TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Informatik

Technische Berichte
Technical Reports

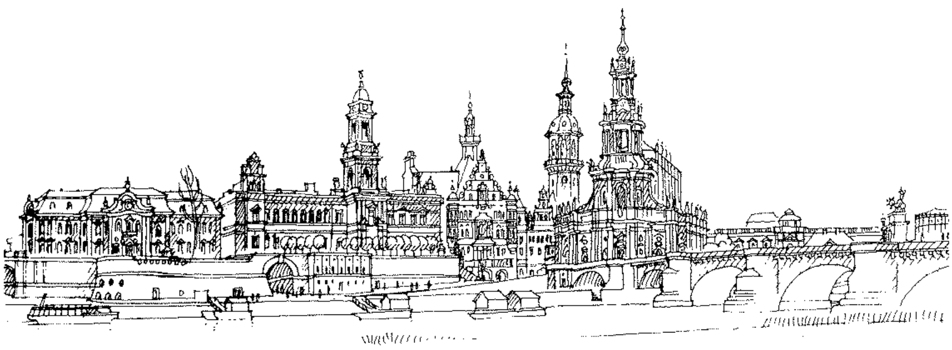
ISSN 1430-211X

TUD-FI-11-05-Sept. 2011

R. Fritzsche, J. Waltsgott, C. Wilke, S. Götz, S. Cech

Institut für Software- und Multimediatechnik

State of the Art: Optimization of Energy Consumption in Storage Systems



Technische Universität Dresden
Fakultät Informatik
D−01062 Dresden
Germany
URL: http://www.inf.tu−dresden.de/

State of the Art: Optimization of Energy Consumption in Storage Systems

Ronny Fritzsche, Johannes Waltsgott, Sebastian Götz,
Claas Wilke, Sebastian Cech

Technische Universität Dresden, Dresden, Germany
{Ronny.Fritzsche, Johannes.Waltsgott, Claas.Wilke,
Sebastian.Cech}@tu-dresden.de, Sebastian.Goetz@acm.org,
WWW home page: http://www.cool-software.org

Abstract. This document illustrates basics and research topics for optimizing the energy consumption in storage systems. Due to the increasing amount of storage systems as main energy consumers in IT infrastructures, we focus on the applicability and relevance for storage subsystems and hierarchical storage systems. Furthermore, we discuss possibilities and economical capabilities of different approaches and illustrate the influence they have on future generations of storage technologies and storage management systems.

1 Introduction and Motivation

In the last decade the energy consumption of IT infrastructures heavily increased. The main reason for this growth consists in the relocation of services into the world wide web and in it's rapid expansion. Thus, today many services are offered online, e. g., music stores, video on demand services and many more. Because of this, more infrastructure is needed to host services and data, which means more energy is consumed. According to [8], energy consumption of data centers has doubled between 2000 and 2005. The reason mainly consists in the dropped costs for low-end and mid-end hardware. Therefore, special storage hardware is not only affordable for big companies, but for small ones and private individuals, too, which use it to provide online services. Further studies of the *American Energy Information Administration (EIA)* and the *International Energy Agency (IEA)* stated that the average worldwide energy consumption per year increased about two percent. This means a doubling every 35 years [19]. Figure 1 illustrates the energy needs since 1980 and gives a forecast to 2030.

Today's efforts for saving energy are mainly limited to the optimization of internal server hardware (e.g., processors, hard disks, ...), which comes to an energy optimization of the servers themselves. For example, in 2005 a single-core processor consumed about 100 W. In contrast, today a core (in idle state) consumes about 3 W of energy [5]. Notably the power consumption of, e.g., disk storage remains largely unchanged. Therefore, server systems are no longer the main energy consumers of an IT infrastructure, but storage systems. The development of new energy optimization techniques for storage systems stagnated,

