Towards Extension of Data Centre Modelling Toolbox with Parameters Estimation

Yulia Berezovskaya, Chen-Wei Yang, Valeriy Vyatkin

Towards Extension of Data Centre Modelling Toolbox with Parameters Estimation

Yulia Berezovskaya[1], Chen-Wei Yang[1], Valeriy Vyatkin[1,2]

[1]Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden
[2]Department of Electrical Engineering and Automation, Aalto University, Finland
yulia.berezovskaya@ltu.se, chen-wei.yang@ltu.se, vyatkin@ieee.org

Abstract. Modern data centres consume a significant amount of electricity. Therefore, they require techniques for improving energy efficiency and reducing energy waste. The promising energy-saving methods are those, which adapt the system energy use based on resource requirements at run-time. These techniques require testing their performance, reliability and effect on power consumption in data centres. Generally, real data centres cannot be used as a test site because of such experiments may violate safety and security protocols. Therefore, examining the performance of different energy-saving strategies requires a model, which can replace the real data centre. The model is expected to accurately estimate the energy consumption of data centre components depending on their utilisation. This work presents a toolbox for data centre modelling. The toolbox is a set of building blocks representing individual components of a typical data centre. The paper concentrates on parameter estimation methods, which use data, collected from a real data centre and adjust parameters of building blocks so that the model represents the data centre most accurately. The paper also demonstrates the results of parameters estimation on an example of EDGE module of SICS ICE data centre located in Luleå, Sweden.

Keywords: Data centre, modelling, power consumption, parameter estimation, Matlab/Simulink

1 Introduction

Data centres are sizable consumers in the energy grid and comparable to industrial plants in terms of energy consumption [1]. Modern data centres use more and more renewable energy sources, which tend to intermittent energy production [2]. Data centres are expected to flexibly respond to changes in energy supply and possible restrictions from the energy grid and remain highly available and reliable. To meet these expectations, the promising approach is flexible control methods, which adapt the data centre energy use based on resource requirements during the data centre operation.

The subject of my PhD study is the development of energy-efficient control in data centres. Any data centre can be thought of as a distributed system of interdependent components, which are too complex for centralised management. As a part of my PhD research, a design of multi-agent control to reach energy efficiency in data centres was presented in [3]. Further elaborating the multi-agent control requires the data centre model, which is expected to substitute a real data centre as a system under control. In our previous work [4], we developed a modular Simulink toolbox, which is formed as a set of blocks for modelling individual data centre components such as servers, and components of a cooling system. The toolbox allows constructing the models of data centres of arbitrary configurations. The models are capable of estimating the energy consumption and predicting the thermal evolution inside the data centre.

Most of the blocks in the toolbox have their internal parameters, which depend on the type of modelled component. The accuracy of modelling results is mainly determined by the selection of those parameters. The current edition of the toolbox uses rough estimates of the parameters. The parameters estimation for the toolbox is not scalable as the parameters are only accurate for a specific datacentre configuration. Any other configuration requires a redefinition of model parameters, which can be achieved by either a) using datasheets of data centre components, or b) extracting the parameter values from the real data. This research work aims to introduce data-driven methods to estimate the modelling parameters for any given datacentre configuration. The work also considers an extension of the toolbox with the proposed methods.

This work deals with parameters of the toolbox block called the Server block. That block is responsible for the modelling of the main IT component. The main contributions of the work:

- presenting two data-driven procedures for parameters estimation: the regression-based estimation and the simulation-based optimisation;
- extension the toolbox with both procedures implemented as Matlab scripts;
- employment of both procedures to real data from the EDGE module of SICS ICE data centre located in Luleå, Sweden [5];
- demonstrating the pros and cons of both procedures.

The rest of the paper is organised as follows. Section 2 provides an overview of related works in the area of parameter estimation for data centre models as well as the relationship of the work to applied artificial intelligence. Section 3 describes the parameters, which require the estimation. Section 4 describes the general procedures for parameters estimation. Section 5 considers the results of parameters estimation and compares two proposed procedures. Section 6 gives the conclusion.

2 Related Works

This section considers works, which concentrate on approaches to modelling and models parameters estimation. The main approaches are white-box modelling and black-box identification [6]. The white-box models use scientific relations for describing the process. Parameters of the processes are known constants or they come from specifications of the modelled components. For example, server power consumption modelling requires power parameters of the corresponding CPU and server fan. From the authors' experience, it is not so easy to get the required parameters for the CPU and the server fan. As an idea of the toolbox is in modelling data centres of any configuration, the modelling approach used in it is a so-called grey-box, which utilises scientific relations and requires additional techniques for parameter estimation.

As this work deals with the server power modelling, we are interested in modelling approach and parameters estimation for such components as CPUs and server fans. In [7], authors provide a detailed survey of works, which concentrate on modelling the energy consumption of all components of data centres including servers and their components.

