Computing (2012) 94:973–994
DOI 10.1007/s00607-012-0211-2

Energy-efficient data centers

Junaid Shuja · Sajjad A. Madani · Kashif Bilal · Khizar Hayat · Samee U. Khan · Shahzad Sarwar

Received: 3 January 2012 / Accepted: 10 August 2012 / Published online: 2 September 2012

Abstract Energy consumption of the Information and Communication Technology (ICT) sector has grown exponentially in recent years. A major component of the today's ICT is constituted by the data centers which have experienced an unprecedented growth in their size and population, recently. The Internet giants like Google, IBM and Microsoft house large data centers for cloud computing and application hosting. Many studies, on energy consumption of data centers, point out to the need to evolve strategies for energy efficiency. Due to large-scale carbon dioxide (CO_2) emissions, in the process of electricity production, the ICT facilities are indirectly responsible for considerable

J. Shuja · S. A. Madani · K. Hayat (✉)
COMSATS Institute of Information Technology (CIIT),
Abbottabad 22060, Pakistan
e-mail: khizarhayat@ciit.net.pk

J. Shuja
e-mail: junaidshuja@ciit.net.pk

S. A. Madani
e-mail: madani@ciit.net.pk

K. Bilal · S. U. Khan
Department of Electrical and Computer Engineering,
North Dakota State University, Fargo, ND 58108-6050, USA
e-mail: kashif.bilal@my.ndsu.edu

S. U. Khan
e-mail: samee.khan@ndsu.edu

S. Sarwar
Punjab University College of IT (PUCIT), The University of Punjab,
Lahore, Pakistan
e-mail: s.sarwar@pucit.edu.pk

amounts of green house gas emissions. Heat generated by these densely populated data centers needs large cooling units to keep temperatures within the operational range. These cooling units, obviously, escalate the total energy consumption and have their own carbon footprint. In this survey, we discuss various aspects of the energy efficiency in data centers with the added emphasis on its motivation for data centers. In addition, we discuss various research ideas, industry adopted techniques and the issues that need our immediate attention in the context of energy efficiency in data centers.

Keywords Energy efficiency · Data centers · Causal data

Mathematics Subject Classification 68-02

1 Introduction

The last two decades have seen a phenomenal growth in the Information and Communication Technology (ICT) sector. But this exponential growth has come at a price since the ICT sector has, not only, become a major consumer of energy but has been an agent in the escalation of fuel prices, in fact, inhibiting the widespread deployment of ICT equipment in some areas. Energy consumption is the largest single operational cost of such environments. Due to the increasing demands of energy from ICT sector, with the accompanying high energy cost and depletion of the natural resources, energy efficiency in ICT sector has gained significant importance. Energy efficiency in this sector has been the focus of significant research for the past 10 years or so [1–6]. The net energy consumption, as well as the per-device energy consumption, estimates of these studies may vary but all of them predict a considerable rise in the energy consumption of the ICT sector in the near future. In the USA alone, the networking devices have contributed to 0.07 % of the total annual electricity consumption, in the year 2000, amounting to more than 6 terawatt-hour (TWh) of electricity [7], the capacity of a typical nuclear reactor unit in terms of cost [8]. Similar consumption figures, for the year 2007, in the context of the UK [4], Italy [9] and Japan [4] are reported to be 0.7, 1 and 4 % of the total electricity consumption, respectively, for Japan the escalation has been 20 % as compared to 2006. In non-industrial countries, like India, it is estimated that the networking devices consume about 5 % of the total electricity production [4]. In 2009, the Internet usage has contributed to 9.4 % of the total electricity sales in the USA. It is estimated that the European Telcos accounted for 21.4 TWh of electricity in 2010 [2,3]. Currently, the ICT sector consumes 3 % of the total electricity production worldwide and the amount is estimated to increase by 15–20 % per year [4].

Another important aspect in the need to go for "green energy" is to minimize the emission of green house gases (GHG) by ICT sector. GHG are the major contributors to the global warming. An annual increase of 6 % is predicted in CO_2 emissions from the ICT equipment which will result in 12 % share of ICT equipment in CO_2 emissions by 2020 [3]. A high energy consumption would imply a large carbon footprint of ICT equipment and vice versa. The details of CO_2 emissions, by the ICT sector, are presented in Table 1 (after [10]).

Table 1 CO_2 emissions in the ICT sector

ICT component	% of total
PCs and monitors	40
Servers	23
Fixed-line Telecoms	15
Mobile Telecoms	9
LAN and office Telecoms	7
Printers	6

In recent years, various IT service providers—like IBM, Microsoft, Google and similar large organizations—have deployed data centers for the provision of cloud computing and grid computing services as well as hosting of Internet applications and scientific research. Typically, such data centers have sizes of the order of thousands of servers and switches. The mushroom growth of such data centers, and expansion of the existing ones, has prompted many in-depth studies related to their energy consumption. In 2007, a public law, in the USA, directed the Environmental Protection Agency (EPA) to conduct an analysis on the data centers' electricity demands. The analysis resulted in the "EPA study" [1] that presented data center energy consumption figures, predictions and recommendations for the energy efficiency. The EPA study estimated that the data centers consumed 61 TWh of electricity in 2006 which was projected to grow to 107 TWh in 2011. The electricity consumption by these data centers was 1.5 % of the total electricity sales in the USA for the year 2006. The worldwide electricity usage, in data centers, experienced an average growth rate of 16.7 % from 2000 to 2005. These averages are significantly high in the Asia-Pacific region (23 %) [4]. About 80 % of this electricity usage growth is attributable to servers, while network devices and storage equipments account for 10 %, each in individual capacity. This is due to the fact that the servers are the most used commodity in a data center. It has been estimated, in the said EPA report, that the annual electricity consumption growth of data centers will be 76 % in 2010. Applying all energy efficiency techniques such as device reduction, dynamic voltage/frequency scaling (DVFS), network port reduction, and improved server and storage efficiency will lead to a saving of 25.1 TWh (36 %) in the current electricity bill [1]. Another 31 TWh of electricity can be saved from reduced device demands, uninterruptible power supply (UPS), transformer and cooling efficiency [4]. Moreover, the EPA study suggests stricter energy efficiency measures to achieve an economy of around 80 % in the electricity costs of data centers. A similar study, by ECONET, estimates 68 % savings in energy requirements of a network by applying green technologies [2]. Greencloud simulations [11] for an energy-aware data center put these savings at 31–37 % for different devices. Table 2 presents the percentage of total electricity consumed by various ICT equipments, in the data centers, according to the EPA study [1]. The figures show that network devices account for 5 % of the total electricity consumed. Kliazovich et al. [11] puts this figure at 30 % for the communication fabric. Another aspect of the large-scale data centers is the cost of cooling such environments. Densely populated data centers need efficient cooling and air flow techniques, since the servers are prone to higher failure rates if operated at high temperatures. Cooling techniques further increase the energy consumption, as an indirect outcome of the extensive use of the ICT equipment. The cost of cooling has