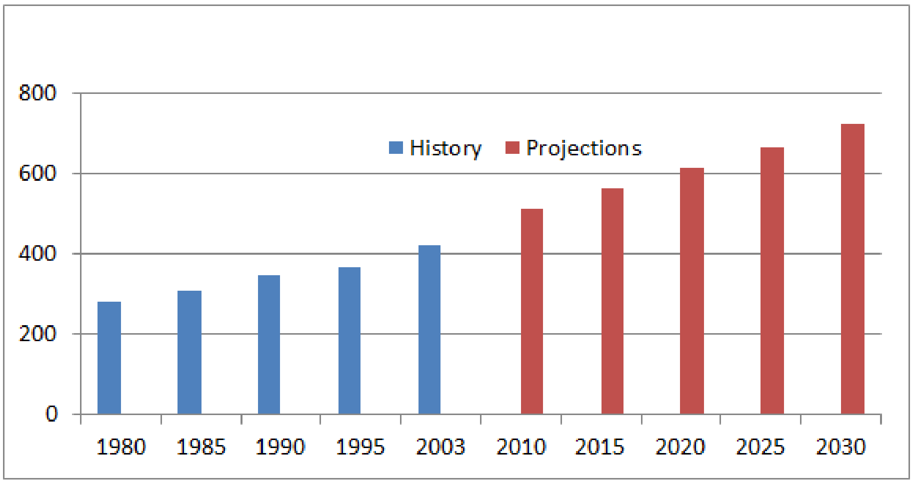


Fig. 1. World Market Energy Consumption 1980-2030 [19].

which made them the new energy consuming leaders. The papers of Guerra, Belluomini, Glider [5] and Raffo [16] illustrate that the storage infrastructure increasingly consumes the biggest amount of energy in a typical data center. For that reason, the challenge of this paper is to show the state of the art and discuss several research papers for optimizing the energy consumption of storage systems and then illustrate trends for future developments. This paper is the base to identify the requirements and to develop a conception for a future storage system.

The remainder of this paper is structured into three sections. Section 2 summarizes different research papers regarding the energy consumption of storage systems. It is divided into four subsections: analyses of data centers (Subsection 2.1), energy optimization techniques of storage hardware (Subsection 2.2), data management aspects (Subsection 2.3) and intelligent algorithms (Subsection 2.4). Thereafter, a common summary is given in Section 3. Afterwards, Section 4 outlines a project specific summary for reuse in the CoolSoftware[1] project. The CoolSoftware project focuses on optimizing software's energy consumption due to energy auto-tuning at runtime. The vision also is an energy management of hardware components. This paper only describes the aspect of optimizing energy consumption of storage systems, but there are further research results regarding hardware energy management and aspects of software's energy consumption illustrated in two additional papers: [20] and [22].

---

[1] http://www.cool-software.org

2 Research Papers

In this section we summarize research papers of the last ten years, which address the energy optimization of storage systems. For a clarity, this chapter is divided into four categories. In Subsection 2.1, we give a summary of some statistics and analyses of big data centers. This will show where energy consumers are located in a typical infrastructure and which potential savings they hold. Thereafter, techniques and approaches for the optimization of individual storage hardware are described in Subsection 2.2. In Subsection 2.3, an abstraction is made by observing the layer above the hardware. We describe optimization techniques which orientate on data. Afterwards, an overview of intelligent algorithms for data reduction is given and their effects on energy savings are discussed in Subsection 2.4.

2.1 Statistical Analyses of Data Centers

Guerra, Belluomini and Glider [5] show a common summary about various research areas and trends. Additionally, an analysis of the energy consumption of data centers is made and their savings potential is outlined. They predict an increasing energy consumption in storage systems. Their thesis is supported by a survey of the *International Data Corporation (IDC)* [7] from 2008. The study describes four trends till 2013:

1. In the next years the unit acquisitions of storage systems will exceed the ones of server systems.
2. Because of faster growing memory needs and slower growing capacities of storage technologies, the acquisitions of additional storage systems will increase.
3. To handle the additional storage requirements data centers will move to 2.5" form factor, which typically needs more energy per GB as comparable devices of the 3.5" class.
4. The performance will improve slower than the capacity.

Furthermore, [5] illustrates an overview about already existing research papers which address the energy optimization of storage systems. For example, a variety of papers exists which concerns about the optimization of single hard drives (cf. Subsection 2.2). However, many of these ideas still have not arrived in industry. Another part of this paper performs an analysis of measurements in data centers. This results in a projection of the energy savings potential of a representative IT infrastructure [5]. Guerra, Belluomini and Glider figured out that hard disks typically are without access for about 20 % of their operation time. Furthermore, the workload of hard disks in a combined storage group is very different, because data is not spread among particular disks by using energy aspects for the distribution itself. At the end of this paper an entire energy savings potential of 40 % to 75 % is calculated if an energy optimization would take place; on hardware side as well as on software side.

