

Towards Simple Models for Energy-Performance Trade-Offs in Data Centres

Boudewijn R. Haverkort and Björn Postema*

University of Twente

CTIT Centre for Dependable Systems and Networks

`b.r.h.m.haverkort@utwente.nl`, `b.f.postema@utwente.nl`

<http://www.utwente.nl/ewi/dacs/>

Abstract. In this paper we advocate the use of simple stochastic models to analyse the energy-performance trade-off in data centres. Recently such trade-offs have received increased attention, however, the tools used to make such trade-offs are largely based on simulation and real-life experiments. Although simulations studies are very helpful, we think that simple analytical models, or models based on stochastic Petri nets (or similar description techniques) can be very fruitful in guiding design processes in the early phases. Similarly, we do think that experimental work is very important, however, its results come “after the fact” in the sense that the system has been built already once the experiments are being performed. Our claim is that the use of simple models early in the design phase provides a very good return on investment. This short paper presents some preliminary models that can be used for early-in-design trade-off analyses.

Key words: Data centres, energy, performance, analytical models, matrix-geometric methods, stochastic Petri nets.

1 Introduction

Recent studies have revealed that ICT equipment consumes 2-3% of all electrical energy, and the trend is that this number is growing [13]. Although other sectors are much more energy hungry, i.e., traditional industries amount for about 40%, transportation

* The work in this paper has been supported by the Dutch national STW project *Cooperative Networked Systems*, as part of the program “Robust Design of Cyber-Physical Systems”.

for about 20%, and residential use for about 12%, still it is important to make ICT less energy hungry. A recent EU forecast states that in 2020 93 TWh is used for ICT, which is the equivalent of 106 million 100 Watt light bulbs burning continuously for a year. Data centres are particularly energy hungry: per 60×60 cm² floor usage in a data centre, the annual CO₂ emission is as much as 1200 kg. Of course, this is no news any more; many project have in the meantime been started to investigate “green ICT”.

Within a data centre, around 50% of the energy is being used for ICT, the remaining part being used for lighting, UPS (uninterruptible power supply), cooling, etc., see Figure 1. Furthermore, it is well-known that energy use for ICT in a data centre is multiplied through the so-called *cascade effect* [7], which states that a 1 Watt saving in CPU power, leads to 0.18 W reduction for DC-DC conversion, 0.31 W reduction in AC-DC power conversion, 0.04 W in power distribution, 1.4 W in UPS, 1.07 W in cooling, and finally 0.1 W in power reduction for switchgear and transformer capacity. In total, this yields a factor 2.84. This is both good and bad news: more energy use for ICT implies more energy use elsewhere (bad news), however, the good news is that savings at the ICT are also translated into savings elsewhere.

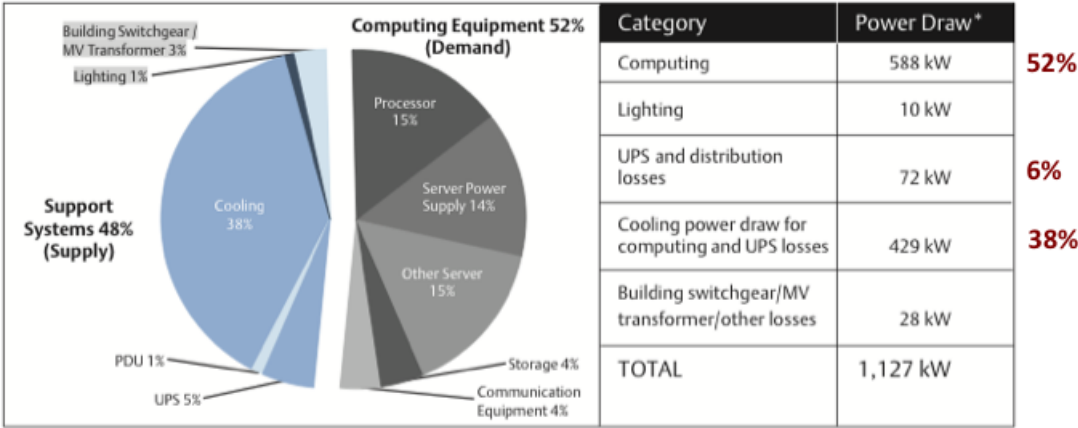


Fig. 1. Energy use within a data centre (picture from [4])

In this paper we advocate the use of simple stochastic models to analyse the energy-performance trade-off in data centres. Recently such trade-offs have received increased attention, however, the techniques used to make such trade-offs are largely based on simulation and experiments. Although simulations studies are very helpful, we think that simple analytical models, or models based on stochastic Petri nets (or similar description techniques) can be very fruitful in guiding design processes in the early

phases. Similarly, we do think that experimental work is very important, however, its results come “after the fact” in the sense that the system has been built already once the experiments are being performed. Our claim is that the use of simple models early in the design phase provides a very good return on investment.