Table 2 Data center electricity consumption

ICT component	2006 electricity usage (TWh)	% of total
Infrastructure	30.7	50
Network devices	3.0	5
Storage	3.2	5
Servers	24.5	40
Total	61.4	100

been variously put by many studies: the optimists quoting it at 25 % [12], whereas the pessimists reporting it to be 45 % [13] of the total energy costs. Such conflicting estimates notwithstanding even the optimistic estimate cannot steal the focus from the *green computing*, *green internetworking* and *green data centers* due aspect of energy economy.

Several performance metrics are used as indicators of a data center's power efficiency [14]. The most commonly used metric is the Power Usage Effectiveness (PUE) that is computed as the ratio of total power consumption by a data center to its IT power consumption. The PUE value of a data center may vary from 2 (legacy) to 1.2 (highly efficient) [15]. Therefore, a PUE value of 1.2 indicates that the data center has applied state of the art energy efficiency techniques. Although Google claims that they achieve an ideal PUE mainly by focusing on the data center's cooling efficiency [15], the most adopted energy efficiency technique in data centers is virtualization [16]. An average PUE of 1.86 from a survey of 22 data centers reveals that most of the data center facilities lack state of the art energy efficiency techniques [15].

The rest of the paper is arranged as follows. Section 2 dwells on the issue of various energy efficiency techniques in the context of Dynamic Power Management (DPM), DVFS and virtualization. The server and network resource management aspects are being dealt in Sects. 3 and 4, respectively. A brief overview of various data center simulation tools is also included in Sect. 5. Section 6 concludes the survey.

2 Energy-efficient techniques

Energy efficiency is seldom a design consideration, when networks and data centers are constructed; the primary focus being better performance and high throughput. Without much overhead, many energy efficiency techniques can be applied at the data centers. These include the DPM or device reduction, DVFS, virtualization and the improvement of port, server, storage, cooling, and electricity supply efficiency [5]. The state of the art energy efficiency techniques, thoroughly dealt in the literature, are DPM, DVFS and virtualization. The focus of this survey is, therefore, these three techniques. Some authors classify DVFS under DPM techniques [17]. In this survey, we will classify DPM and DVFS as separate and non-overlapping techniques. DPM is a technique in which the devices are powered on/off dynamically, while in DVFS, the voltage/frequency of each device is changed dynamically. The basic premise of all the energy efficiency techniques is that the data centers are built with redundant and over provisioned resources to meet the workloads rarely experienced in practice. On the whole, the average workload on the ICT equipment and on the data center remains at 30 % of the peak workload [18].

2.1 Dynamic power management

The concept of DPM was first proposed by Benini et al. [19]. DPM is a set of techniques in which devices are dynamically powered off according to the current load. The number of active electronic components is restricted to a minimum while conforming to the required performance and QoS criteria. Typically, the electronic components, both as a whole and as interconnected devices, experience non-uniform workloads during the course of their operation. Future workload of these components can be predicted, with some degree of uncertainty, using probabilistic techniques on the basis of the recent workload history. An efficient power management solution exhibits the following characteristics [6]:

- It is implementable in software, hardware, firmware or VM.
- It is limited to a component, server, cluster or data center.
- It is based on a local or global optimization metric.
- It has local or global resource distribution management.

Dynamic Power Manager or Autonomic Power Manager (APM) is a framework that dynamically powers on/off electronic devices after predicting the future workloads [6,20–22]. Khargharia et al. [21] have proposed a theoretical framework for optimizing the per watt performance of a data center at each level of hierarchy—cluster level, server level and device level. The Autonomic Manager (AM) works in four phases, viz. monitoring, analysis, planning, and execution wherein each phase is dependent on the previous workload history. In order to comply the QoS requirements, the AM calculates the minimal power state by solving the constraint optimization problem. The constraints of the optimization problem are the energy consumed during a state transition while executing a task and the time taken for the transition. This optimization problem is first solved at the highest level of the hierarchy followed by the same at the lower levels. The approach employed a discrete event modeling and simulation technique to show that 72 % of the power savings can be achieved, by adopting the framework, as compared to the static power management techniques. Mastroleon et al. [20] have used a dynamic programming approach, to solve the power-delay trade-off, that considers the workload and thermal status of the systems. The aftermentioned model has a job buffer, a CPU pool and a thermal environment. The objective function is formulated to process all jobs while minimizing the process time. A dynamic programming recursion applied to compute the optimal parameters considers three types of costs: cost of increasing job buffer, the cost of increasing the number of servers and the cost of the configuration changes that are required during the transition of the system from one state to another. Jiang et al. [22] have proposed a similar power management system that uses embedded sensors with the computational models and workload schedulers. The sensors are placed inside the data center to collect real-time data of the operating environment such as the temperature, the humidity and the air flow and send it to the online servers. The data are used by the AM to make decisions on cooling and the air flow. The data are also integrated with the job scheduling process to analyze the heat and air distribution in the data center. The authors present a description of the implementation modus operandi of the framework using a GridMap/iZone programming system. Raghavendra et al. [6] present a power management framework that consists of (a) an efficiency controller that optimizes per-server power consumption, (b) a server manager that limits the maximum thermal power consumption of the server and (c) a group manager that implements maximum thermal power consumption at rack and data center level. The aftermentioned managers work to manage individual and group power budgets inside the data center. A virtual machine controller is used to consolidate the workload on a minimum number of machines so that the rest of the devices can be turned off. The power management and energy efficiency frameworks, discussed above, have only taken into account the processor and memory resources. There are, however, two frameworks worth to mention over here with regard to the autonomic power management of network resources in a data center. The first one is by Heller et al. [23] who propose the ElasticTree, a network-wide power manager for data centers. Their proposed strategy keeps only those links, to be turned on, which are required for the current workload. The rest of the links are powered off. The traces from different data centers, in their study, show that data center workloads show predictable behavior at different times of the week. That is why the scheme relies on the historical data for prediction. The power manager consists of the optimizer, routing manager and power control. The optimizer finds the minimum subset of a network that accommodates the current traffic requirements. The power controller decides, on the basis of the power manager's output, which ports, linecards and switches should be powered on. The routing manger decides which paths should be followed. OpenFlow[1] is used to evaluate ElasticTree on testbeds formed using switches from various vendors. Three testbeds have been formed using different tree architectures; SNMP and NetFlow[2] have been used to provide the traffic data while NetFPGA[3] has been employed for the traffic generation. Energy savings ranging from 25 to 62 % were achieved for various traffic patterns. Switches from different vendors take 30–180 s to pass from the powered-off state to the powered-on state, which means that a significant difference in the traffic pattern from the predicted one may lead to the loss of data and increased latency. In the second work, Costa et al. [24] have proposed the GREEN-NET framework based on three principal components. The Energy Aware Resource Infrastructure (EARI) component switches off unused resources, predicts their next reservation and aggregates all the reservations. The resource management system component uses the open source OAR[4] to control the resources and job allocations. The trust evaluation component selects the nodes that can be delegated to the task of working as a proxy for the sleeping node(s). The Grid5000 platform [25] has been used to evaluate the proposed system. Another aspect of the energy efficiency in networks is that they are not energy proportional. Energy overheads are introduced by fans and transceivers. The energy consumption does not vary with the traffic load on the