The work [8] demonstrates that the CPU power consumption depends linearly on its utilisation. This work describes the experiments on specific CPUs to get their power parameters. However, it is not always possible to conduct experiments with each type of CPU presented in the modelled data centre. In [9], the authors also deal with the linear relationship between the CPUs power consumption and its utilisation, but there are no clear recommendations on parameters selection.

For the server fans, there also exists works, which consider modelling their power consumption. In [9], the authors present the cubic polynomial model of fan power consumption, but there is no explanation on how its parameters were obtained. The work [10] describes the experiment on a specific server fan to get parameters for its power model, which is a cubic polynomial. However, it is not always possible to conduct experiments with each type of server fan utilised in the modelled data centre.

Thus, this work is inspired by the necessity to have reliable procedures for parameters estimation in the server power model. The work presents estimation procedures based on data about CPUs utilisation, server fans speed and power consumption of servers. Those data are collected from the real data centre. One of the suggested procedures is based on a regression model, which is a common machine learning algorithm. So the paper relates to artificial intelligence systems in way of using machine learning algorithms.

3 Parameters Estimation

This work deals with parameters of the Server block, which estimates the power consumption of the server as a sum of the power of its main components: CPUs and server fans. The total power consumption of the server can be calculated by (1), which is combination of equations for CPU and server fan power consumption used in the toolbox [4].

$$P_{SRV} = \sum_{i=1}^{n}\left(P_{CPU,idle} + \left(P_{CPU,max} - P_{CPU,idle}\right)Util_i\right) + \sum_{i=1}^{m}\frac{P_{SF,max}}{RPM_{max}^3}RPM_i^3\ . \quad (1)$$

Here, n is number of CPUs on the server; $Util_i$ is utilisation of ith CPU; m is number of server fans on the server; RPM_i is rotation speed of ith server fan.

From equation (1) inner parameters of the server block: the CPU power consumption in idle mode ($P_{CPU,idle}$); the CPU peak power consumption ($P_{CPU,max}$); the server fan peak power consumption ($P_{SF,max}$); the server fan maximum rotation speed (RPM_{max}).

This work suggests two ways for parameter estimation: regression-based estimation and simulation-based optimisation.

The *regression-based estimation* performs the following steps:

1. constructing the regression model which reflects the relationship between the values of the independent variables (predictors) and the explanatory variable (responses);
2. planning the experiments on the real facility to measure values of the predictors and responses;
3. pre-processing measured data;
4. determining the coefficients of the regression model;
5. statistical analysis of the results.

The *simulation-based optimisation* performs the following steps:

1. collecting data from the real facility;
2. constructing a model of a component whose parameters require estimation;
3. setting inputs of the model to the data from the real facility;
4. running the optimisation method, which minimises the mean deviation among modelling results and real values (cost function) and takes parameter values corresponding to the minimum as required ones.

The following subsections discuss the implementation of both ways for estimation parameters of building blocks in the data centre modelling toolbox. This section deals with estimating the parameters from the equation (1). The model takes the assumption that all CPUs on the server are identical to each other, the same works for all the server fans. It means that all the CPUs and all the server fans have the same parameter values.

3.1. Regression-based Estimation

Based on (1) the server power can be presented as linear regression model:

$$P_{SRV} = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 \,. \tag{2}$$

In (2) regression coefficients represent the server block parameters:

$$a_0 = n \cdot P_{CPU,idle};\ a_1 = P_{CPU,max} - P_{CPU.idle};\ a_2 = \frac{P_{SF,max}}{RPM_{max}^3}. \tag{3}$$

In (2) predictors represent the server block inputs:

$$X_1 = \sum_{i=1}^{n} Util_i;\quad X_2 = \sum_{i=1}^{m} RPM_i^3 \,. \tag{4}$$

To find regression coefficients, the measured values of predictors (X_1 and X_2) and response (P_{SRV}) are required. For that aim, the experiment in the real data centre is conducted. An experimenter can control the CPUs utilisation directly. Whereas server fans speeds depend on the corresponding CPUs temperature, thus they are uncontrollable by the experimenter.

The experiment plan:

1. set the experiment duration (T);
2. split the experiment time into two or more periods;
3. set the utilisation of all CPUs so that it is 100% and 0% during the first and second periods (if there are other periods the utilisation can have arbitrary value but it should be constant during the period);
4. measure the rotation speed of all server fans and power consumption of all servers during the periods;
5. save measurements as time-series.

The result of the experiment is measured data, such as the servers power consumption and the rotation speed of server fans saved as separate time-series for each server and server fan.

Before starting the calculation of parameters, the measured data requires pre-processing.

Measurements pre-processing:

1. remove explicit outliers from the data set;
2. smooth out the data in the data set by filtering;
3. carry out centring the predictors and the response to get rid of intercept in the regression model;
4. perform standardisation of the predictors and the response in the regression model.