Another paper of Poess [15] investigates different energy saving techniques (like clock speed control for processors, power modes in the operating system, energy efficient power supplies and more) which are used in practice. The authors evaluate their influence on the performance in data centers and on its applications. The work is motivated as there are many approaches for energy savings of different system components but there are only few studies about their practical usage in real scenarios. Also almost no data about the scalability of those techniques (for a very huge amount of data) is available. In [15] the distribution of energy consumers in an IT infrastructure is shown, which reveals that the main part (79.1 %)[2] is used by the storage subsystem. This paper also focuses on special hardware, like solid state disks (SSD), and data compression. At the end, a classification regarding the performance and the energy consumption is made, which is used for a rating of the energy optimization. It is shown that SSDs comparatively need less energy at average performance in comparision to traditional hard drives and data compression enables additional energy savings, too. If less data has to be stored, fewer I/O accesses are required leading to a lower energy consumption for a drive. The decreasing performance is balanced by energy savings.

2.2 Energy Optimization Techniques of Storage Hardware

A series of further papers directly focuses on storage hardware systems. Some of them try to optimize the data distribution in a RAID group. A big field of research concentrates on *Dynamic Speed Control for Power Management (DRPM)*[3] [6,18]. It is a technique to dynamically control the rotation speed of hard disks w.r.t. their workload. Traditional approaches often only allow a shutdown of media if a specific time without access elapsed (idle time). According to Tamilarasan, Shankarapani and Qin [18] this only works as far as the load of the hard disk is low and the idle time between different accesses is very high. Typically, this is the case in private and home environments, but it is not applicable for storage systems in an IT infrastructure. DRPM strikes a balance by automatically controlling the rotation speed of hard disk systems. A small number of revolutions helps to optimize the energy usage, because the rotation engine sill consumes most of the energy. Thus, a decreasing rotation speed means a square increasing energy saving, whereas the performance loss grows linear [18].

Further work like Weddle, Oldham and Qian [21] discusses the optimization of RAID systems. Because of their performance and reliability they are widespread among data centers. RAID systems are constructed in a way so that data at most is equally distributed between all hard disks. Because of the energy saving by shutdown or DRPM of single hard disks this approach is only partly suitable. Instead, it is necessary to distribute the data in the storage system w.r.t.

[2] The calculation refers to an analysis of a TPC-C benchmark in the proceedings of the VLDB 1(2): 1229-1240 (2008)

[3] Dynamic Rotations Per Minute

the workload. So rather than to achieve a uniform distribution it is required to get a distribution which enables a shutdown or spin down of hard disks with a small load. In [21] this approach is called *PARAID (Power Aware RAID)*. Regarding to the authors, a reduction of the energy usage from up to 34 % is possible. The advantage of this idea is that the performance and reliability of traditional RAID configurations mostly is conserved. PARAID distributes data above different hard disks. This is done by using free space on each media to build up several overlapping hierarchical RAID configurations. Each configuration includes a different number of hard disk systems. Regarding the workload, hard disks can be shutdown or spun down by changing the RAID configuration. For example, if the workload of the storage system increases the configuration is changed, additional hard disks are activated and requested data is made available from these new disks, too. The advantage on the one hand is an energy saving and on the other hand a bandwidth increase (by using additional hard disks).

Particularly in the area of enterprise storage there is still the problem of scarce propagation of relevant energy optimization strategies [10], because the focus lies on high availability of data meaning special requirements for the underlying storage systems. Thus, the problem is that storage hardware can hardly or not shut down, because the idle time between single requests is too low, implying a high performance loss. To address this problem, Narayanan and Donnelly [10] analyzed different storage systems in an IT infrastructure over a period of a week. The results show that the idle times are low indeed, but nevertheless there are lots of them. Based on this insight the procedure of *Write Off-Loading* [10] was developed. It tries to influence the write and read behavior to maximize idle times between accesses and thereby increases the efficiency of other procedures like spin down or DRPM. In consequence, Write Off-Loading allows a write access to shutdown hard disks by caching the active data on another place in the storage scenery. This happens transparently in the background on block level. By this method allows energy savings of 45 % to 60 %.

2.3 Data Management

In the research field of data management many papers were published which discuss the efficient and especially the energy-optimized deposit of information. Here one question is very important. Namely, how does the data distribution influences the energy consumption in a storage system?