1 http://www.openflowswitch.org.

2 http://www.cisco.com/web/go/netflow.

3 http://tinyurl.com/ygcupdc.

4 http://code.google.com/p/google-summer-of-code-2009-oar/.

devices. Although research has been carried out on efficient, energy-proportional links and switches [8,26,27], the real exploitation of this potential lies in the implementation of IEEE 802.3az standard.[5] The IEEE 802.3az standard introduces a Lower Power Idle (LPI) state that recognizes long periods of inactivity to put the device in a low energy state. This technique is known as the Adaptive Link Rate (ALR), stated by many [28,29] as the future of "Green Networking". The ALR technique reduces the energy consumption of a switch by either using it in the LPI state (power off) or by changing the data rate of a link according to the current bandwidth requirement. The technique requires a policy to decide link state switching and an ALR mechanism to implement the policy decisions. The commonly used mechanisms for the ALR are:

1. the IEEE 802.3 Auto-Negotiation protocol,[6]
2. the ALR Medium Access Control (MAC) frame handshake [30] and
3. the IEEE 802.3az Energy-Efficient Ethernet.[7]

The ALR policies can be categorized on the basis of (a) the link sleep/shutdown [8,31,32] and (b) the data/link rate scaling [33,30]. Most of the proposed ALR techniques use the link sleep because it saves more energy as compared to the link rate scaling. Nedevschi et al. [27] present a comparison between the link sleep and link rate scaling and conclude that the link sleep strategy leads to more energy savings when the links are idle. A detailed comparison of the various ALR techniques can be found in [29].

2.2 Dynamic voltage/frequency scaling

This technique is based on the fact that power consumption in a processing chip depends on voltage supplied and is described by the following equation,

$$P = V^2.f \tag{1}$$

where P is the power consumed, V is the voltage and f is the corresponding frequency. Therefore, by reducing the voltage or the switching frequency, the power consumption can be reduced. The frequency reduction only applies to the CPU power since bus, memory, and disks do not depend on the CPU frequency. Moreover, the hardware support is essential to implement the DVFS technique. The Advanced Configuration and Power Interface (ACPI) specification,[8] is an OS-independent power management and configuration standard adopted by most of the manufacturers, nowadays. ACPI defines four power states for a server of which two, the G_0 and G_3, pertain to the powered-on and powered-off states, respectively. The states G_1 and G_2 are further divided into sub-states based on which components are powered on. Most of the DVFS schemes depend on the ACPI standard implementation.

[5] http://grouper.ieee.org/groups/802/3/az.

[6] http://standards.ieee.org/getieee802/download/802.3-2008_section2.pdf.

[7] http://grouper.ieee.org/groups/802/3/az.

[8] http://www.intel.com/ial/powermgm/specs.html.

DVFS has been explored in both server [34–37] and switch domains [38]. PowerNap [34] designs a system that transitions between the active state and the idle state, dynamically, in response to the workload detected through the Network Interface Controller. The LPI state minimizes the response time of the workload that is high in case of sleep mechanisms, but as the systems do not consume energy proportional to the workload, idle state still consumes 70 % of the energy. This issue is addressed in PowerNap by introducing the concept of Redundant Array of Inexpensive Load Sharing (RAILS) that consists of multiple power supply units (PSUs). The PSUs do not draw power for the idle servers, thus saving idle power usage. The design is evaluated over a testbed of blade servers that shows 70 % power usage reduction. Horvath et al. [35] have proposed a DVFS technique for a multi-tier web server architecture in which higher-tier servers invoke services of lower-tier servers. It is assumed that there are soft constraints of end-to-end delay. The DVFS is applied until the end-to-end delay exceeds a threshold value. The convexity of the utilization-delay function implies a common rule of thumb: increase the frequency of the most utilized server and decrease the frequency of the under-utilized server. A prototype of the multi-tier web servers has been built to demonstrates 30 % power savings. Pouwelse et al. [36] propose a central controller that controls the clock speeds and processor voltages with the help of a power-aware application that communicates its future demand to the controller. The tasks are ordered according to the deadlines and priorities. The task with a lower priority is scheduled over an interval such that it has lowest CPU utilization. Thus, low priority tasks are scheduled first so that they can be preempted for higher priority tasks that may arrive later. The algorithm is evaluated with a variable voltage processor, supporting Linux drivers, clock scheduling daemon and power-aware application. The ensued results demonstrate around 50 % power savings. Shang et al. [38] gave the concept of DVFS in the communication fabric wherein each router port predicts the future link load according to the workload history and adjusts the frequency of the link accordingly. The local knowledge usage avoids the communication overhead. The link and input buffer utilization are used as input parameters for the DVFS policy. To evaluate the proposed system, the authors formed an 8×8 mesh network of 64 routers (1 GHz each). The results showed that on average $4.6\times$ power can be saved using the proposed technique while bearing a small increase in the network latency and a decrease in the network throughput.

