Berezovskaya, Yulia; Yang, Chen Wei; Vyatkin, Valeriy

Towards Extension of Data Centre Modelling Toolbox with Parameters Estimation

Published in:
Technological Innovation for Applied AI Systems

DOI:
10.1007/978-3-030-78288-7_18

Published: 01/07/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Berezovskaya, Y., Yang, C. W., & Vyatkin, V. (2021). Towards Extension of Data Centre Modelling Toolbox with Parameters Estimation. In L. M. Camarinha-Matos, P. Ferreira, & G. Brito (Eds.), *Technological Innovation for Applied AI Systems : Proceedings of 12th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2021* (pp. 189-196). (IFIP Advances in Information and Communication Technology; Vol. 626). Springer. https://doi.org/10.1007/978-3-030-78288-7_18

Towards Extension of Data Centre Modelling Toolbox with Parameters Estimation

Abstract. Modern data centres consume a significant amount of electricity and require energy-saving techniques. 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 such experiments may violate safety and security protocols. The model accurately estimating the energy consumption in data centres and predict its thermal behaviour is required. This work suggests a toolbox for data centre modelling which is implemented as 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, parameter estimation

1 Introduction

Nowadays, data centres consume a significant amount of electricity, which is comparable to industrial consumption [1]. At the same time, the data centres constitute a part of the energy grid. Moreover, modern data centres use more and more renewable energy sources [2]. Therefore, the aim of data centre operators is in sustaining high availability and reliability of the data centres at the intermittent energy production intrinsic to renewable energy, and constraints from the energy grid.

The promising way here is flexible control methods, which adapt the data centre energy use based on resource requirements at run-time. Such methods are, for example, smart distribution of data centre workload considering variations of electricity price; resource management techniques such as virtualisation, scheduling, consolidation; partial or complete deactivating the idle servers and the dynamic voltage and frequency scaling (DVFS). In addition, such methods as adjusting the rotation speed of server fans to the actual temperature of the corresponding CPUs, adjusting the utilisation of cooling units inside the data centre to the actual IT-load, and adaptive control of the joint action of free-cooling and ordinary air-cooling improve the data centre energy efficiency [3].

Those control methods require tuning and examining. The examining procedure uses closed-loop co-simulation, in which a model of a data centre substitutes the real facility. 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. Moreover, those values work only for modelling the specific data centre but any other one requires new parameter values, which ensure accurate modelling. Thus, the toolbox needs additional procedures, which provide appropriate parameter values. One possible direction here is in using the datasheets for data centre components; another one is in extracting the parameter values from real data. This work deals with the second way. The paper contributes with:

- new ideas for parameter selection (description),
- their implementation in the Simulink toolbox,
- comparing the results,
- summary on which way is more preferable (ads/cons).

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. Section 3 describes the parameters, which require the estimation. Section 4 describes the general procedures for parameters estimation. Section 5 considers results of parameters estimation and compares two proposed procedures. Section 6 gives the conclusion.

2 Related Works

3 Parameters description

3.1 Server block parameters

In the toolbox, the Server block is responsible for modelling of the main IT component. The block 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].

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

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}).

3.2 Cooling system parameters

. (2)

4 Parameters estimation

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

The *equation-based estimation* performs the following steps: analysis of the equations in which parameters appear and detecting the components of the equations, which can be measured on a real facility; planning the experiments on the real facility, which provide the measurements of the components; preprocessing measured data; calculating the parameters from the equation using the measurements; statistical analysis of the results.

The *simulation-based optimisation* performs the following steps: collecting data from the real facility; constructing a model of a component whose parameters require estimation; setting inputs of the model to the data from the real facility; 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.

4.1 Server block parameters

This subsection 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.

The equation-based estimation. Analysis of the equation (1) found out that the Server block needs following inputs: utilisation of all CPUs and rotation speed of all server fans. When planning the experiments on the real data centre, it is necessary to keep in mind that 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.

