The University of Louisiana at Lafayette

Center for Advanced Computer Studies
P.O. Box 44330
Lafayette, Louisiana 70504-4330
(337) 482-6338
awlewis@cacs.louisiana.edu

November 16, 2011

Professor Tom Conte
Editor-in-Chief
ACM Transactions on Architecture Code Optimization

Dear Professor Conte:

We thank the reviewers for the time and effort required for their thorough reviews of our submission and greatly appreciate their candid and valuable comments, which clearly strengthen our manuscript. We have taken all comments into full consideration, addressing them in the revised manuscript, as detailed in sequence. Please find separate response documents, one for each of the reviewers, in the enclosed attachments to this cover letter.

Respectfully Yours,

Adam Wade Lewis

encl: (1) Response to Reviewer #1, (2) Response to Reviewer #2, and (3) Response to Reviewer #3

TACO-2010-27: Review Response: Reviewer #1

First of all, we greatly appreciate your time and effort put toward your candid and informative review, which helps to improve the quality of our submission. The following lists our responses to your concerns and comments made on our initial manuscript. Hopefully, we have addressed all your concerns and comments to your satisfaction.

First Comment/Concern:

> *First, the authors have to address how their work is related to the papers I have listed. Since power modeling of servers is a well-trodden area, the authors have to highlight how their contribution advances the state of the art/knowledge.*
> *Second, related to the above point, since the per component models are not new, the novelty has to be in the way the power modeling itself is done and the additional benefits and insights such a modeling approach will provide. While the idea of a Chaotic Attractor Predictor appears to be novel, none of the quantitative results are convincing that such a predictor is an improvement over the prior approaches. The authors either need to quantitatively compare, or at least provide a strong qualitative discussion about how the CAP predictor advances the state-of-the-knowledge on power modeling.*

Responses:

We agree with your point, recognizing that there is a large body of prior work pertinent to this submission. However, this area is yet to be well understood, despite being well-trodden. While the use of linear auto-regression for constructing power models is widespread in earlier work, it is demonstrated in our work and also supported by recent publications [Kansal et al. 2010; Tsirogiannis et al. 2010; McCullough et al. 2011; Hsu and Poole 2011] that significant issues exist with linear regression for power modeling. It is a major goal of our work to take a different approach aiming to arrive at better power modeling by considering the problem from a dynamic standpoint to better capture the chaotic non-linear dynamics introduced by hardware and OS measures. To address your concerns and comments on this point, we have made changes to Section 1 (Paragraphs 1 and 3) and Section 2 (Paragraphs 4, 5, and 6) for clarifying our position in this regard.

Second Comment/Concern:

> *There is a large body of literature on modeling power consumption (especially servers) that the authors have overlooked.*
> *There has been a lot of work on modeling the power and temperature of the various components considered in this paper [Skadron et al., ISCA'03], [Gurumurthi et al., ISCA'05], [Liu et al., DAC'08]. [Mesa-Martinez, ASPLOS'10]. The models used by the authors are much more simplistic those in these prior work.*
> *The authors' claim that there is no work that has looked at the power profiles of NUMA based architectures is incorrect. See [Ware et al., HPCA'10].*
> *There has also been prior work on using statistical methods to model power consumption [Powell et al., HPCA'09].*
> *Finally, there has been prior work on modeling the power and temperature of servers [Heath et al., ASPLOS'06].*

Responses:

Thank you for pointing out those relevant publications, which are mostly included in Section 2 with appropriate discussion. Specifically, they are handled as follows.

[**Skadron et al. 2003; Skadron et al. 2004**] Agreed. This reference and its discussion have been added to Section 2 of this revision.

[**Liu et al. 2008**] Agreed. This reference has been discussed in Section 2. It should be noted that our work uses the same power model as this reference, with both utilizing the DRAM power model found in [Micron, Inc. 2007].

[**Gurumurthi et al. 2005**] Agreed. This reference plus associated discussion has been added to Section 2 and Section 3.3 (Paragraph 1).

[**Ware et al. 2010**] Agreed. This reference plus associated discussion has been added to Section 2. Like [Ware et al. 2010], our manuscript also considered [Brochard et al. 2010] and [Rajamani et al. 2010] as relevant prior work.

[**Mesa-Martinez et al. 2010**] Agreed and noted in Section 2 that [Mesa-Martinez et al. 2007] was equally relevant to our work as well.