2.3 Virtualization

Virtualization is a technique that allows the sharing of one physical server among multiple virtual machines (VM), where each VM can serve different applications. The CPU and memory resources can be dynamically provisioned for a VM according to the current performance requirements. This makes virtualization perfectly fit for the requirements of energy efficiency in data centers. The resources of a VM can be provisioned dynamically as the workload demands change for an application. Virtualization is the most adopted power management and resource allocation technique used by the data center operators [16].

Virtualization is implemented in both the server and switch domain but with different objectives. Server domain virtualization usually achieves energy efficiency by sharing limited resources among different applications. While virtualization in the network domain, on the other hand, aims to implement logically different addressing and forwarding mechanisms on the same physical infrastructure. This latter type of virtualization does not provide for energy efficiency [39]. In effect, network resources are burdened by the virtualization techniques. Live migration of VMs in the data center is an active area of research as data have to be transferred from one physical host to another, generating significant amount of traffic [40]. Meng et al. [41] have proposed a VM placement strategy that minimizes the distance between those VMs which have large mutual bandwidth. The VMs are assigned to the hosts in close proximity in order to reduce the distance. The NP-hard VM placement problem is approximated using a two-tier algorithm that takes the traffic matrix between the VMs and cost matrix between the hosts as input. The algorithm partitions the servers into clusters based on the cost between clusters. The VMs are then partitioned into VM clusters in such a manner that minimizes the inter-cluster traffic. Stage et al. [42] discuss the impact of VM live migration on the network resources. A migration scheduler determines the optimal schedule for the migrations, based on the knowledge of their duration, starting time and deadline. The optimal scheduler schedules the live migrations in such a way that the network is not congested by the VM live migration load. The live migrations are also fulfilled in time. Beloglazov et al. [43] have proposed that live migration of VMs can be used to concentrate the jobs on a few physical nodes so that the rest of the nodes can be put in a power saving mode. The allocation, of new requests for VMs, is done by sorting all the VMs in a Modified Best First Decreasing (MBFD) order with respect to the current utilization. The VM is then allocated to a host based on the least deterioration in the power consumption among the hosts. The current allocation of VMs is optimized by selecting the VMs to be migrated on the basis of heuristics related to utilization thresholds. If the current utilization of a host is below a threshold, then all the VMs from that host should be migrated and the host is put in the power saving mode. A similar approach [44] achieves energy efficiency with the help of Limited Look Ahead Control (LLC). The LLC predicts the next state of the system by a behavioral model that depends on the current state, environment input and control input. A profit maximization problem, based on the Service Level Agreement (SLA), is formulated to calculate the maximum number of physical hosts that can be powered off.

In SecondNet [49], a central Virtual Data Center (VDC) manager controls all the resources and VM requests. When the VDC manager creates a VM for the VDC, it assigns the VM, a VDC ID and a VDC IP address, reserves the VM-to-VM and VM-to-core bandwidths, as mentioned in the SLA for the application using the VM. Simulations demonstrate that the system provides a guaranteed bandwidth and high network utilization. A study to measure the impact of virtualization on network parameters—such as throughput, packet delay, and packet loss—has been conducted by Wang et al. [45]. The study is carried out on the Amazon EC2[9] data center where each instance

[9] http://aws.amazon.com/ec2/.

Table 3 Energy efficiency techniques in data centers

Energy efficiency technique	Application	Description and references	*Pros* and *cons*
DPM	Server	Dynamically power off servers according to current load [21]	A framework required to coordinate, predict future workload
	Network	Dynamic network shutdown [11]	Network shutdown degrades performance and is considered taboo
DVFS	Server	Power consumption based on $V^2.f$, can be reduced [34]	A framework required to change voltage, frequency according to future workload. Provides negligible energy savings as compared to DPM
	Network	ALR [8]	A policy required to decide LPI states, link rate. ALR is yet to be discussed in data center context. ALR achieves negligible energy savings
Virtualization	Server	Dynamically provision resources according to QoS requirements, [16]	Most widely used, live migration of VMs effect network resources

of the data center is a Xen VM.[10] Processor utilization and TCP/UDP throughput are measured by CPUTest and TCP/UDPTest programs, respectively. The packet loss is measured by the Badabing tool [46]. The results show an unstable TCP/UDP throughput and a very high packet delay among EC2 instances. The above discussion concludes that although virtualization is a widely adopted power management and resource allocation technique, the impact of VM live migrations has to be considered to reduce the network traffic generated by such activity. Furthermore, live migrations can be adopted to reduce VM-to-VM traffic and such a criterion should be built into the current VM-based energy efficiency solutions. Table 3 summarizes the energy efficiency techniques discussed above.

3 Server resource management

The best option for server resource management, in a data center, is the virtualization. The *pros* and *cons* of virtualization have been discussed in Sect. 2.3. In this section, we will focus on techniques, other than virtualization, for server resource management. The cost of cooling for a densely populated data center may amount to around half of its total energy expenditure [47]. Different techniques have been proposed in the literature in order to lower the energy cost related to the cooling aspect of data centers [48–54]. Abbasi et al. [48] have proposed a temperature- aware server provisioning scheme. The data center is modeled as racks of blade servers with either their front panels or back panels facing each other. The aforementioned is called a hot/cold aisle arrangement. The data center energy consumption is modeled as the

[10] http://xen.org/.

