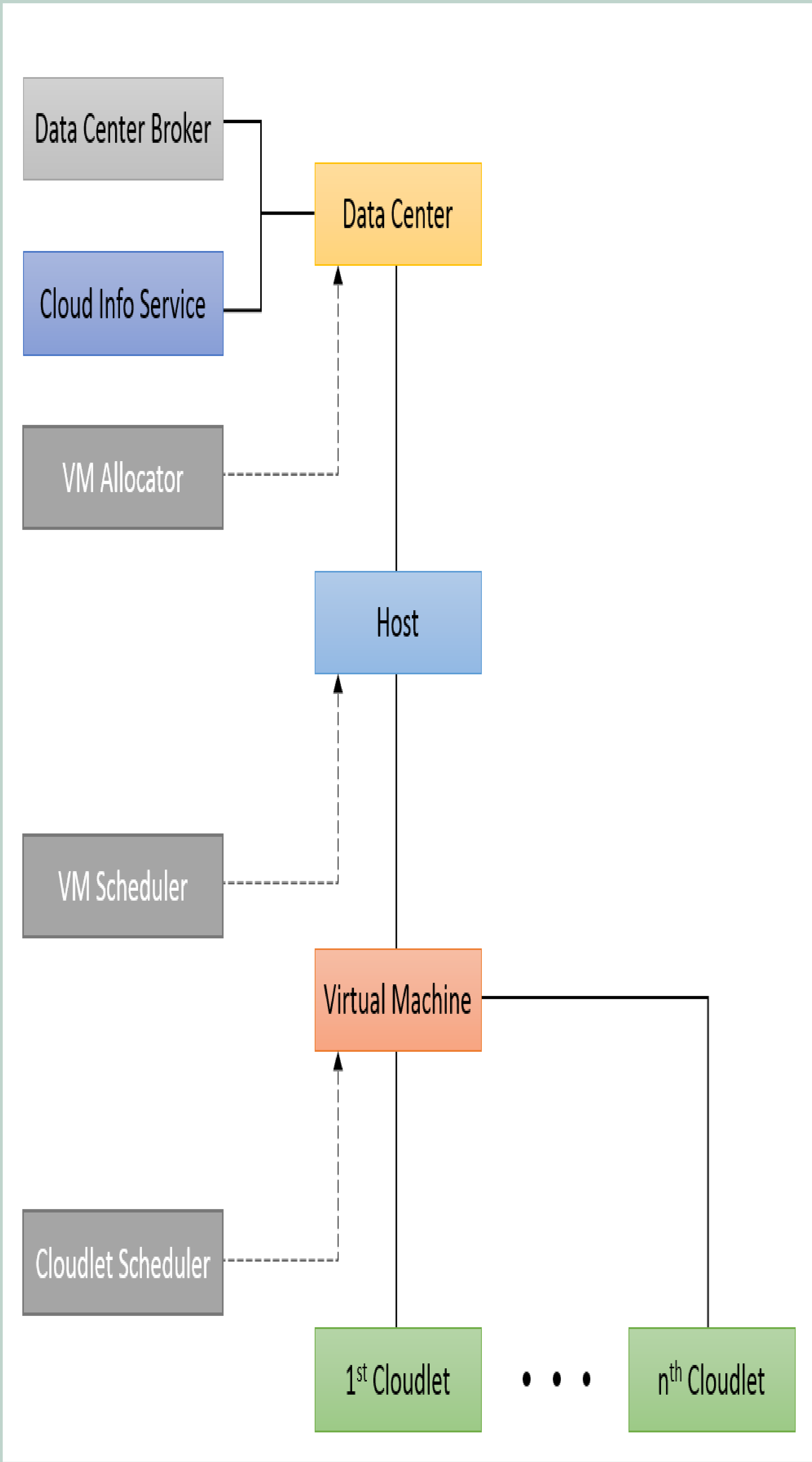
$$P(t) = \alpha_m P_m + \sum_{i=1}^{n} \alpha_i P_i$$

where $\alpha_m + \sum_{i=1}^{n} \alpha_i = 1$ and P_n is the LPM predicted value for different application services.

- The following flow chart illustrates the technique to be used for predicting the next-time recourse utilization when multiple application services are using the recourses within the same cloud datacenter.



IMPLEMENTATION



REFERENCES

- M. Dabbagh, B. Hamdaoui, M. Guizani, and A. Rayes, "Towards energy-efficient cloud computing: Prediction, consolidation, and overcommitment," IEEE Network Magazine, vol. 29, no. 2, 2015.
- Buyya, R., Beloglazov, A., & Abawajy, J., "Energy-efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges," Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), Las Vegas, USA, July 12-15, 2010.
- Choi, K., Soma, R., & Pedram, M. (2005). Fine-grained dynamic voltage and frequency scaling for precise energy and performance tradeoff based on the ratio of off-chip access to on-chip computation times. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, 24(1), 18-28.
- A. Beloglazov and R. Buyya. "Energy Efficient Resource Management in Virtualized Cloud Data Centers," 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, Melbourne, Australia, 2010, pp. 826-831.
- Shi, Y., Jiang, X., & Ye, K., An energy-efficient scheme for cloud resource provisioning based on cloudsim, 2011 IEEE International Conference on Cluster Computing (CLUSTER), pp. 595-599, 2011.
- Magi, Carlo, Jouni Pohjalainen, Tom Bäckström, and Paavo Alku. "Stabilised weighted linear prediction," Speech Communication 51, no. 5 (2009): 401-411.