At last, when all data are prepared, it is possible to calculate the parameters. For each server, parameters are found using matrix form of linear regression model (5).

$$P_N = X_N \cdot \alpha. \tag{5}$$

In (5) P_N is a centred and standardised vector of measured values of server power consumption; X_N is a centred and standardised matrix of measured values of CPU utilisation (the first column), and cube rotation speed of the server fans (the second column); α is a vector of parameters. Traditionally, the method of least squares is used to obtain the values of α, which is reduced to solving a system called normal equations (6) with a positive definite symmetric matrix.

$$X'_N \cdot P_N = (X'_N \cdot X_N) \cdot \alpha. \tag{6}$$

The vector α can be determined as in (7), as the system matrix is small, it is only 2×2.

$$\alpha = (X'_N \cdot X_N)^{-1} \cdot X'_N \cdot P_N. \tag{7}$$

Use the vector α to calculate coefficients represented by (3).

Statistical analysis of the results is in testing the hypothesis that coefficients calculated for different servers are from the same distribution with the same mean value.

Section 5 demonstrates the implementation and results of the equation-based estimation procedure for data measured in the real data centre.

3.2. Simulation-based Optimisation

The simulation-based optimisation is another procedure to estimate server parameters. This procedure minimises the cost function (8), which estimates the mean deviation among modelling results and real values [12], [13].

$$J\left(P_{CPU,idle}, P_{CPU,max}, P_{SF,max}, RPM_{max}\right) = \frac{1}{2m}\sum_{t=1}^{m}\left(P_{SRV,real}(t) - P_{SRV,model}(t)\right)^2. \tag{8}$$

In (8), $P_{SRV,real}(t)$ is the real server power at time t; $P_{SRV,model}(t)$ is the server power calculated by the model at the same time t; and m is the number of timestamps.

The parameters estimation process goes through the following steps. The first step is in the server model preparation:

(1) the initial value, lower and upper bounds of all parameters are determined;

(2) the time-series with data about the CPU usage, the rotational speed of corresponding local fans, are set as input values of the auxiliary model;

(3) the server model provides modelled values of the server power consumption, which is used for calculation the cost function (8).

The second step consists in running the optimisation process (Matlab function: fmincon) which runs the server model with the current parameters to calculate the cost function value, then generate new parameter values and reruns the server model until the global minimum for cost function is found. The parameter values corresponding to the found minimum is considered as the desired parameter values.

The next section demonstrates the implementation and results of the simulation-based optimisation procedure for data measured in the real data centre.

4 Results and Discussion

To evaluate the proposed earlier procedures for the server parameters estimation, the data were collected in the Edge module of the SICS ICE facility located in Luleå, Sweden [5]. Table 1 demonstrates the profile of the CPUs utilisation during the data collection experiment.

Table 1. The CPUs utilisation profile during the data collection experiment.

Time:	0-3 h	3-6 h	6-9 h	9-10 h	10-11 h	11-12 h
Utilisation:	100 %	50 %	75 %	100 %	0 %	80 %

Fig. 1 demonstrates the results of the server parameters estimation with the equation-based and the simulation-based procedures. Both procedures shows the similar results. The mean absolute error (MAE): for the equation-based procedure is 5.5 W; for the simulation-based optimisation is 5.1 W. The mean absolute percent error (MAPE): for the equation-based procedure is 5.1 %; for the simulation-based optimisation is 4.8 %.

Both procedures demonstrate quite realistic results, and they are added as scripts to the toolbox so that they can be used for parameters estimation. Comparing the MAE and MAPE for both procedures the best one can be selected, and parameters obtained with it can be applied as the server model parameters.

It is worth mentioning, that the equation-based estimation procedure takes much less time than the simulation-based optimisation. However, for the server fan, it can estimate only the composite parameter $\left(\frac{P_{SF,max}}{RPM_{max}^3}\right)$, when the simulation-based optimisation is able to estimate all parameters separately. In addition, the equation-based estimation can be used only if corresponding equation is quite easy such as linear equation. Thus, the toolbox provides with both procedures, so the decision, which one should be used, is made in each individual case relying on timing and accuracy requirements.