aggregate of processing and cooling energy. The server provisioning is modeled as a subset of servers for a specific time length (usually 1 h) that minimizes the total energy and also provides the maximum CPU utilization. To solve the thermal-aware server provisioning problem, minimax optimization, least re-circulated heat and Computer Power Server Provisioning (CPSP) heuristics are used. A testbed of 30 blade servers, with 4 cores each, was used to evaluate the heuristics. Results show that the minimax heuristic provides the best energy savings. In [49], the researchers have given a similar approach for coordinated job and cooling management in data centers. The intuition behind their highest thermostat setting (HTS) algorithm is that the higher thermostat settings of the Computer Room Air Conditioner (CARC) lead to low energy consumption for the cooling purposes. Researchers [50] have explored the benefits of three schemes for the management of Information Technology (IT) and Cooling Technology (CT) from the aspect QoS and energy efficiency: (a) baseline, (b) coordinated and (c) un-coordinated management. The data center is modeled as a coupled network consisting of a computational network (cyber dynamics) and thermal network (physical dynamics). The computational network is modeled as a first-order queuing system with workload defined in terms of job flow rate. The thermal network is modeled in terms of input and output temperatures of each computing node. The input temperature of a node is the amount of heat received from all of the other nodes while the output temperature of the node is the amount of heat produced by a node. The baseline strategy implements no coordination for the management of IT and CT domains and is beneficial in case when the CT expenses are much smaller than the IT expenses. The uncoordinated strategy represents the working of a typical data center where CT and IT are managed by separate controllers. Because the uncoordinated strategy considers IT and CT as two separate optimization problems, it cannot insure that all of the cyber-physical QoS requirements are met. In the coordinated strategy, the controller adjusts the number of servers according to the predicted and actual workloads while considering QOS and energy efficiency of both CT and IT in a single optimization problem. The data center management strategies are evaluated on a small data center consisting of 32 racks of servers and 4 computer room air conditioning (CRAC) units. The results of the three strategies evaluated in terms of PUE show that while all strategies have similar PUE in case of higher utilization of data center, the PUE is much lower for the coordinated management strategy when the utilization is lower. Wang et al. [51] have studied the joint optimization of energy consumed by computing and cooling units under different prioritized constraints. The optimization problem finds the minimum of cooling and computing energy under three constraints: the data center utilization should provide the threshold performance, the respective component temperatures should be within the safety limits and power consumed by the cooling and computing servers should be within the budget. These constraints can be prioritized as soft or hard constraints according to the preferences of the data center management. Although several DPM techniques focusing on server resource management have been discussed in the Sect. 2.1, one such technique is worth mentioning in this section. Gong et al. [55] have studied the issue of energy-aware server provisioning and load dispatching for connection-intensive applications. They have used real trace data from a data center hosting Windows Live Messenger application. The clients connect to the data center via the data center front-end servers which act as load dispatchers. The load dispatchers evenly distribute the incoming login requests to the connection servers (CS), which authenticates the client. The CS have two hard constraints on them: $L_{\max}$ is the maximum number of login requests a CS can facilitate per second and $N_{\max}$ is the total number of connections the CS has to maintain. Based on the forecasted number of login requests L_{tot} and number of connections N_{tot} for the next period of time t, number of required servers are calculated based on the constraints that for each server $L \leq L_{\max}$ and $N \leq N_{\max}$. Three server provisioning algorithms are described: hysteresis-based provision, forecast-based provision and short-term load forecasting. The server load provisioning algorithms are tested on a testbed of 60 CS and a load dispatcher with real data center traces. The trade-off between energy efficiency and server initiated disconnections (SID) is studied. It is concluded that a hybrid approach achieves the best energy SID trade-off. The hybrid approach distributes load evenly when the load increases and concentrates load to minimum number of CS when the load decreases. Le et al. [56] provide a geo-distribution optimization problem of data centers based on the factors of electricity prices, time zones and green electricity production with limited carbon foot-print. Valancius et al. [57] propose an unconventional model of data centers, called the Nano data centers (NaDa), based on the geo-distribution of data center elements. Services are delivered by the servers placed at the network edge. These servers are controlled by a single ISP and hosted behind the gateways and DSL modems hosted by it. The access to the services, hosted on these servers, is provided by a managed peer-to-peer infrastructure. Cooling is not required in such data centers as the servers are widely distributed. The model has been tested for a Video-on-Demand (VoD) application and the results show about 20–30 % of energy savings as compared to the conventional data centers; memory units, of the order 4 GB, are sufficient to satiate the memory requirements of the data-intensive VoD applications. Another study about the geo-distribution of data center resources can be found in [58].

Berral et al. [59] have used machine learning techniques to predict the future workload based on the models of the hosted applications and machine behaviors. A dynamic scheduler calculates the impact of each job scheduling on power and performance using machine learning models before allocating the resources. Two genetic algorithms have been proposed in [60] for the minimization of both energy and makespan of tasks. Lesser the makespan of a job, lower is its energy consumption. For genetic algorithms, the initial population is created using the *minimum completion time* and *longest job to fastest resource* heuristics. A goal programming approach to solve the multi-optimization problem has been proposed by Khan et al. [61]. It has been shown that the goal programming approach, for the minimization of energy and makespan, converges to the compromised Pareto-optimal solution. The Nash Bargaining Solution (NBS) has also been proposed for this problem in computational grids [62]. All the machines of the grid arrive at a task allocation decision that minimizes the energy consumption and makespan.

3.1 Research issues

As pointed in Sect. 2.3, live migrations of VMs in data center have a network overhead. The optimal migration of VMs that limit the overhead on the network resources is an active area of research [40]. Moreover, live migrations of VMs may lead to non-optimized VM placement. In this scenario, two VMs that have large mutual bandwidth may be placed at distant physical machines increasing the network communications significantly [41]. Moreover, the optimal placement of VMs such that it minimizes the cooling cost is a topic of active research. Therefore, the problem of optimal placement of VMs under the constraints of energy efficiency, cooling efficiency, QoS, SLAs, network overheads and data center topology has many research issues to be explored.

As the heat is constantly dissipated from servers, device failure rate increases which affects the reliability and availability of the systems. Cooling data center environments almost double the cost of energy. The joint optimization problem of minimizing thermal and computing energy under the constraints of budget, threshold component temperatures, component utilization and required performance metrics is an active area of research [48–52].

The research focus in data centers resource allocation has been on meeting SLA and providing high performance with little attention to minimizing energy [63]. Due to the variability of workload, the thought of energy efficiency is considered taboo as it can lead to SLA violations. Different techniques of workload prediction have their own *pros* and *cons* [55]. While history-based prediction is the most commonly deployed technique, it is not efficient for all workload scenarios. The penalties paid in terms of business impact due to violation of SLA can be huge. Therefore, intelligent workload prediction strategies have to be formulated [52].

4 Network resource management

The commonly adopted network architectures, in data centers, are multi-tier, that is, two-tier (2T), three-tier (3T) and three-tier high-speed architecture (3Ths) [64]. The 3T architecture is the most common in large-scale data centers wherein the core layer connects the data center to the Internet backbone, the aggregation layer provides diverse functionalities (such as content switching, SSL and firewall) and the access layer connects the internal data servers placed in a rack-blade assembly. There are multiple links from one tier to another which, along with the multiple internal servers, ensure availability and fault tolerance in the data center, though at the cost of some redundancy. The 2T architecture does not have the aggregation layer. The 3T architecture is graphically depicted in Fig. 1 (after [13]).

