2015 4th International Conference on Computer Science and Network Technology (ICCSNT 2015)

Workload-Based Prediction of CPU Temperature and Usage for Small-Scale Distributed Systems

Raju Ahmed Shetu[1], Tarik Toha[2], Mohammad Mosiur Rahman Lunar[3], Novia Nurain[4], and A. B. M. Alim Al Islam[5]
[1, 2, 4, 5]Department of CSE, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[3]Department of CSE, University of Nebraska, Lincoln, USA
Email: {[1]shetu2153, [2]tarik.toha}@gmail.com,[3]mlunar@cse.unl.edu, [4]novia@cse.uiu.ac.bd, [5]alim_razi@cse.buet.ac.bd

***Abstract*—The recent boost in the usage of high-performance computing systems in small research environments, such as those found at many universities, stipulates the need of small-scale distributed systems. Owning to the rapid growth in both computing power and heat, development of proper thermal and resource management becomes crucial concern of the research community along with the vendors to ensure efficiency for such systems. Moreover, an accurate and relatively fast strategy is needed for adaptation of different sizes of workload in such systems. Therefore, in this paper, we focus on developing simple prediction models of CPU temperature and usage for the systems. We investigate impacts of macro-level parameters such as the number of machines and different sizes of workload on CPU temperature and usage via real experiment. Our experimental results reveal that for a certain size of workload, the variation in CPU temperature and usage is minimal in response to a change in the number of machines, which does not hold in the reverse way. Hence, we develop workload-based prediction models for CPU temperature and usage. We evaluate the accuracy of our models by comparing the values calculated based on these models against the measurements found from real implementation.**

Keywords*—*Distributed systems, CPU temperature, CPU usage, prediction models.

I. Introduction

Distributed systems are widely used in high-performance computing environments having hundreds or thousands of networked machines connected together for processing Terabytes or larger data sets. However, the recent proliferation of various promising applications such as automated banking system [1], tracking roaming cellular telephones [2], air-traffic control systems [3], global positioning systems [4], research and development projects [5], etc., boost the emergence of small-scale distributed systems. Unfortunately, systems such as processing log files, word count, sorting numbers, etc., [6] that deal with larger data sets fail to scale down to smaller environments comprising dozen of machines posing computing challenges. Hence, the usage of a distributed system in small-scale research environments, especially those available at many universities and even commercial sectors in third-world countries, captures interests of the research communities now-a-days.

Since both computing power and heat are growing for distributing systems, effective thermal management becomes evident to ensure efficacy and efficiency of the systems. Additionally, proper usage of resources such as CPU clocks and memory is necessary to ensure reliable QoS. Therefore, in this paper, we develop mathematical models for capturing trends in CPU temperature and usage for distributed systems mostly focusing on its small-scale implementation. The proposed models can be used to forecast effectively the future CPU temperature and usage. Such predictions can be used to derive temperature-efficient resource allocation policies and workload scheduling algorithms considering QoS expectations.

There exist several CPU temperature prediction and management techniques in the literature [7]–[9]. These techniques derive mathematical models based on observed temperatures generated from the execution of specific applications using auto-regressive moving average and generic polynomial function. Besides, studies in [10], [11] perform temperature prediction based on simplified thermodynamic laws and micro-level parameters such as core temperature/utilization, and airflow velocity. However, these approaches rely on steady-state thermal models, which fail to reflect the evolution of temperature due to a change in the sizes of workload. Therefore, in this paper, we focus on proposing a mathematical model for performing temperature prediction based on macro-level parameters such as sizes of workload and the number of machines instead of the micro-level parameters.

On the hand, recent studies in [16], [17] focus on estimation of CPU usage. These studies perform predictions using machine learning (ML) techniques based on provided training data set. However, these techniques often suffer low prediction fidelity due to insufficient training data. Therefore, in this paper, we devise a mathematical model for CPU usage prediction that can overcome this limitation.