Below, we will first sketch the bigger picture in energy savings for data centres, followed by a simple model-based approach that will allow us to investigate design trade-offs regarding energy and performance, thereby using simple Markovian models based on stochastic Petri nets.

2 Possible energy reduction steps

In this section we briefly discuss a number of ways to decrease energy use in data centres [7], that range from very practical (“moving boxes”) to more advanced (requiring elaborate sensory equipment) and more software-oriented.

1. Data centre ICT equipment:

- The use of *low voltage processors* will directly decrease the power usage with some 30%, thereby not necessarily impacting performance, although detailed studies have to be made to ascertain that.
- The use of *high-end power supplies*, with efficiency 90% instead of the typical 70% that is standard for low-end power supplies in consumer computing equipment. Furthermore, the power supplies should be chosen such that they, under normal load circumstances, operate at their optimum.
- The use of *blade servers*, that use multiple processing boards with shared IO, fans, and power supply.
- The use of emerging techniques for *green networking*; in the current data centre literature there appears to be a focus on computing only.

2. Data centre (power) management software:

- The use of *advanced power management software*, that make that servers or server groups can be switched off completely while still meeting the performance requirements; this will be elaborated upon in the next section.
- Advanced *server virtualisation* software can be used to increase server utilisation and to reduce the number of active servers.

3. Data centre power supply:

- Higher voltage AC *power distribution* within a data centre can decrease overall power usage, as higher voltage transport is more efficient than low voltage

transport; the EU is doing better here (with their standard 240V than the US with standard 110V).

- The use of *more efficient UPS systems*, that do avoid the double conversion, from the external AC source, to the DC storage and buffering, and the AC end-use.

4. Data centre cooling:

- The use of better *spatial arrangement of servers* and cooling (use of hot/cold aisles) and higher room temperatures (28 vs. 20 degrees Celsius) and *variable capacity precision cooling* (instead of simple overall room cooling) to cool just there where it is needed.
- *Per server/system monitoring and control* of temperature, humidity, etc., to further increase cooling efficiency; this requires the installation of an advanced (wireless) sensor system.

Of the above energy reduction steps, we will focus below on an approach that allows for the analysis and effect of dynamic power management software, server virtualisation and per-server monitoring, as these lend themselves very well for a model-based approach.

3 A model-based approach

As stated in the introduction, most work on data centre energy efficiency focuses on simulation and experimentation. The focus in this short paper is on analytical and numerical techniques. Recent work along these lines is still rather limited, but interesting (and partly similar) work on using Markov chains can be found in [8,10,14]. Another interesting approach based on stochastic Petri nets is [3], in which the effect on the energy usage of on-demand creation and deletion of virtual machines is studied.

The basic idea of our models is illustrated in Figure 2: a data centre serves a job stream from the outside world, buffers the incoming jobs and subsequently schedules and executes them, on the basis of the job requirements and the system-internal state information that is available. What that state information exactly is, largely depends on the data centre: it might involve only information on job queue lengths or server utilization, but can also include information on temperature and humidity in parts of the data centre, or information in networking bottlenecks, to give just a few examples.

Of course, the model in Figure 2 cannot be used for any computation; it has to be made more concrete, e.g., in the form of the simple (infinite-state) stochastic Petri net [16,17] given in Figure 3. In this model, jobs arrive according to a Poisson process