The 10 GE links are replaced by 100 GE while 1 GE links are replaced by 10 GE links in 3Ths architecture. The 3Ths architecture provides higher bandwidth in the core and aggregation layer of the data center. A higher bandwidth also helps to reduce the number of switches in the core and aggregation layer but it comes with a price of higher energy consumption and higher switch costs. Multi-tier network architectures limit the size of a single L2 layer domain to around 4,000 servers. The size of the IP subnet within a domain is further limited to a few hundred servers due to the overhead of the broadcast traffic. VLANS are used to further divide a domain into subnets. These multi-tier architectures are widely adopted but they are inherently dense, inflexible and non-scalable for a wide range of applications that need thousands of servers to work

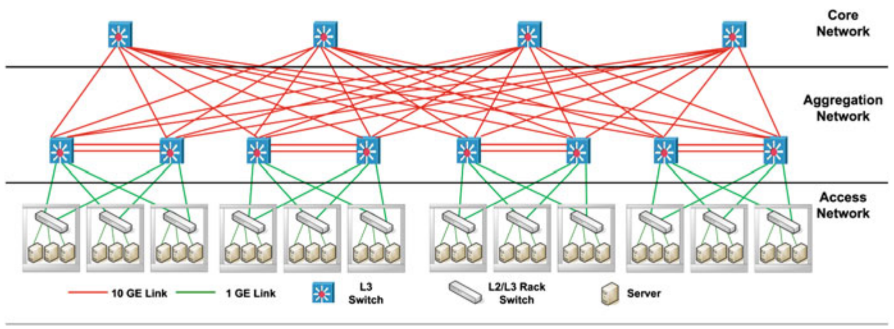

Fig. 1 Three-tier (3T) network architecture

with, for example, MapReduce [65] and Hadoop.[11] When an application expands over several L2 domains, the resources are fragmented. Furthermore, multi-tier network architectures do not provide server to server connectivity and servers have to route data through L3 routers to communicate with each other. Server to server connectivity with large and scalable L2 domains is the basic requirements of new state-of-the-art Internet applications.

4.1 Network architectures

Most of the proposed network architectures for data centers focus on scalability, fault-tolerance and high-speed connectivity for server-to-server and server-to-client throughput. Some of these network architectures try to achieve a 1:1 oversubscription ratio [12]. The oversubscription ratio is defined as the ratio of total bandwidth provided by a layer to the total bandwidth provided by its higher layer. For example, a rack of 48 servers with 1 Gbps port each, having twenty 1 Gbps uplinks to the aggregation switches, has an oversubscription ratio of 2.4—limiting the per server available bandwidth to 416 Mbps. Conventional data center architectures have oversubscription ratio greater than 1, at every tier, that significantly limits the theoretical throughput of the data center. To achieve 1:1 oversubscription ratio and full aggregate bandwidth, the state-of-the-art network architectures have the following salient features:

1. Accommodation of more switches at aggregation and core layers.
2. The use of commodity switches instead of high-end enterprise switches.
3. Be scalable to accommodate more links by adopting variants of mesh, Clos and hierarchical architecture and
4. Building a routing protocol to support the architecture.

DCell [66] is a recursively built network architecture where high-level DCells are formed from several low-level fully connected DCells. DCell Fault-tolerant Routing protocol (DFR) is a single-path, recursively built routing protocol to support the recursive architecture of DCell. When a packet arrives at a node, it checks whether it is

[11] http://hadoop.apache.org/.

the destination and if so, it sends the packet to the higher layer. Otherwise, it checks whether the packet belongs to the same DCell and if so, broadcasts the packet within the DCell; otherwise it uses the DCellRouting protocol to calculate the destination path. BCube [67] is a modular data center network architecture based on the DCell. BCube supports server-to-server communication in all forms: one-to-all and all-to-all. The network architecture has additional ports attached to each server and Commodity off-the-shelf (COTS) mini-switches as crossbars to support server-to-server communication. The BCube is a recursively defined structure with BCube_0 a subset of n servers connected to an n-port switch and BCube_k consists of n Bcube_{k-1}. Servers are named according to their location in the BCube structure. A routing protocol, BSR, is defined to support this architecture. BSR decides, which path a packet should follow from the source to the destination, after probing the network and encoding the path in the packet header. The source decides the routing path and the intermediate servers and switches have to route the packet accordingly, without being involved in the routing decisions. The routing protocol uses a Breadth First Search (BFS) strategy to find parallel paths to the destination. The BCube protocol resides at layer 2.5, between the TCP/IP protocol driver and the Ethernet Network Driver Interface Specification (NDIS). A testbed, of 16 servers and eight 8-port mini-switches, is used to evaluate the performance of BCube. The results show greater per-server and server-to-server throughput than that of tree architectures.

Monsoon [68] is a mesh-like network architecture built from low cost commodity switches. Load balancing servers are used to distribute the load on the servers and avoid congestion within the data center. Commodity hardware is used to scale out, rather than scale up, the data center. Server-to-server communication is provided by MAC-in-MAC encapsulation.[12] This results in the overhead of encapsulation/decapsulation at each hop. Monsoon requires that each switch maintain an entry, in its forwarding table, for every other switch. Switches track the IP and MAC address of their adjacent switches and advertise them in a Link State Advertisement (LSA) to perform the required encapsulation/decapsulation. Programmable switches are required to control the forwarding table of switches. A fat-tree topology is proposed in [12,69]. Al-Fares et al. [69] proposed a Clos-based fat-tree k-ary architecture using the commodity switches. In fat-tree topology, all switching elements are identical, allowing the use of commodity switches at all the tiers. A single routing path is defined between any two hosts in this scheme. This may lead to bottlenecks, which are avoided by adopting an IP forwarding technique. Greenberg et al. [12] proposed VL2 with similar characteristics. VL2 uses Clos network topology to provide high server-to-server bandwidth, and valiant load balancing to provide load distribution where each server randomly chooses the path of its flow from the available paths. Portland [70] is a scalable and fault-tolerant data center network fabric built from pods. The Portland protocol works on layer-2 forwarding mechanisms. Each switch is assigned a pseudo MAC (PMAC) address. All forwarding is done on the basis of PMAC. The switch maintains the forwarding table of MAC addresses of its attached servers. A Location Discovery Protocol is used by the switches to establish their position within the topology. Although most

[12] http://www.ieee802.org/1/pages/802.1ah.html.

of the network architectures described in the above section overcome the drawbacks of multi-tier architectures, they are not adopted in industry, mainly due to the support of hardware industry (Cisco) for the multi-tier architecture.