Our contributions in this paper are as follows:

- We investigate the variations in CPU temperature and usage in response to a change in the number of machines for different sizes of workload via real experiment.
- Based on our investigation, we propose two different mathematical models for workload-based prediction of CPU temperature and usage.
- Finally, we evaluate accuracy of our proposed models by comparing the values calculated based on these models against the measurements obtained from real implementation.

II. Related Work

There are a number of studies found in the literature that perform thermal prediction for data centers using mathematical models. These models are mostly developed through exploiting auto-regressive moving average (ARMA) [7], [8] and generic polynomial function [9]. Here, the predictions of ARMA model are fairly accurate, however, there lies a possibility that the predictor could miss some of the prediction opportunities during training phase. Besides, studies in [10], [11] propose thermal model emulating simple thermodynamic laws, velocity of airflow, and CPU temperature/utilization. Additionally, the study in [12] utilizes a neural network to learn and predict temperature under static workload. However, all these approaches ([10]–[12]) fail to adapt to the dynamic workload (i.e., different sizes of workload). Moreover, such approaches suffer from low prediction fidelity due to insufficient training data that may occur in case of critical thermal emergencies.

Furthermore, several existing studies [13], [14] utilize a popular tool called CFD modeling for capturing the fluid patterns in the applications to eventually predict CPU temperature. Besides, the study in [14] proposes a thermal forecasting model, called ThermoCast, to predict temperature around servers in data center based on principles from thermodynamics and fluid mechanics. However, these models depend on several restrictions on airflow dynamics and fail to resemble the diversity of data center environments. Nonetheless, the study in [15] introduces a workload-independent temperature predictor based on the band limited property of the temperature. This predictor is not sensitive to a change in the size of workload.

On the other hand, recent studies [16], [17] perform resource (CPU usage) estimation and prediction using machine learning techniques based on provided training data sets. However, again, these techniques often suffer low prediction fidelity in case of having insufficient training data.

It is quite evident from the above discussion, most of the existing studies in the literature focus on developing thermal and usage prediction models based on micro-level parameters (core temperature/utilization, airflow velocity, dynamic voltage and frequency scaling per core and processor) and static workload. Consequently, the impacts of macro-level parameters such as dynamic workload (i.e., different sizes of workload) and the number of machines on the prediction model are yet to be devised in the literature. Therefore, in this paper, we explore the variations in CPU temperature and usage in response to a change in the number of machines for different sizes of workloads to develop workload-based prediction models.

III. Experimental Analysis

In our study, we conduct experiments on a custom-built small-scale distributed system. In our experiment, we focus on CPU temperature and CPU usage. First, we briefly illustrate our experimental setup and then we present our results.

A. Experimental Setup

We conduct our experiment in a laboratory environment. Here, our experimental setup consists of 16 independent computers connected through a wired network. Each of the computers possesses a CPU of Intel Core 2™Duo CPU E4600 having 2.40GHz clock speed and a memory (RAM) of 2GB. We utilize Ubuntu 14.04 (x86) as the operating system. Here, as all the computers in our setup have the same configuration, our setup mimics a homogeneous setting. To further clarify our setup, we present a snapshot of the setup in Fig. 1.

We use Hadoop framework for conducting our experiment. Here, we perform an operation of counting words from a number of files. We vary the total size of the files, i.e., the size of workload, used in one experimental run from 2.9GB to 58.7GB. Additionally, we vary the number of machines in our distributed system from 6 to 16. For each workload and for each set of machines, we perform two experimental runs and take averages of the runs as our experimental results for CPU temperature and usage. We use Psensor [18] to monitor the CPU temperature, which supplies the average temperature of all the cores in the processor. Psensor also provides the measurements of the CPU usage. Next, we present our results obtained through this setup.

B. CPU Temperature and Usage for Varying Sizes of Workload

First, we vary the number of machines for different sizes of workload and analyze corresponding change in CPU temperature. We present outcome of such variation in Fig. 2. This figure shows that a certain number of machines can handle a maximum size of workload. For example, the setup of 6 machines can handle up to 29.3GB workload and it stops operating beyond this size. However, even there is a maximum bound on workload size for a certain number of machine, the figure demonstrates variations in CPU temperature over different sizes of workload. Nonetheless, an interesting finding of the figure is that for a certain size of workload, the variation in CPU temperature is minimal in response to a change in the number of machines. In addition, these results confirm the validity of our approach of predicting CPU temperature based on the sizes of workloads.