Nightingale [11] introduced a new file system, named *BlueFS*, which was developed to decrease the energy consumption of the underlying storage technology. Therefore, BlueFS divides the file system into hierarchical caches and replicates data on different storage technologies. If information is read, BlueFS decides what storage technology is the best in relation to the expected energy consumption. So it takes advantage of the energy saving states of the underlying technology. An energy saving of up to 55-76 % is achieved.

A number of further work focuses on how to use the available energy saving states of hard disk systems more effectively. Liu, Cheng and Guan [9] describe the employment of modern storage technologies, like flash storage, as caches for hard disks. The advantage of these new technologies is a reduced energy consumption. For example, flash storage has no rotation engine. By using these technologies as caches for frequently used files it is now possible to relieve the underlying hard disks and set them to an energy saving state, because the number of accesses is drastically reduced. According to [9], an energy saving of up to 90 % is achieved and additionally an improvement of the access time is enabled, because flash storage usually offers a faster data access than traditional hard disks. Other papers addressing the same topic are [1,2,17,24]. They capture this approach and describe it for different file systems and IT infrastructures. For example, David and Devaraj [2] developed an intelligent algorithm to manage the cache storage approach more energy efficient then, e.g., traditional cache algorithms like LRU[4]. Additionally, an overview is given about how often specific files are accessed on a hard disk to substantiate the cache approach with flash storage. The result is that only a fractional part of all stored documents is simultaneously active. However, flash storage is essentially more expensive than traditional disk storage. Thus, it is necessary to evaluate the optimal relation between the capacity (of the flash storage) and the cost saving by energy optimization.

A last big group of research papers focuses on RAID systems. The already outlined approaches have the problem that they are not readily available for RAID systems, because those have a very high load (depending on the RAID level) leading to a small number of usable idle times. Otoo and Rotem [12] especially address this problem. The so far discussed ideas of the disk power management (DPM) try to shutdown hard disks as often as possible. This implies additional costs for the time when an access to offline data is attempted, due to the resulting spin up of the regarding media. However, especially in a server environment where a lot of data is stored in RAID systems, the idle times do not become high enough so that the DPM approach gets profitable. Hence, Otoo and Rotem [12] introduced an approach which exchanges information between different RAID combinations on block level. This happens by monitoring the load of particular hard disks in such a way that blocks with a high access rate are shifted to special RAID groups which are constantly running. Thereby, a higher idle time at the majority of the left RAID groups could be reached, what allows a shutdown or spin down of them. This method is similar to this one where flash storage is used as cache. But in contrast [12] offers the advantage to use the existing infrastructure instead of extending it. According to [12], energy savings of up to 50 % are reached and the arising limitations for the access time stay small. There are some similar papers focussing this topic [13,23], which substantiate these results.

[4] Least Recently Used

2.4 Intelligent algorithms

One last relevant research area is concerned with algorithms for the deposition of information on a storage system and how to design it in a more intelligent and energy efficient way.

Pinheiro, Bianchini and Dubnicki published a representing paper [14] for a variety of work which concentrates on the reduction of data in a storage system. The so called *DeDuplication* of data allows a reduction of them what results in a lower access rate for information. And a decrease of this rate means an increase of the energy savings. [14] illustrates this procedure and subsequently evaluates it. According to [14], such data reduction approaches obtain energy savings between 20 % and 61 %.

3 Summary

In this paper the state of the art and the latest research results in the area of energy optimization in storage systems were outlined. In the whole topic of energy optimization of IT hardware the storage systems only make up a small part. However, this small part becomes more important. Especially the fact that hardware components already get constructed very energy efficient means that the storage systems increasingly became the largest energy consumers in an IT infrastructure. The research done in this area is in its beginnings but continues to rise. A lot of papers tried to reuse traditional procedures like DRPM and revive them for the purpose of energy optimization in complex storage systems and storage groups. But a lot of new ideas arose, which focus on the problem of high availability of RAID systems or try to optimize the data distribution by using intelligent algorithms. Thus, the research potential in the area of the energy optimization for storage systems is huge whereby the development of such systems has just begun.

4 CoolSoftware specific summary

The CoolSoftware project focuses on building an energy auto-tuning runtime environment for software components [4]. In this section, we want to summarize the related work discussed in this paper and relate it to the CoolSoftware project.