The aforementioned data center network architectures provide energy savings by deploying commodity switches instead of enterprise/high-end switches. Commodity switches are cheaper than enterprise switches and also consume lesser energy. Moreover, these network architectures over provision the resources and deploy multiple links between nodes. The over-provisioned resources provide energy efficiency by supporting energy-efficient routing protocols and resource consolidation. Detailed comparison of data center architectures can be found in [71,72].

4.2 Network protocols

Most of the data center architectures discussed in the above section use custom routing protocols. Shang et al. [48] propose an energy-aware routing protocol that provides efficient routing using fewer network devices and places the rest of the devices in the sleep mode for energy efficiency. The minimum subset of network devices needed for efficient routing is found by first calculating the total throughput of the data center. Then, network devices are gradually removed from the topology until the throughput decreases to a minimum tolerable performance level. The devices removed from the topology are switched off. Chiaraviglio et al. [73] have proposed a mathematical model of energy-efficient communication between a content provider and an ISP. Many researchers [74–77] have discussed the drawbacks of Equal-Cost Multiple-Path (ECMP) routing in data centers. As most of the network architectures deployed in data centers are tree based, multiple equal cost paths exist between any two pairs of systems. With ECMP, the router calculates the hash of a packet based on its header and assigns a route to a hash value. Packets belonging to the same flow (header) are routed to the same link to avoid an out-of-order delivery that may result in unbalanced traffic on various links, in case of data centric flows. The remedies, related to the ECMP routing, have been discussed in [74–77]. Energy-aware routing is a NP-hard problem as formally proved by many researchers, for instance [78].

4.3 Research issues

Kliazovich et al. [13] argue that network load consolidation can result in substantial energy savings. But the network traffic consolidation techniques are usually treated as an unaffordable luxury by the data centers where the supreme business objective is to provide maximum throughput at minimum latency. The exercise of shutting down switches and consolidating traffic on other switches can have severe effects on network throughput, latency, congestion and routing protocols. These issues have not been focus of any research so far. We are particularly interested in the case of causal data in such scenarios. Causality implies that the data sent in an order should be received in the same order. Communication protocols like TCP/IP make sure that the causality of data is complied with. The concept of causality is associated with the content sharing websites and teleconferencing data. Content sharing websites produce Data Intensive

Table 4 Energy efficiency-related research in data center networks

Project name	Goal	QoS	Simulations
DENS [13]	Server consolidation while avoiding hotspots	Yes	GreenCloud [11]
ElasticTree [23]	Keep minimum subset of switches alive	No, predicted traffic may conflict with actual traffic	OpenFlow[a]
GREEN-NET [24]	Advance network resources reservation	No, as application demands may change	Grid 5000 [25]
A routing protocol [48]	Keep minimum switches on for tolerable throughput	Partially, as throughput is considered but latency is not	Not mentioned

[a] http://www.openflowswitch.org

Workloads (DIW) that have little computation but larger bandwidth requirements. Despite the fact that web applications buffer the data, it is considered to be lost after a threshold latency time. As mentioned earlier, the Dynamic Network Shutdown (DNS) technique increases the latency and decreases the throughput of the network, therefore, affecting QoS requirements of causal data. Different traffic scheduling techniques can be applied to address this issue. Preemptive scheduling can be used to preempt the non-causal data and provide priority to the causal data. This technique has two drawbacks. First, all the data in a data center may be causal and preemption might not be a choice. Second, preemption increases latency of the preempted data in any case. The concept of advanced network resource allocation can also be used under this scenario. Internet applications can reserve the network resources for a DIW and those resources may not be part of the DNS exercise. These techniques can be adopted from the wireless transmission domain [79]. Table 4 summarizes the research regarding energy efficiency in data center networks.

5 Simulation tools

Most of the researchers have used testbeds to evaluate their proposed schemes of energy efficiency in data centers [12,36,70,74]. The main disadvantage is, however, the issue of scalability of the testbed to the real world data center environments that are many times larger than the testbed environment. Ripcord [80] provides a modular platform for data center networking and monitoring that employ a 160-node cluster forming fat-tree architecture with programmable switches configured with OpenFlow. The visualization of network states is provided by a GUI. Implementation of some of the routing algorithms presented in previous sections has also been provided by Ripcord. Three well-known data center energy-aware simulation environments are MDCSim [81], CloudSim [82] and GreenCloud [11]. A comparison of the three environments is provided in [11] which indicates that GreenCloud is by far the best choice for energy-aware data center simulations. GreenCloud has several advantages over MDCSim and CloudSim simulators. GreenCloud is an extension of NS2,[13] a simulator that captures

[13] http://www.isi.edu/nsnam/ns/l.

the packet-level simulation details, while MDCSim and CloudSim are event-based simulators that do not capture packet-level details. Both MDCSim and CloudSim do not provide GUI support, while NS2-based simulators come with the support of Nam animator.[14] GreenCloud implements the detailed communication model capturing the dynamics of various protocols, while MDCSim and CloudSim only implement limited communication models. Energy consumption can be measured in GreenCloud very accurately as energy levels are changed on packet arrival, task execution and completion. CloudSim provides no energy model while MDCSim provides only rough estimates. GreenCloud also implements three power saving algorithms in the form of DPM, DVFS and DPM + DVFS. MDCSim and CloudSim are event-based simulators and can simulate a data center of 100,000 nodes easily while GreenCloud can simulate a data center of a few thousand nodes. As GreenCloud has the advantage of packet-level detailed communication modeling and support of a GUI, its usage is preferred in energy-aware data center simulations.

6 Conclusion

A typical data center consumes a large amount of electricity due to its redundantly built architecture. The redundant elements of a data center can be placed in low-power states to save energy. Workload consolidation is done on a minimum number of servers according to the predicted workload. The contemporary research focuses on the issues of workload prediction, virtualization and autonomic power management. Consolidation of traffic on minimum number of switches has received little attention. We have discussed both server and switch devices for energy efficiency purposes while dwelling on the research issues regarding the energy efficiency in data center networks. The case of causal data needs special attention as DNS techniques affect the latency of causal data. Message queue preemption at switches and wireless transmission scheduling techniques are proposed as a solution for such scenarios.