To further illustrate the impact of the size of workload on CPU temperature, we present Fig. 3. This figure again confirms significant response for changing the size of workload while a minimal responds to a change in the number of machines. Additionally, we plot the impact of a change in the size of workload on CPU usage in Fig. 4. This figure demonstrates clear, however similar, trend on the variation in CPU usage in response to an increase in the size of workload. Next, we propose two different models to capture the two different trends we observe in Fig. 3 and Fig. 4 in response to an increase in the size of workload.

IV. Proposed Models for Workload-based Prediction of CPU Temperature and Usage

We propose the following equation for predicting CPU temperature based on the size of workload:

$$t = P_1x^2 + P_2x + P_3 \quad (1)$$

Here, t denotes CPU temperature in °C and x denotes the size of workload in GB. We formulate this equation from the data presented in Fig. 3 using Matlab. The equation suggests that the CPU temperature of a small-scale distributed system can be modeled as a polynomial of the size of workload, where degree of the polynomial is 2. Besides, we find conditions on the coefficients of the polynomial for our case as follows:

Fig. 1: Laboratory environment for conducting our experiments

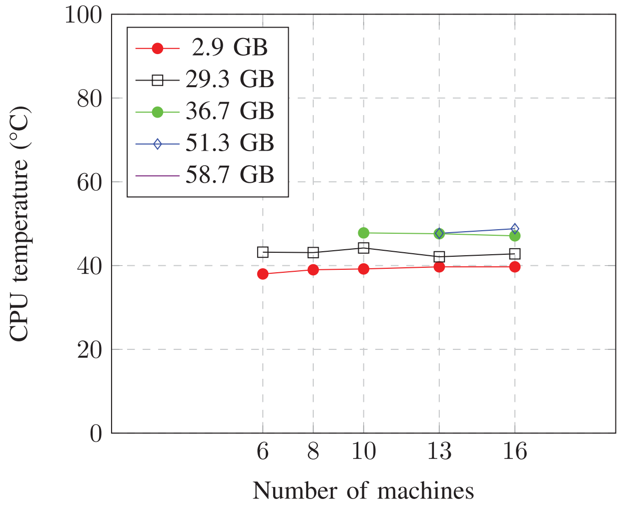

Fig. 2: Variation in CPU temperature in response to a change in the number of machines with different sizes of workload

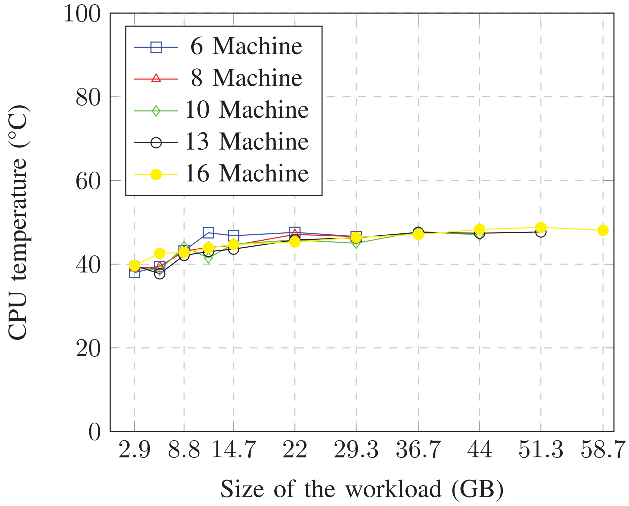

Fig. 3: Variation in CPU temperature in response to a change in the size of workload for different number of machines

$$P_1 = -0.003092, (-0.004831 \leq P_1 \leq -0.001353)$$
$$P_2 = 0.321, (0.2145 \leq P_2 \leq 0.4276)$$
$$P_3 = 40.05, (38.82 \leq P_3 \leq 41.27)$$