In this literature study we focused on storage systems. In contrast, CoolSoftware concentrates on energy optimization by an intelligent distribution of software components. Nevertheless, in CoolSoftware the hardware is monitored and controlled by so called *Resource Managers* [3] to achieve energy optimization. With these, it is possible to put specific IT infrastructure components in different energy states and to reflect on them. The energy optimization within these states and within the hardware subsystem is done by the concerned subsystem itself. For this reason the discussed approaches in this paper reveal different starting points for CoolSoftware.

4.1 Hardware manipulation

Some approaches intersect with the objective of CoolSoftware. For example, mechanisms like DRPM (Subsection 2.2) try to manipulate the hardware to find a tradeoff between energy savings and performance loss. In terms of CoolSoftware this could be interesting when trying to use less systems resources for specific tasks and software components and still tempt to save more energy. Certainly, DRPM is only one way in the tier of the storage system to influence the hardware's energy consumption. So it's conceivable that particular storage hardware is shutdown or (partially) restricted in it's performance for appropriate workloads if a fractional part of this performance still suffices to solve a given task. In CoolSoftware these performance requirements for software components have to be described and evaluated. Thereby, two general possibilities appear for saving energy with limited performance:

- On way is to use the given hardware without controlling it so that software components could run on hardware components with just acceptable performance.
- Another way is to control the hardware components so that used components could be influenced in such a way that they just offer a limited performance spectrum for a given task.

4.2 Intelligent algorithms

Papers in this field of research provide different ideas on how to deal with the distribution of data. The input for CoolSoftware could be to adapt these approaches for distributed software components, so that maybe some of these algorithms can be reused or modified.

Another result of Section 2.4 is that data reduction is an important possibility to lower energy costs. Less data means less energy consumption in the storage system. Now, the challenge is to consider how this is useful for CoolSoftware. For example, it is imaginable to only choose those variants of software components which produce a small amount of data in the storage system or which use less resources than other similar components.

4.3 Caching

A bigger part of the literature is concerned in introducing fast flash storage as cache for storage subsystems, because flash storage consumes substantially less energy than traditional storage technologies. Suggestions for intelligent caching and buffering mechanisms can be derived for CoolSoftware. They could be used as a layer between the software components and the storage subsystems (energy aware storage middleware). The advantage is a lower energy consumption and a better performance. However, at this time these aspects are not focused by the CoolSoftware project.

Acknowledgments The project CoolSoftware is part of the Leading-Edge Cluster "Cool Silicon", which is sponsored by the Federal Ministry of Education and Research (BMBF) within the scope of its Leading-Edge Cluster Competition. We would like to thank our advisors Prof. Aßmann and Prof. Meißner and the Silicon Saxony e.V.

References

1. Colarelli, D., Grunwald, D.: Massive arrays of idle disks for storage archives. Proceedings of the 2002 ACM/IEEE conference on Supercomputing pp. 1–11 (2002), http://portal.acm.org/citation.cfm?id=762761.762819&type=series
2. David, F., Devaraj, C., Zhu, Q., Li, Z., Zhou, Y., Pei, C.: Reducing Energy Consumption of Disk Storage Using Power-Aware Cache Management. 10th International Symposium on High Performance Computer Architecture (HPCA'04) pp. 118–118 (2004), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1410070
3. Götz, S., Wilke, C., Schmidt, M., Cech, S., Waltsgott, J., Fritzsche, R.: THEATRE Resource Manager Interface Specification v. 1.0. Tech. rep., TU Dresden (2010)
4. Götz, S., Wilke, C., Schmidt, M., Cech, S., Aßmann, U.: Towards energy auto tuning. In: Proceedings of First Annual International Conference on Green Information Technology (GREEN IT). pp. 978–981 (2010)
5. Guerra, J., Belluomini, W., Glider, J., Gupta, K., Pucha, H.: Energy proportionality for storage: impact and feasibility. SIGOPS Oper. Syst. Rev. 44(1), 35–39 (2010)
6. Gurumurthi, S., Sivasubramaniam, A., Kandemir, M., Franke, H.: DRPM: dynamic speed control for power management in server class disks. 30th Annual International Symposium on Computer Architecture pp. 169–179 (2003), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1206998
7. IDC: The Real Costs to Power and Cool All the World's External Storage. IDC Report, Doc (2008)
8. Kolbasuk McGee, M.: Data Center Energy Consumption Has Doubled Since 2000 – Data Center – InformationWeek (2007), http://www.informationweek.com/news/global-cio/showArticle.jhtml?articleID=197006210
9. Liu, S., Cheng, X., Guan, X., Tong, D.: Energy efficient management scheme for heterogeneous secondary storage system in mobile computers. In: Proceedings of the 2010 ACM Symposium on Applied Computing - SAC '10. p. 251. ACM Press, New York, New York, USA (2010), http://portal.acm.org/citation.cfm?doid=1774088.1774141
10. Narayanan, D., Donnelly, A., Rowstron, A.: Write off-loading. ACM Transactions on Storage 4(3), 1–23 (Nov 2008), http://portal.acm.org/citation.cfm?doid=1416944.1416949
11. Nightingale, E., Flinn, J.: Energy-efficiency and storage flexibility in the blue file system. Proceedings of the 6th conference on Symposium on Opearting Systems Design and Implementation - Volume 6 p. 25 (2004), http://portal.acm.org/citation.cfm?id=1251279
12. Otoo, E., Rotem, D., Tsao, S.C.: Dynamic Data Reorganization for Energy Savings in Disk Storage Systems. Scientific and Statistical Database Management 6187, 322–341 (2010), http://www.springerlink.com/index/2031167TR22878JQ.pdf