[**Powell et al. 2009**] While similar to our manuscript coverage, this reference resorted to linear regression and evidenced much of ill-behavior reported in our manuscript, as noted in Section 2 of this revision.

[**Heath et al. 2006**] This work has been included and discussed in Section 2 of the revised manuscript.

TACO-2010-27: Review Response: Reviewer #2

First of all, we greatly appreciate your time and effort put toward your candid and informative review, which helps to improve the quality of our submission. The following lists our responses to your concerns and comments made on our initial manuscript. Hopefully, we have addressed all your earlier concerns and comments raised earlier to your satisfaction.

First Comment/Concern:

> *I'd like the authors to discuss the time interval used in the evaluation (for the collection of the PeC and measurement of power). This is also needed in the appendix which has equations for energy, without discussing the time interval to which the equations apply. The time interval determines the possible applications (power capping) for using the power estimation. For example, actuators that use the parameters must work on the same time scale to be effective.*

Responses:

Agreed. A time interval of 5 seconds was used for our evaluation. This value is driven by the sampling accuracy of the system utilities employed to collect the data. Discussion about concerning this fact was included in Section 4.2.2 (Paragraph 3, Page 16) and in the Appendix (Paragraph 1, Page 23).

Second Comment/Concern:

> *I could not follow the discussion on calibrating the CAP. The text claims that p=100 and 4 workloads were run to calibrate the model. I take this to mean that the 4 workloads together were divided into 100 time intervals (of many seconds or minutes?) and the PeC values and measured power were averaged over each interval to calibrate the model. Is this correct? It would be helpful to add a couple of sentences to be clear how you obtain the 100 vectors required for calibration.*

Response:

Agreed. More discussion was added to Section 4.2.2 (Paragraph 2, Page 16) for clarifying this process.

Third Comment/Concern:

> *The definition of E_{em} is very strange. E_{em} is defined as a component of server power, but the equation seems to contain the entire servers power in the form of $V(t) * I(t)$, which contains CPU power (and other components) which have been already accounted for. In fact, the fan power itself should be part of $V(t) * I(t)$, the DC power, so it is not clear why they are added together in the equation for P_{elect}.*

Response:

Agreed. Section 3.5 (Page 10) has been adjusted to clarify this issue.

Fourth Comment/Concern:

> *Please define SMPS.*

Response:

SMPS is an acronym for *Switched Mode Power Supply.*

Fifth Comment/Concern:

> *I was confused by Section 3.6 which claims that power is controlled, but the rest of the paper does not come back to how power is controlled or how controlling power is relevant for developing the model.*

Response:

Agreed. Section 3.6 has been deleted.

Sixth Comment/Concern:

> *I'd like a longer discussion of the step-wise process in section 5 for predicting the distant future. I thought that the function $\hat{f}$ was for estimating the power of the system for interval k, based on PeC measurements made during time interval k. How does one use this to get power estimates for which PeC has not yet been measured? Are the values for PeC also being predicted?*

Response:

Agreed. Section 5 (Paragraph 1, Page 18) and 5.2 (Paragraph 2, Page 20) have been adjusted to clarify this process.

Seventh Comment/Concern:

> *I found the equation for $E_{intel_{proc}}$ (in the appendix) surprising because the coefficients for the temperature of each core is so different (more than a factor of 70x). It looks like core 1 does not contribute much to the power of the processor, which is hard to believe. I would expect them to be more balanced as in the AMD processor. If all the workload were scheduled on core 1 (core 0 disabled), would these equations still hold?*

Responses:

An interesting observation, as we suspect that this is an indication of an architectural difference between the processors and how Solaris carries out load-balancing. It seems that the OS design favors balancing workload in the fashion observed by you. Furthermore, it is very difficult to reason what is implied by these equations, given the difficulty in interpreting the physical meaning of regression coefficients. We would point you to [McCullough et al. 2011] for a discussion of this topic. Additional text has been added to the Appendix of this revision (Paragraph 2, Page 24), summarizing some of this work and how it is compared with our case.

TACO-2010-27: Review Response: Reviewer #3

First of all, we greatly appreciate your time and effort put toward your candid and informative review, which helps to improve the quality of our submission. The following lists our responses to your concerns and comments made on our initial manuscript. Hopefully, we have addressed all your earlier concerns and comments raised earlier to your satisfaction.