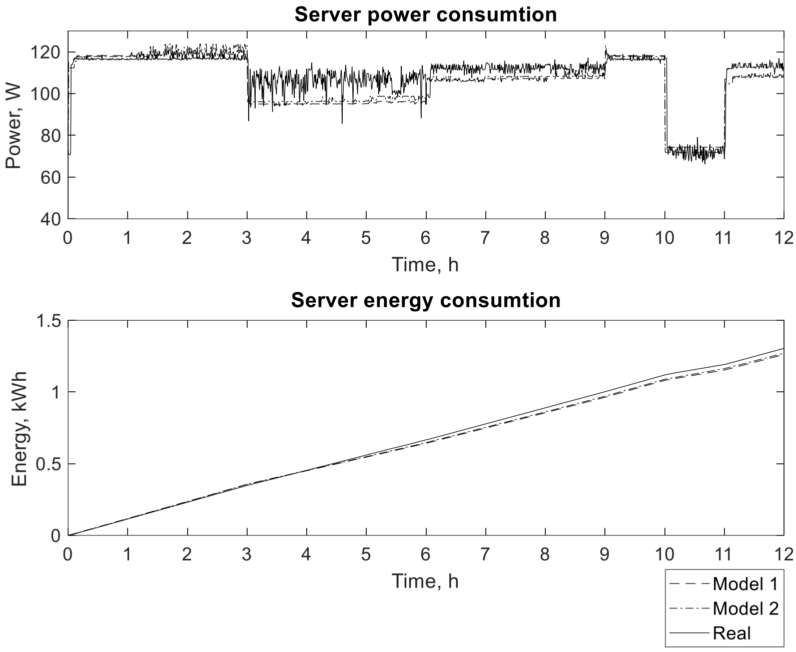

Fig. 1. Comparing the modelling results and real data: server power consumption and server energy consumption.

5. Conclusion

In this paper, we proposed and implemented two procedures for parameters estimation at server power modelling. The toolbox [4] has been extended with two scripts implementing the estimation of the parameters. The procedures have been employed to the estimation of parameters of server power model. Section 5 demonstrates the results of parameters estimation using the proposed procedures. The models constructed with the estimated parameters demonstrate that the calculated power consumption of the server is close to the real data.

The future work is going to be developed in two directions. The first one is in implementation of additional techniques for parameters estimation such as on-line parameters estimation, and utilising neural networks for estimation parameters. The second one is in the estimation of parameters for all building blocks in the data centre modelling toolbox [Ber]. The special interest here is parameters determining the thermal behaviour of data centres, namely temperature evolution of all CPUs and air inside the data centre. The idea here is to extend the toolbox with all possible techniques of parameters estimation and each time at model constructing use those, which demonstrate most accurate results at modelling.

Acknowledgments. This project has been funded by partners of the ERA-Net SES 2018 joint call RegSys (www.eranet-smartenergysystems.eu) – a network of 30 national and regional RTD funding agencies of 23 European countries. As such, this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 775970.

References

[1] P. Bertoldi, "A Market Transformation Programme for Improving Energy Efficiency in Data Centres," in Proceedings of the MELS: Taming the Beast, ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 2014.

[2] C. Koronen, M. Åhman and L. J. Nilsson, "Data centres in future European energy systems—energy efficiency, integration and policy," Energy Efficiency, vol. 13, pp. 129-144, 2020.

[3] Y. Berezovskaya, C.-W. Yang and V. Vyatkin, "Towards Multi-Agent Control in Energy-Efficient Data Centres," in IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, 2020.

[4] Y. Berezovskaya, C.-W. Yang, A. Mousavi, V. Vyatkin and T. B. Minde, "Modular Model of a Data Centre as a Tool for Improving Its Energy Efficiency," IEEE Access, vol. 8, pp. 46559-46573, 2020.

[5] URL of SICS ICE Data Centre located in Luleå, Sweden: https://www.ri.se/en/ice-datacenter/about-ice-datacenter (accessed at 11.02.2021).

[6] B. Sohlberg and E. W. Jacobsen, "Crey Box Modelling - Branches and Experiences," IFAC Proceedings Volumes, vol. 41, no. 2, pp. 11415-11420, 2008.

[7] M. Dayarathna, Y. Wen and R. Fan, "Data Center Energy Consumption Modeling: A Survey," IEEE Comunications Surveys & Tutorials, vol. 18, no. 1, pp. 732-794, 2016.

[8] X. Fan, W.-D. Weber and L. A. Barroso, "Power Provisioning for a Warehouse-Sized Computer," in 34th ACM International Symposium on Computer Architecture, San Diego, 2007.

[9] R. Basmadjian, N. Ali, F. Niedermeier, H. de Meer and G. Giuliani, "A Methodology to Predict the Power Consumption of," 2011.

[10] Z. Wang, C. Bash, N. Tolia, M. Marvah, X. Zhu and P. Ranganathan, "Optimal Fan Speed Control for Thermal Management of Servers," in Proc.of the ASME/Pacific Rim Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Systems, MEMS, and NEMS (InterPACK), 2009.

[11] D. C. Montgomery, Design and Analysis of Experiments, Wiley, 2013.

[12] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge, U.K.: Cambridge Univ. Press, 2004.

[13] J. J. Schneider and S. Kirpatrick, Stochastic Optimisation, Berlin: Springer-Verlag Heidelberg, 2006.