Next, we propose the following equation for predicting CPU usage based on the size of workload:

$$u = a_0 + \sum_{i=1}^{4} a_i cos(i\omega x) + b_i sin(i\omega x) \quad (2)$$

Here, u denotes CPU usage in % and x denotes the size of workload in GB. We formulate this equation from the data presented in Fig. 4 using Matlab. The equation suggests that, unlike the previous case of CPU temperature, CPU usage of a small-scale distributed system can be modeled as a complex equation combining both *sine* and *cosine* terms. Here, we find conditions on the coefficients used in the equation for our case as follows:

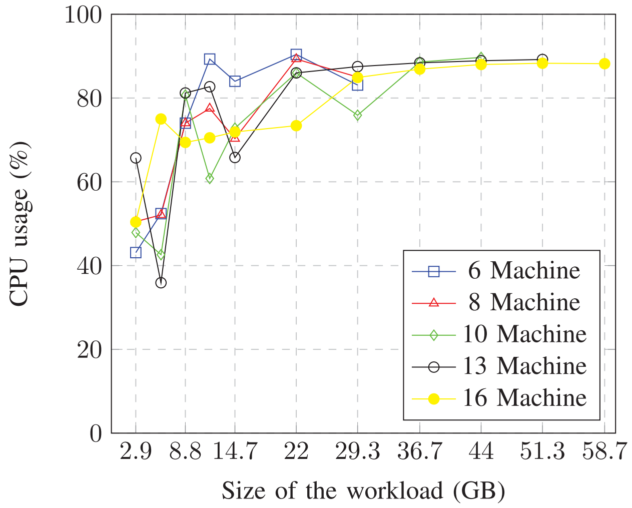

Fig. 4: Variation in CPU Usage in response to a change in the size of workload for different number of machines

$$\begin{aligned}
a_0 &= -9.298 \times 10^{10}, (-4.238 \times 10^{15} \leq a_0 \leq 4.237 \times 10^{15}) \\
a_1 &= 1.469 \times 10^{11}, (-6.717 \times 10^{15} \leq a_1 \leq 6.717 \times 10^{15}) \\
b_1 &= -2.379 \times 10^{10}, (-9.494 \times 10^{14} \leq b_1 \leq 9.494 \times 10^{14}) \\
a_2 &= -7.074 \times 10^{10}, (-3.265 \times 10^{15} \leq a_2 \leq 3.265 \times 10^{15}) \\
b_2 &= 2.352 \times 10^{10}, (-9.417 \times 10^{14} \leq b_2 \leq 9.418 \times 10^{14}) \\
a_3 &= 1.893 \times 10^{10}, (-8.887 \times 10^{14} \leq a_3 \leq 8.887 \times 10^{14}) \\
b_3 &= -9.888 \times 10^{9}, (-3.982 \times 10^{14} \leq b_3 \leq 3.981 \times 10^{14}) \\
a_4 &= -2.146 \times 10^{9}, (-1.035 \times 10^{14} \leq a_4 \leq 1.035 \times 10^{14}) \\
b_4 &= 1.604 \times 10^{9}, (-6.509 \times 10^{13} \leq b_4 \leq 6.51 \times 10^{13}) \\
\omega &= -0.005126(-29.21 \leq \omega \leq 29.2)
\end{aligned}$$

V. Validation of the Proposed Models

To validate accuracy of our proposed models, we simultaneously plot values achieved through the models and values obtained from our real experiments in the same graph. While presenting the graphs, we retain data for 13 and 16 machines to make matching between the data obtained from the model and the data obtained from our real experiment clearly visible.

Fig. 5 presents the graph containing data on CPU temperature. This graph exhibits a close match between the data obtained from the model and the data obtained from our real experiment. Besides, Fig. 6 presents the graph containing data on CPU usage. This graph shows that our model can mostly capture the trend exhibited by CPU usage. Here, model exhibits little deviation from the real data only for smaller sizes of workload.