First Comment/Concern:

> *Majority of the paper is devoted to introducing the power model, subsystem models, CAP, and a variety of other issues such as timeseries forecasting, linearity issues, DC vs. AC power distribution and alternative performance counters for improving power estimation. After these a few examples of estimated and total power is shown for some SPEC CPU benchmarks, which do not help us (i) validate any of the assumptions for subsystem models; and (ii) build any intuition to how the CAP better understands and predicts power behavior.*
> *I recommend reducing some of the lengthy discussions at the beginning of this work and removing some tangential points (i.e., AC vs. DC, additional counters). It would be much more useful to use the remaining space to expand on the observed results and to provide some insights to why the CAP method is a better model for system power estimation.*

Responses:

We have attempted to address the concerns and comments made by all reviewers in this revision. Per the concerns of another reviewer, this revision adds a few short paragraphs in Sections 1 and 2 to clarify the unique contributions of our work, and to include and discuss extra prior articles. According to your recommendation, we have trimmed down the first two sections in our original manuscript, including the removal of Section 3.6 and its associated references.

Second Concern/Comment:

> *First, I feel the title of this work is somewhat vague. What does "time-series approximation of energy consumption estimation" mean? Do you fit an estimated energy consumption to a timeseries to predict future power behavior? Are you proposing CAP as a timeseries method for estimated power, or are you proposing an instantaneous energy estimation method based on performance counters; or both? This is not clearly explained in the paper. Second, what is "based on server workload"? It seems the evaluations are only for Spec CPU benchmarks.*

Responses:

We propose an instantaneous method, approximated by CAP, for estimating power consumption. Adjustments have been made to Sections 1 (Paragraphs 2, 3, and 4 on Page 2) and 2 (Paragraphs 1, 2, 5, 6, and 7 on Pages 3–6) to clarify the relationship between the title and the contents.

Third Concern/Comment:

> *What do you specifically propose for NUMA that prior work omits? How do you evaluate the improvements compared to a non-NUMA model?*

Responses:

Our model relies on tracking data movement across the system as a means for estimating energy consumption. The model component for processors involves estimating the amount of data transmitted on those paths between cores. Adjustments have been made to Sections 1 (Paragraph 3, Page 2) and 3.1 (Paragraph 2, Page 7).

Fourth Concern/Comment:

> *I am also not sure whether the complexity of CAP is commonly necessary for system power estimation. Several prior system-level power estimation methods show very good accuracy with very simple models, such as utilization-based estimations. Why do they not work well in this scenario?*

Responses:

While CAP is more complex than its earlier counterparts (for example, [Economou et al. 2006]), it gains an advantage of exhibiting a better fit to the dynamics of the system. Our model captures both the non-linear and chaotic aspects of the underlying dynamic system better than others. We have enhanced Section 1 (throughout Pages 2 and 3) and Section 2 (Paragraphs 5, 6, and 7 on Pages 5 and 6) of this revision for clarification, highlighting the need of a chaotic model to estimate power consumption more accurately even over short and varying execution time intervals.

Fifth Concern/Comment:

> *You highlight that your approach accounts for thermal effects in power estimation. Does this only pertain to processor power estimation, or for overall system?*

Response:

It focuses on the overall system. This revision has clarified this point in Section 1 (Paragraphs 3 and 4).

Sixth Concern/Comment:

> *The introduction mentions that prior work "assumes a linear relationship between dependent variables and previous data points". Do you mean (i) they predict future system behavior based on a linear model; or (ii) they use a linear combination of system measurements to predict instantaneous power? For (i) there are some prior studies that use statistical or pattern based models to predict future system power/performance behavior, as it is commonly known that dynamically-varying workload behavior impacts system behavior, which does not necessarily vary in a linear way.*

Responses:

We meant the first case in your above list. Additional discussion and representative references have been included to Section 1 (Paragraph 1, Page 3) and Section 2 (Paragraph 5, Pages 5 – 6) in an attempt to clarify this issue.

Set of Concerns and Comments:

1. P_{proc}

> 1. *The derived processor power model (intel) shows a 100X difference in sensitivity to two thermal sensors. Is there any reason for this discrepancy?*

2. *The model seems quite oblivious to workload and very dependent on temperature. Considering temperature does not vary instantaneously, while power does, does this model remain valid for different workload intensities? For example an idle period after a long running cpu burn can result in a higher estimated power than a long-running idle period. Or the order of different workload intensities can impact the estimated power.*

Response:

Your observation about the sensitivity of the linear models to temperature, in fact, highlights a point that we are showing about those models: failing to take into account the chaotic and non-linear dynamics of the system, one will see odd and unexpected behavior out of the models, which over-compensate values in certain time windows aiming to fit a straight-line approximation to a non-linear curve. We have added discussion to Section 2 (Paragraphs 5 and 6 on Page 4) and the Appendix of this revision (Paragraph 4, Page 24) discussing this point.