13. Pinheiro, E., Bianchini, R.: Energy conservation techniques for disk array-based servers. In: Proceedings of the 18th annual international conference on Supercomputing - ICS '04. p. 68. ACM Press, New York, New York, USA (2004), http://portal.acm.org/citation.cfm?doid=1006209.1006220
14. Pinheiro, E., Bianchini, R., Dubnicki, C.: Exploiting redundancy to conserve energy in storage systems. ACM SIGMETRICS Performance Evaluation Review 34(1), 15 (Jun 2006), http://portal.acm.org/citation.cfm?doid=1140103.1140281
15. Poess, M., Nambiar, R.O.: Tuning servers, storage and database for energy efficient data warehouses. IEEE 26th International Conference on Data Engineering (ICDE) pp. 1006–1017 (2010), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5447806
16. Raffo, D.: Is storage top energy hog in data centers? (2007), http://searchstorage.techtarget.com/news/article/0,289142,sid5_gci1285060,00.html
17. Ruan, X., Manzanares, A., Majors, J., Qin, X.: ECOS: An energy-efficient cluster storage system. In: 28th International Performance Computing and Communications Conference. pp. 79–86. IEEE (Dec 2009), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5403814
18. Tamilarasan, A., Shankarapani, M.K., Qin, X., Mukkamala, S., Sung, A.H.: Integrating energy efficiency and security for storage systems. In: International Conference on Systems, Man and Cybernetics. pp. 2396–2400. IEEE (Oct 2008), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4811653
19. Time for Change: Prediction of energy consumption world-wide (2007), http://timeforchange.org/prediction-of-energy-consumption
20. Waltsgott, J., Götz, S., Fritzsche, R., Cech, S., Wilke, C.: State of the Art : Hardware Energy Management. Tech. Rep. TUD-FI-11-06-Sept. 2011, TU Dresden (2011)
21. Weddle, C., Oldham, M., Qian, J., Wang, A., Reiher, P., Kuenning, G.: PARAID: A gear-shifting power-aware RAID. ACM Transactions on Storage (TOS) 3(3), 13 (2007), http://portal.acm.org/citation.cfm?id=1288783.1289721
22. Wilke, C., Götz, S., Cech, S., Waltsgott, J., Fritzsche, R.: Aspects of Software's Energy Consumption. Tech. Rep. TUD-FI-11-04-Sept. 2011, TU Dresden (2011)
23. Zhu, Q., Chen, Z., Tan, L., Zhou, Y., Keeton, K., Wilkes, J.: Hibernator. ACM SIGOPS Operating Systems Review 39(5), 177 (Oct 2005), http://portal.acm.org/citation.cfm?doid=1095809.1095828
24. Zong, Z., Briggs, M., O'Connor, N., Qin, X.: An Energy-Efficient Framework for Large-Scale Parallel Storage Systems. In: International Parallel and Distributed Processing Symposium. pp. 1–7. IEEE (Mar 2007), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4228287