VI. Applications of Our Models

In this paper, we focus on developing a workload-based mathematical model for predicting CPU temperature. Such forecasting model can be essential for preventing critical thermal emergencies such as server shutdown owing to over heating and cooling system failure. Our prediction model can also be helpful for devising dynamic thermal management scheme that manages fan speed by allocating the workload intelligently. Another promising application of our thermal predictor is constructing a stable scheduler that reduces the cooling costs by delineating a better thermal evolution. Besides, we envision airflow management system, HVAC (heating, ventilation and air conditioning) system design along with

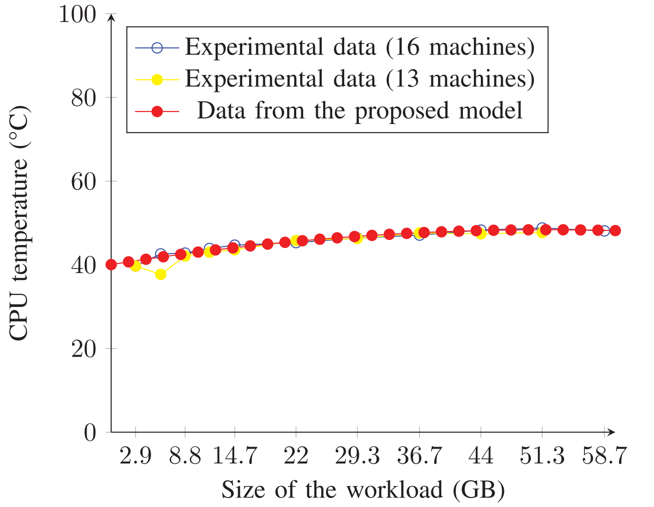

Fig. 5: Validation of the model proposed for CPU temperature

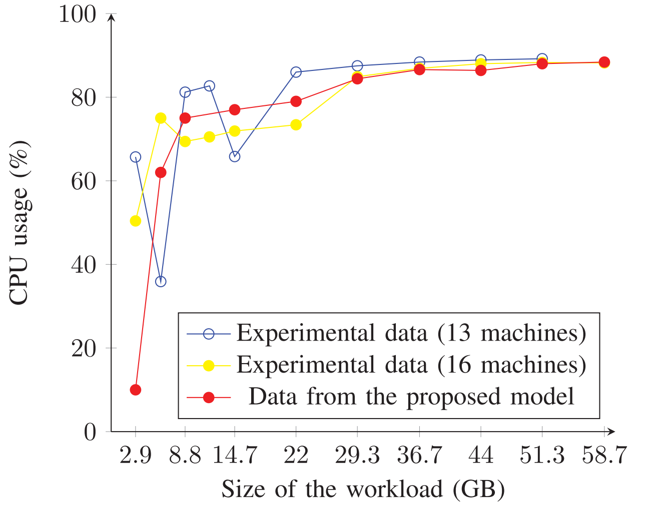

Fig. 6: Validation of the proposed model for CPU usage

workload distribution and optimization mechanisms as another potential application domain for our proposed workload-based CPU temperature predictor.

Furthermore, various workload management strategies (i.e., dynamic workload scheduling) requires a prediction module in order to determine whether or not to add more resources, allow or reject a new incoming workload, and rearrange the order of the workload. Our proposed workload-based CPU usage prediction model can be helpful for the generation of such management strategies.

VII. Conclusion and Future Work

The recent advancement of diverse promising applications such as automated banking system, air traffic control, research and development projects, etc., bolsters the emergence of small-scale distributed systems. Workload-based predictions of CPU temperature and usage are yet to be done in the literature even though such predictions are needed for cooling management, workload management, policy making, etc. Therefore, in this paper, we focused on developing workload-based prediction models for CPU temperature and usage. Here, we confirm accuracy of our proposed models by comparing the values estimated from our models against the measurements obtained from a real implementation. In future, we plan to make our proposed models more rigorous through investigating impacts of other parameters such as external temperature, number of cooling fans, and dynamic voltage and frequency scaling per core and processor, etc.