2. E_{hdd} **&** E_{board}

I think the HDD power would also depend random vs. sequential writes and the involvement of disk cache. For the board, I expected more or less constant power, do you see any significant variations from different workloads?

Responses:

E_{hdd}**:** There are three approaches to approximating hard disk energy consumption: an analytical model as proposed in [Gurumurthi et al. 2005], an interrupt-based model as proposed in [Bircher and John 2011], and the operating system metric approach as adopted in our work. Our model represents a reasonable first-order approximation to this value.

E_{board}**:** It is possible to treat this value as a constant (like what was done in [Bircher and John 2011]). Given that the value may vary from motherboard to motherboard, it is our position, however, to consider E_{board} as one model component, with its value determined through experimentation. This position is now noted appropriately in this revision (Section 3.4, Page 10).

Seventh Concern/Comment:

As I have mentioned previously, it is also quite hard to gauge the accuracy of the subsystem models without any validation, and particularly for some of the components Spec CPU would be limiting as it is designed not to exercise components below memory.

Responses:

The three important points to be emphasized in this work include:

1. the necessary requirement for chaotic behavior is that the model can be expressed in terms of a dynamic system with differential equations,
2. this work focuses on the quality of the time series approximation provided by the model, and
3. it seems statistically risky to place too much weight on the physical interpretation of auto-regressive model components, regardless of whether the model is linear or non-linear (see [McCullough et al. 2011]).

In terms of utilizing the SPEC benchmarks, it is important to remember that we are looking at the system energy consumption. While SPEC CPU 2006, as a performance benchmarking suite, focuses on components above memory, its execution does exercise the entire system when virtual memory, benchmark bookkeeping, and other interfaces to the operating system are considered as well. In that regard, we have included some description in Section 5.1 (Paragraph 1, page 16–17) to outline the criteria followed for selecting our benchmarks.

Eighth Concern/Comment:

> *The observations on the chaotic nature of power behavior are quite interesting. Are the two conditions with Lyapunov and Hurst parameter sufficient to consider power behavior chaotic? Can you provide some intuition behind these? I cannot see how the two requisites (high sensitivity to initial condition & dense period orbits) hold necessarily true for power behavior?*

Response:

Agreed. Discussion has been added to Section 2 (Paragraphs 5, 6, and 7), stating the necessary conditions for exhibiting chaotic behavior. Also, references on power electronics which inspired this work are included in Section 1 (Paragraph 4, Page 3) ([Hamill et al. 1997; Tse and Di Bernardo 2002])

Ninth Concern/Comment:

> *Power measurements: The measured power for both systems vary between 60-70W from idle-active. This seems rather low for both servers, and the delta seems very low for the entire system including CPUs, memory, and so on. Could you please provide more details on the system configurations and what is reported.*

Response:

We have consistently seen this kind of measured power draws on the class of commodity servers used in our experimentation, and the measured readings are in line with the data provided by server manufacturers. Our measurements were confirmed by using multiple power measurement devices from different vendors.

Tenth Comment:

> *Please consider reviewing the paper for language. Here are a few typos that caught my eye:*
>
> - *valid valid* -->*validate*
> - *CAT* -->*CAP*
> - *5.2: number of past observations –¿ number of future observations*

Responses:

Thanks. Those typos have been corrected in this revision.

REFERENCES

BIRCHER, W. AND JOHN, L. 2011. Complete System Power Estimation using Processor Performance Events. *IEEE Transactions on Computers Pre-print*.

BROCHARD, L., PANDA, R., AND VEMUGANTI, S. 2010. Optimizing Performance and Energy of HPC Applications on POWER7. *Computer Science – Research and Development 25*, 135–140. 10.1007/s00450-010-0123-3.

ECONOMOU, D., RIVOIRE, S., KOZYRAKIS, C., AND RANGANATHAN, P. 2006. Full-System Power Analysis and Modeling for Server Environments. In *Proc. of the 2008 Workshop on Modeling Benchmarking and Simulation*.

GURUMURTHI, S., SIVASUBRAMANIAM, A., AND NATARAJAN, V. 2005. Disk Drive Roadmap From the Thermal Perspective: A Case For Dynamic Thermal Management. In *Proc. of the 32nd Int'l. Symp. on Computer Architecture*. 38 – 49.