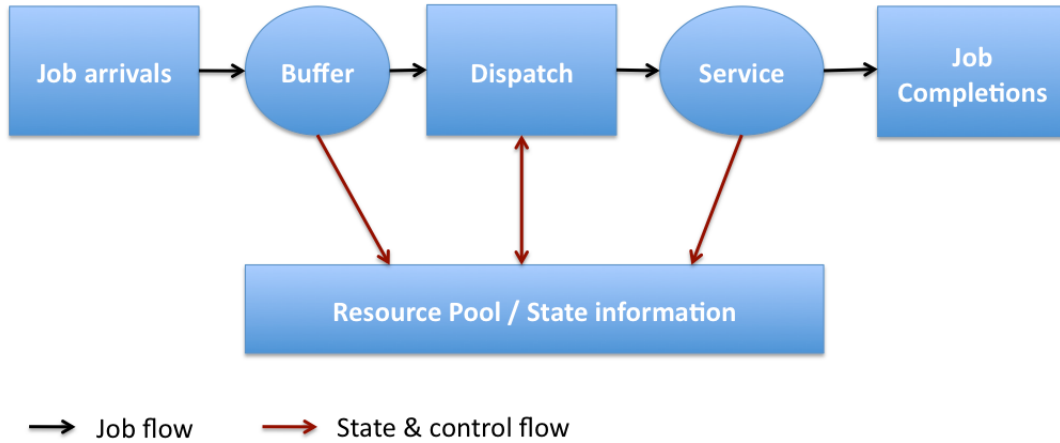


Fig. 2. The basic model for energy/performance trade-offs in data centres

(transition **Arrivals** with rate λ) and are buffered in place **Buffer**. If the server is switched off upon arrival (token in place **Off**), it has to be switched on first (via transition **Setup**), leading to an extra delay (transition **Delay** with rate α) and extra energy use (as long as there is a token in place **Booting**) before actual processing (token in place **Processing** can start (transition **Service**). Once the processing finishes, with rate μ , the server is moved to place **Idle** where it will not stay when there are other jobs buffered (since transition **Start** will fire immediately in that case). If there are no other jobs waiting to be served, the server will stay idle for some amount of time, before it is switched to a lower-energy state (via transition **Release**). The time-out value (exponential rate β of transition **Release**) in relation to the setup delay (exponential rate α) as well as the energy usage parameters (non-zero reward rates “E” for places **Booting**, **Processing**, and **Idle**), allow for making a trade-off between energy usage and performance requirements. Energy usage is modelled using these *rate rewards*, with the following semantics: as long as there is a token in place, say, **Processing**, then energy is used with rate $E_{\text{Proc}} = 200$. Of course, this could be extended to also include *impulse rewards*, that is, impulses of energy being used instantaneously when certain transitions fire.

Note that the above is the simplest possible model, but that it can be easily made more advanced by adding multiple servers, by using more advanced workload models (such as multi-class arrival patterns and multi-class service times), using phase-type distributions, the use of deterministic timing, etc.

Once the model has been specified, it can be solved for either its steady-state behaviour, or its time-dependent (transient) behaviour. In the former case, performance

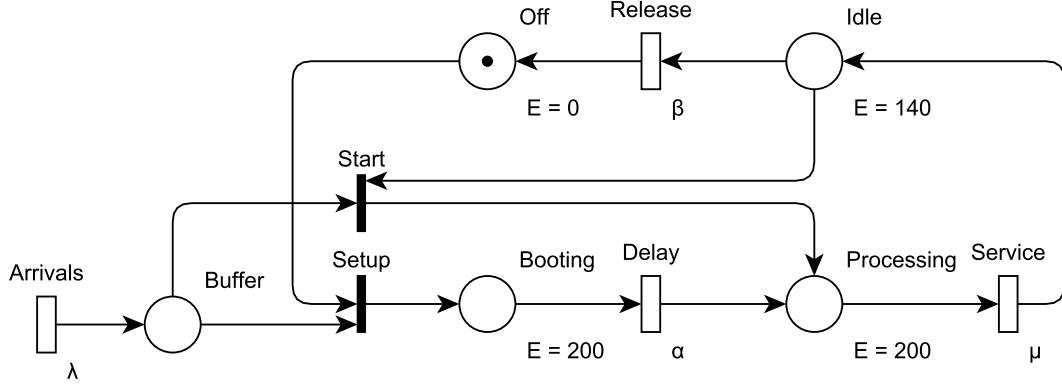


Fig. 3. An infinite-state stochastic Petri net for energy/performance trade-offs in data centres

measures such as throughput, mean delay, server utilisation, etc., can be computed, as well as the overall energy-usage rate, that is, the power consumption. In case of a transient analysis, say, for some finite period $[0, t)$, the expected cumulative number of jobs processed and the expected amount of energy used for that can be computed. In its full generality, we might need to use discrete-event simulation to evaluate the model, however, in restricted cases, a numerical analysis can be performed via the automatically generated underlying Markov chain (see [11]). In particular, for the model at hand, we used the Möbius toolset to evaluate the models fully numerically [5].