Acknowledgment

This work has been conducted at and partially supported by Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh. Besides, this work has been partially supported by the Ministry of Education, Government of the People's Republic of Bangladesh.

References

[1] Ginter, Karl L., Victor H. Shear, Francis J. Spahn, David M. Van Wie, and Robert P. Weber. "Trusted infrastructure support systems, methods and techniques for secure electronic commerce, electronic transactions, commerce process control and automation, distributed computing, and rights management." U.S. Patent 8,185,473, issued May 22, 2012.

[2] Stewart, Brett B., James Thompson, and Kathleen E. McClelland. "Distributed network communication system to provide wireless access to a computing device at a reduced rate." U.S. Patent 8,588,130, issued November 19, 2013.

[3] Qu, J., Yang, J., Liu, H., and Wu, R. "Clock Synchronization in Air Traffic Control System Using NTP". International Journal of Digital Content Technology and its Applications, vol. 7(1), pp. 344, 2013.

[4] Kopetz, H. "Real-time systems: design principles for distributed embedded applications". Springer Science and Business Media, 2011.

[5] Remington, J. "Distributed systems in small scale research environments: Hadoop and the EM algorithm." Doctoral dissertation, Colorado State University, 2011.

[6] Mapreduce tutorial, "http://hadoop.apache.org/common/docs/ current." Last accessed on October 23, 2010.

[7] Coskun, A., Rosing, T., and Gross, K. "Utilizing Predictors for Efficient Thermal Management in Multiprocessor SoCs." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 28(10), pp. 1503–1516, 2009.

[8] Coskun, A., Rosing, T., and Gross, K. "Proactive Temperature Management in MPSoCs." In International Symposium on Low Power Electronics and Design, pp. 165–170, 2008.

[9] Yeo, I., Liu, C., and Kim, E. "Predictive Dynamic Thermal Management for Multi core Systems." In Design Automation Conference, pp. 734–739, 2008.

[10] Heath, T., Centeno, A. T., George, P., Ramos, L., Jaluria, Y., and Bianchini, R. "Mercury and freon: Temperature emulation and management for server systems." In Proceedings of the 12^{th} International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), pp. 106–116, 2006.

[11] Ramos, L., and Bianchini. R. "C Oracle: Predictive thermal management for data centers." In IEEE 17^{th} International Symposium on High Performance Computer Architecture (HPCA), pp. 111–122, 2008.

[12] Moore, J., Chasey, J., and Ranganathanz, P. " Weatherman: Automated, online, and predictive thermal mapping and management." In Proceedings of the 3^{rd} IEEE ICAC, pp. 155–164, 2006.

[13] Choi, J., Kim An, Y., Srebric, J., Wang, Q., and Lee. J. "Modeling and managing thermal profiles of rack mounted servers with thermostat." In Proceedings of the 13th International Symposium on High-Performance Computer Architecture, pp. 205–215, 2007.

[14] Li, L., Liang, C. J. M., Liu, J., Nath, S., Terzis, A., and Faloutsos, C. "Thermocast: a cyber physical forecasting model for data centers." In ACM KDD, 2011.

[15] Ayoub, R., Indukuri, K., and Rosing, T. S. "Temperature Aware Dynamic Workload Scheduling in Multisocket CPU Servers." IEEE Transactions on computer aided design if integrated circuits and systems, vol. 30(9), pp. 1359, 2011.

[16] Li, Jiexing, et al. "Robust estimation of resource consumption for SQL queries using statistical techniques." Proceedings of the VLDB Endowment vol. 5(11), pp. 1555–1566, 2012.

[17] Ganapathi, Archana, et al. "Predicting multiple metrics for queries: Better decisions enabled by machine learning." Data Engineering, 2009. (ICDE'09). IEEE 25th International Conference on. IEEE, 2009.

[18] Psensor, "http://wpitchoune.net/blog/psensor/". Last Accessed on October 23, 2015.