HAMILL, D., DEANE, J., AND ASTON, P. 1997. Some Applications of Chaos In Power Converters. In *IEEE Colloquium on Update on New Power Electronic Techniques (Digest No: 1997/091)*.

HEATH, T., CENTENO, A. P., GEORGE, P., RAMOS, L., JALURIA, Y., AND BIANCHINI, R. 2006. Mercury and Freon: Temperature Emulation and Management for Server Systems. In *ASPLOS-XII: Proc. of the 12th Int'l Conference on Architectural Support for Programming Languages and Operating Systems*. ACM, New York, NY, USA, 106–116.

HSU, C.-H. AND POOLE, S. 2011. Power Signature Analysis of the SPECpower_ssj2008 Benchmark. In *Proc. of the 2011 IEEE Int'l. Symp. on Performance Analysis of Systems and Software*. 227 –236.

KANSAL, A., ZHAO, F., LIU, J., KOTHARI, N., AND BHATTACHARYA, A. A. 2010. Virtual Machine Power Metering and Provisioning. In *Proc. of the 1st ACM Symp. on Cloud Computing*. SoCC '10. ACM, New York, NY, USA, 39–50.

LIU, S., MEMIK, S. O., ZHANG, Y., AND MEMIK, G. 2008. A Power and Temperature Aware DRAM Architecture. In *Proc. of the 45th Annual Design Automation Conference*. DAC '08. ACM, New York, NY, USA, 878–883.

MCCULLOUGH, J. C., AGARWAL, Y., CHANDRASHEKAR, J., KUPPUSWAMY, S., SNOEREN, A. C., AND GUPTA, R. K. 2011. Evaluating the Effectiveness of Model-Based Power Characterization. In *USENIX ATC '11*.

MESA-MARTINEZ, F., NAYFACH-BATTILANA, J., AND RENAU, J. 2007. Power Model Validation Through Thermal Measurements. *Proc. of the 34th Int'l Conf. on Computer Architecture*, 302–311.

MESA-MARTINEZ, F. J., ARDESTANI, E. K., AND RENAU, J. 2010. Characterizing Processor Thermal Behavior. *SIGARCH Comput. Archit. News 38*, 193–204.

MICRON, INC. 2007. Calculating Memory System Power for DDR3. Tech. Note TN41_01DDR3 Rev.B, Micron, Inc. August.

POWELL, M., BISWAS, A., EMER, J., MUKHERJEE, S., SHEIKH, B., AND YARDI, S. 2009. CAMP: A Technique to Estimate Per-structure Power At Run-time Using A Few Simple Parameters. In *Proc. of the IEEE 15th Int'l. Symp. on High Performance Computer Architecture*. 289 –300.

RAJAMANI, K., RAWSON, F., WARE, M., HANSON, H., CARTER, J., ROSEDAHL, T., GEISSLER, A., SILVA, G., AND HUA, H. 2010. Power-performance management on an ibm power7 server. In *Proceedings of the 16th ACM/IEEE international symposium on Low power electronics and design*. ISLPED '10. ACM, New York, NY, USA, 201–206.

SKADRON, K., STAN, M., HUANG, W., VELUSAMY, S., SANKARANARAYANAN, K., AND TARJAN, D. 2003. Temperature-aware Microarchitecture. In *Proc. of the 30th Int'l. Symp. on Computer Architecture*. 2 – 13.

SKADRON, K., STAN, M. R., SANKARANARAYANAN, K., HUANG, W., VELUSAMY, S., AND TARJAN, D. 2004. Temperature-aware Microarchitecture: Modeling and Implementation. *ACM Transactions on Architecture and Code Optimization 1,* 1, 94–125.

TSE, C. AND DI BERNARDO, M. 2002. Complex Behavior In Switching Power Converters. *Proceedings of the IEEE 90,* 5 (May), 768 –781.

TSIROGIANNIS, D., HARIZOPOULOS, S., AND SHAH, M. A. 2010. Analyzing the Energy Efficiency of A Database Server. In *Proc. of the 2010 Int'l. Conf. on Management of Data*. SIGMOD '10. ACM, New York, NY, USA, 231–242.

WARE, M., RAJAMANI, K., FLOYD, M., BROCK, B., RUBIO, J., RAWSON, F., AND CARTER, J. 2010. Architecting for Power Management: The IBM POWER7 Approach. In *Proc. of the IEEE 16th Int'l. Symp. on High Performance Computer Architecture*. 1 –11.