We now provide an example of the type of trade-offs that can be made. We have used the following parameters (taken from [2,4,9]) for the model: $\lambda = 0.007$ (this is a low rate, but we are only modelling one server here), $\alpha = 0.01$ (giving a mean boot time of 100 s), μ is taken from the set $\{0.01, 0.025, 0.1, 1.0\}$, and the power levels for processing and booting are 200 W, and for idling 140 W. Finally, we let the time-out rate β increase from 0.01 to 1.0 (in steps of 0.01). Figure 4 presents a typical (mean) power-response time trade-off, with on the x -axis the mean response time, and on the y -axis the mean power used, for the four different service rates μ (from bottom-left to top-right: 1.0 (purple), 0.1 (blue), 0.025 (green) and 0.01 (red)). On every curve each cross signifies a different value for the time-out rate β : on the left-most end of each curve the value $\beta = 0.01$ (long average delay before shut-down) and towards the right larger values for β (quicker shut-down); the right-most point is the case $\beta = 0.01$. Clearly, if the time-out delay is, on average, larger (towards the left end of the curves), the server is, upon job arrivals, most often still switched-on but idle, hence, job processing can quickly start (no booting delay), resulting in a lower mean delay, at the cost of

higher mean power usage. Conversely, if an idle server is more quickly switched off (towards the right on the curves), a larger mean delay is perceived, but a lower mean power-usage is the gain reached.

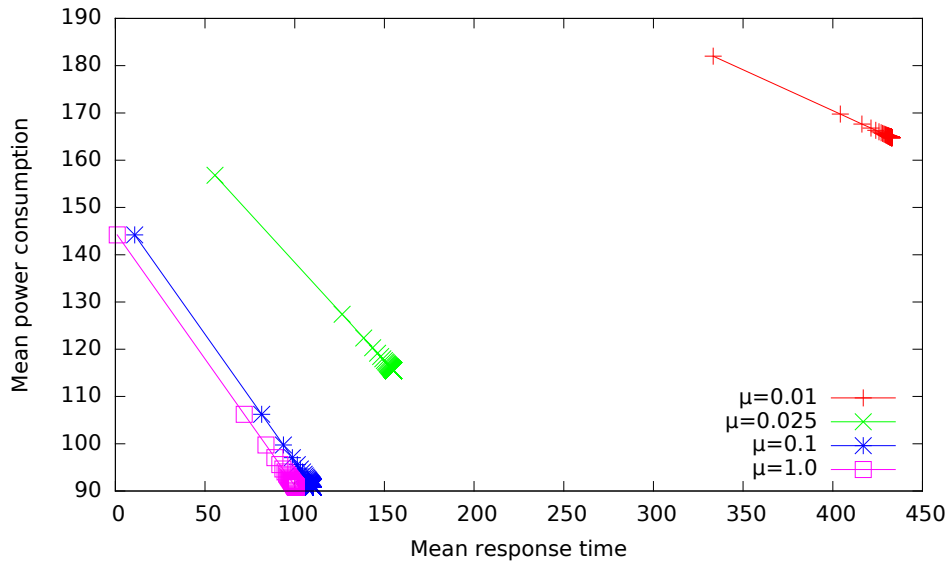


Fig. 4. Mean response time against mean power consumption, using the simple data centre model, for four different service rates

We finish this paper by noting that the model of Figure 3 has as special feature that its underlying Markov chain exhibits a repetitive structure, as depicted in Figure 5, such that efficient matrix-geometric methods can be employed for its solution. In this figure, states of the form (o, n) ($n = 0, 1, 2, \dots$) signify states in which the server is idle and there are n jobs in the system, states of the form (p, n) ($n = 1, 2, \dots$) signify states in which the server is processing a job and there are in total n jobs in the system, and state $(i, 0)$ signifies that the server is idle, and that there are no jobs in the system (any more). In three coloured blocks, we indicate the rate rewards to be associated with each of the states. By solving the underlying CTMC, using matrix-geometric methods, the individual steady-state probabilities can be computed, from which mean performance measures such as throughput and delay can be derived. It should be noted that the type of model sketched here, is not new as such. In the mid-1990's, work has been done on connection management in network systems, in which efficient trade-offs had to be found between network connectivity time (hiring of bandwidth) and connection set-up delays; cf. [12,16,17]. The additional parameter of consideration here

is energy; the models developed then, can be extended easily towards the needs we have now.

