

PoWER - Prediction of Workload for Energy Efficient Relocation of Virtual Machines

Kashifuddin Qazi, Yang Li, and Andrew Sohn

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ

Abstract

Virtual Machines (VM) offer data center owners the option to lease computational resources like CPU cycles, Memory, Disk space and Network bandwidth to end-users. An important consideration in this scenario is the optimal usage of the resources (CPU cycles, Memory, Block I/O and Network Bandwidth) of the physical machines that make up the cloud or 'machine-farms'. At any given time, the machines should not be overloaded (to ensure certain QoS requirements are met) and at the same time a minimum number of machines should be running (to conserve energy). The loads on individual VMs residing on these machines is, in fact, not absolutely random. Certain patterns can be found that can help the data center owners arrange the VMs on the physical machines such that both of the above conditions are met (minimum number of machines running without any being overloaded). In this work we propose a framework, PoWER that tries to intelligently predict the behavior of the cluster based on its history and then accordingly distributes VMs in the cluster and turns off unused Physical Machines, thus saving energy. Central to our framework are concepts of Chaos Theory that make our framework indifferent to the type of loads and inherent cycles in them as opposed to other current prediction algorithms. We also test this framework on our testbed cluster and analyze its performance. We demonstrate that PoWER performs better than another FFT-based time series method in predicting VM loads and freeing resources on Physical Machines for our test loads.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

SoCC'13, 1–3 Oct. 2013, Santa Clara, California, USA.

ACM 978-1-4503-2428-1/13/10.

<http://dx.doi.org/10.1145/2523616.2525938>

1 Design and Evaluation

Currently, the most promising methods to predict VM resource workloads for consolidation or resource management are methods like those based on FFT [6]. The problem we see with these tools is that even though they give a general idea of the most dominant cycle in the load, they ignore the fact that several other cycles could exist within the load. For example, even after observing a large number of weekly cycles, the major cycle calculated is the daily cycle, i.e. the VM has low loads every night and high loads throughout the rest of the day. However, this view ignores the fact that the VM has a light load during weekends compared to the rest of the week. Missing the weekend cycles (and their periods of low loads) could mean lower level of consolidation where more energy saving could have been potentially possible. Additionally, for loads that do not have an obvious cycle (like most real world loads), these methods fall short.

An evaluation of trace-data collected from NASA's website [3] and the World Cup website [5] shows that the loads are chaotic. We propose that using predictions based on Chaos Theory can achieve better results.

PoWER uses Chaos Theory based predictions using [1], [2], and [4]. Using these predictions, it employs live migration to pack as many VMs as possible in individual physical machines such that no overloads take place for at least a certain time in the future. Because of this consolidation, physical machines (and resources) are freed up in the cluster. These physical machines can then be either turned off (to save energy) or used to lease resources to new VMs.

Based on our initial experiments on 48 VMs, PoWER has a prediction accuracy of 89% as opposed to an FFT-based method which has an accuracy of 75%. Additionally, during consolidation it is able to free up about 80% of the total possible physical machines that could be freed up, as compared to 60% by the FFT-based method. Finally, on an average, our framework has 0% overloaded physical machines as opposed to 10% by FFT-based method.

2 References

- [1] Basharat, A., and Shah, M. 2009. Time series prediction by chaotic modeling of nonlinear dynamical systems. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1941–1948 (Oct. 2009).
- [2] Fraser, A. M., and Swinney, H. L. 1986. Independent coordinates for strange attractors from mutual information. *Physical review A*, 33(2):1134 (Feb. 1986).
- [3] Dumoulin, J., and Arlitt, M. "nasa web site http request logs" <http://http://ita.ee.lbl.gov/>, 1995.
- [4] Cao, L. 1997. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, 110(1):43–50 (Dec. 1997).
- [5] Arlitt, M., and Jin, T. 1998. "1998 world cup web site access logs" available at <http://www.acm.org/sigcomm/ita/> (Aug. 1998).
- [6] Gong, Z., Gu, X., and Wilkes, J. 2010. Press: Predictive elastic resource scaling for cloud systems. In *2010 IEEE International Conference on Network and Service Management (CNSM) pages 9–16* (Dec. 2010).