Detailed consideration of equation (1) shows that at the same CPUs utilisation the change of the server power consumption depends only on change of the server fans speeds. The server fan parameter can be calculated with (3):

$$\frac{P_{SF,max}}{RPM_{max}^3} = \frac{P_{SRV,2} - P_{SRV,1}}{\sum_{i=1}^{m}\left(RPM_{2,i}^3 - RPM_{1,i}^3\right)} . \quad (3)$$

When the server fan parameter is known, it is possible to calculate both parameters of the CPU using (4) with only one difference: the CPU power consumption in an idle mode ($P_{CPU,idle}$) based on the data corresponding 0% utilisation, and the CPU peak power consumption ($P_{CPU,max}$) based on data corresponding 100% utilisation.

$$P_{CPU,idle/max} = \frac{1}{n} \cdot \left(P_{SRV} - \frac{P_{SF,max}}{RPM_{max}^3} \cdot \sum_{i=1}^{m} RPM_i^3\right) . \quad (4)$$

When there is an idea of how to calculate the parameters, collecting the data can be planned. For that aim, the experiment in the real data centre requires.

The experiment description:

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 preprocessing.

Measurements preprocessing:

1. split all time-series into subseries so that each one corresponds to constant CPUs utilisation;
2. remove explicit outliers from the subseries;
3. smooth out the data in the subseries by filtering.

At last, when all data are prepared, it is possible to calculate the parameters. For each server, it is necessary to implement the following procedure.

Parameters calculation:

1. take a random sample of time points for each of the subseries;
2. use server power values and fans speed corresponding to selected time points to calculate sets of values of server fan parameter by (3);
3. make sure that all sets of parameter values have the same mean value (test statistical hypothesis of means equality [5]), otherwise additional pre-processing of measurements is required;
4. take the mean value as the server fan parameter;
5. use the server fan parameter value and random sample from subseries corresponding to 0% CPU utilisation to calculate CPU power consumption at idle mode by (4);

6. use the server fan parameter value and random sample from subseries corresponding to 100% CPU utilisation to calculate peak CPU power consumption by (4).

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

The simulation-based optimisation is another procedure to estimate server parameters. This procedure minimises the cost function (5), which estimates the mean deviation among modelling results and real values [6], [7].

$$J\left(P_{CPU,idle}, P_{CPU,max}, P_{SF,max}, RPM_{max}\right) = \frac{1}{2m}\sum_{t=1}^{m}\left(P_{SRV,real}(t) - P_{SRV,model}(t)\right)^2 . \quad (5)$$

In (5), $P_{SRV,real}(t)$ 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 time-series with data about the real server power consumption is set as a sample for comparison with modelling results.

The second step consists in running the optimisation process 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.

4.2 Cooling system parameters

5 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 [ICE DC]. 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 %.

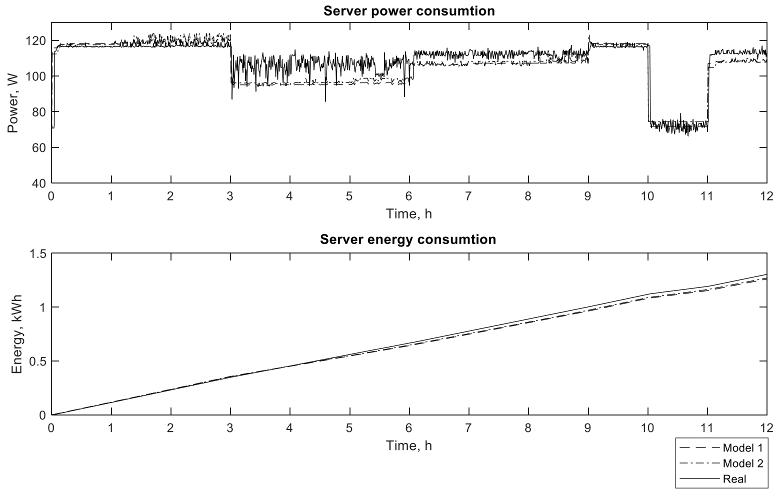

Fig. 1. Comparing the modelling results and real data: server power consumption and server energy consumption.

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. 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.

6 Conclusion

Acknowledgments. The heading should be treated as a 3rd level heading and should not be assigned a number.

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.