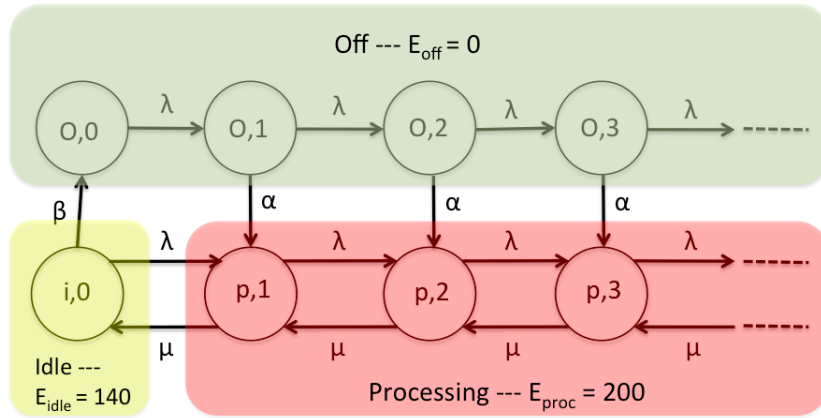


Fig. 5. The underlying infinite-state CTMC for the basic model for energy/performance trade-offs in data centres

4 Conclusion

In this short paper we have proposed the use of analytical and numerical models for the evaluation of, especially, dynamic power management strategies for data centres. We have shown the basic model structure, and given a concrete example of the simplest possible model. Starting from the generic model, many more advanced models can easily be developed and evaluated, using one of the many available stochastic Petri net tools. We also provided a concrete example of the type of trade-offs that can be made.

References

1. T. Bostoen, J. Napper, S. Mullender, Y. Berbers. Minimizing energy dissipation in content distribution networks using dynamic power management. *Proc. Third IEEE Int'l. Conference on Cloud and Green Computing*, pp.203–210. 2013.
2. L.A. Barroso and U. Hölzle. The case for energy-proportional computing. *IEEE Computer* **40**(12): 33–37, 2007.
3. D. Bruneo, A. Lhoas, F. Longo, A. Puliafito. Analytical evaluation of resource allocation police in green IaaS clouds. *Proc. Third IEEE Int'l. Conference on Cloud and Green Computing*, pp.84–91. 2013.
4. G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, W. Vogels. Dynamo: Amazons highly available

- key-value store. *Proceedings of 21st ACM SIGOPS Symposium on Operating Systems Principles*, pp.205–220, 2007.
5. G. Clark, T. Courtney, D. Daly, D. Deavours, S. Derisavi, J.M. Doyle, W.H. Sanders, and P. Webster. The Möbius Modeling Tool. *Proceedings of the 9th International Workshop on Petri Nets and Performance Models*, IEEE CS Press, 2001, pp. 241-250. See also <https://www.mobius.illinois.edu/>.
 6. N. El-Sayed, I. Stefanovici, G. Amvrsiadis, A.A. Hwang. Temperature management in data centres: Why some (might) like it hot. *Proc. ACM Sigmetrics*, pp. 163–174, 2012.
 7. Emerson Network Power. Energy Logic: Reducing data centre energy consumption by creating savings that cascade across systems. 2009.
 8. A. Gandhi, M. Harchol-Balter. How data size impacts the effectiveness of dynamic power management. *Proc. 49th IEEE Allerton Conference on Communication, Control & Computing*, 2011.
 9. A. Gandhi, M. Harchol-Balter, M. Kozuch. Are sleep states effective in data centers? *Third IEEE International Green Computing Conference*, pp.1-10, 2012.
 10. A. Gandhi, S. Doroudi, M. Harchol-Balter, A. Scheller-Wolf. Exact analysis of the M/M/k/setup class of Markov chains via recursive renewal reward. *Proc. ACM Sigmetrics*, pp.153-166, 2013.
 11. B.R. Haverkort. *Performance of computer-communication systems: A model-based approach*. John Wiley & Sons, 1998.
 12. G. Heijenk, B.R. Haverkort. Design and evaluation of a connection management mechanism for an ATM-based connectionless service. *Distributed Systems Engineering* **3**: 53–67. IEE/IOP Publishing, 1996.
 13. J.G. Koomey. Worldwide electricity used in data centres. *Environmental Research Letters* **3**: 1–8, 2008.
 14. P.J. Kühn, M. Mashaly. Performance of self-adapting power-saving algorithms for ICT systems. *Proc. IFIP/IEEE International Symposium on Integrated Network Management*, pp. 720–723, 2013.
 15. G. Mone. Redesigning the data centre. *Communications of the ACM* **55**(10): 14–16, 2012.
 16. A. Ost, B.R. Haverkort. Analysis of windowing mechanisms with infinite-state stochastic Petri nets. *ACM SIGMETRICS Performance Evaluation Review* **26**(2): 38-46, 1998.
 17. A. Ost. Performance of communication systems: A model-based approach with matrix-geometric methods. Ph.D. thesis, RWTH Aachen University, Department of Computer Science. Published with Springer Verlag, 2001.