References

1. Brown R (2007) Report to congress on server and data center energy efficiency public law 109–431. Environ Prot 109:431
2. Bolla R, Davoli F, Bruschi R, Christensen K, Cucchietti F, Singh S (2011) The potential impact of green technologies in next-generation wireline networks: is there room for energy saving optimization? IEEE Commun Mag 49(8):80–86
3. Webb M (2008) SMART 2020: Enabling the low carbon economy in the information age. Tech. rep, Climate Group on behalf of the Global eSustainability Initiative (GeSI)
4. Koomey JG (2008) Worldwide electricity used in data centers. Environ Res Lett 3(3):1–8 (IOPscience)
5. Masanet E, Brown R, Shehabi A, Koomey J, Nordman B (2011) Estimating the energy use and efficiency potential of U.S. data centers. Proc IEEE 99(8):1440–1453
6. Raghavendra R, Ranganathan P, Talwar V, Wang Z, Zhu X (2008) No power struggles: coordinated multi-level power management for the data center. SIGOPS Oper Syst Rev 42(2):48–59
7. Gupta M, Singh S (2003) Greening of the internet. In ACM SIGCOMM
8. Gupta M, Singh S (2007) Using low-power modes for energy conservation in ethernet lans. In: 26th IEEE international conference on computer communications, INFOCOM 2007, pp 2451–2455

[14] http://www.isi.edu/nsnam/nam/l.

9. Bolla R, Bruschi R, Lombardo C, Suino D (2011) Evaluating the energy-awareness of future internet devices. In: IEEE 12th international conference on high performance switching and routing (HPSR), 2011, pp 36–43
10. Zeadally S, Khan S, Chilamkurti N. (2011) Energy-efficient networking: past, present, and future. J Supercomput, pp 1–26. doi:10.1007/s11227-011-0632-2
11. Kliazovich D, Bouvry P, Audzevich Y, Khan S (2010) Greencloud: a packet-level simulator of energy-aware cloud computing data centers. In: IEEE conference on global telecommunications, 2010, GLOBECOM 2010, pp 1–5
12. Greenberg A, Hamilton JR, Jain N, Kandula S, Kim C, Lahiri P, Maltz DA, Patel P, Sengupta S (2009) Vl2: a scalable and flexible data center network. In: Proceedings of the ACM SIGCOMM 2009 conference on data communication, SIGCOMM '09. ACM, New York, pp 51–62
13. Kliazovich D, Bouvry P, Khan S (2010) Dens: data center energy-efficient network-aware scheduling. In: Green computing and communications (GreenCom), 2010 IEEE/ACM international conference on and international conference on Cyber, physical and social computing (CPSCom), pp 69–75
14. Chen D, Henis E, Kat RI, Sotnikov D, Cappiello C, Ferreira AM, Pernici B, Vitali M, Jiang T, Liu J, Kipp A (2011) Usage centric green performance indicators. SIGMETRICS Perform Eval Rev 39(3), pp 92–96. doi:10.1145/2160803.2160868
15. Hoelzle U, Barroso LA (2009) The datacenter as a computer: an introduction to the design of warehouse-scale machines. Morgan and Claypool Publishers, San Rafael. http://www.morganclaypool.com/doi/pdf/10.2200/s00193ed1v01y200905cac006
16. Chernicoff D (2009) The shortcut guide to data center energy efficiency. Realtime Publisher, New York. http://nexus.realtimepublishers.com/sgdcee.php?ref=gbooks
17. Valentini GL, Lassonde W, Khan SU, Min-Allah N, Madani SA, Li J, Zhang L, Wang L, Ghani N, Kolodziej J et al (2011) An overview of energy efficiency techniques in cluster computing systems. Clust Comput. doi:10.1007/s10586-011-0171-x
18. Liu J, Zhao F, Liu X, He W (2009) Challenges towards elastic power management in internet data centers. In: 29th IEEE international conference on distributed computing systems workshops, 2009, ICDCS workshops '09, pp 65–72
19. Benini L, Bogliolo A, De Micheli G (2000) A survey of design techniques for system-level dynamic power management. In: IEEE transactions on very large scale integration (VLSI) systems, 8(3), pp 299–316
20. Mastroleon L, Bambos N, Kozyrakis C, Economou D (2005) Automatic power management schemes for internet servers and data centers. In: IEEE global telecommunications conference, 2005, GLOBECOM '05, p 5
21. Khargharia B, Hariri S, Szidarovszky F, Houri M, El-Rewini H, Khan S, Ahmad I, Yousif M (2007) Autonomic power performance management for large-scale data centers. In: IEEE international symposium on parallel and distributed processing symposium, IPDPS 2007, pp 1–8
22. Jiang N, Parashar M (2009) Enabling autonomic power-aware management of instrumented data centers. In: IEEE international symposium on parallel distributed processing, IPDPS 2009, pp 1–8
23. Heller B, Seetharaman S, Mahadevan P, Yiakoumis Y, Sharma P, Banerjee S, Mckeown N (2010) Elastictree: saving energy in data center networks. In: Proceedings of the 7th USENIX conference on networked systems design and implementation (NSDI'10). USENIX Association, Berkeley, CA
24. Da Costa G, Gelas JP, Georgiou Y, Lefevre L, Orgerie AC, Pierson JM, Richard O, Sharma K (2009) The green-net framework: energy efficiency in large scale distributed systems. In: IEEE international symposium on parallel distributed processing, 2009, IPDPS 2009, pp 1–8
25. Cappello F, Caron E, Dayde M, Desprez F, Jegou Y, Primet P, Jeannot E, Lanteri S, Leduc J, Melab N, Mornet G, Namyst R, Quetier B, Richard O (2005) Grid'5000: a large scale and highly reconfigurable grid experimental testbed. In: The 6th IEEE/ACM international workshop on grid computing, p 8
26. Gunaratne C, Christensen K, Nordman B, Suen S (2008) Reducing the energy consumption of ethernet with adaptive link rate (alr). IEEE Trans Comput 57(4):448–461
27. Nedevschi S, Popa L, Iannaccone G, Ratnasamy S, Wetherall D (2008) Reducing network energy consumption via sleeping and rate-adaptation. In: Proceedings of the 5th USENIX symposium on networked systems design and implementation, NSDI'08. USENIX Association, Berkeley, pp 323–336
28. Bianzino AP, Chaudet C, Rossi D, Rougier JL (2010) A survey of green networking research. CoRR. arXiv:1010.3880v1

29. Bilal K, Khan S, Nasro-Min-Allah Madani S, Hayat K, Khan MI, Kolodziej J, Wang L, Zeadally S (2012) A survey on green communications using adaptive link rate, to cluster computing. J Clust Comput, pp 1–15. doi:10.1007/s10586-012-0225-8
30. Anand H, Reardon C, Subramaniyan R, George A (2006) Ethernet adaptive link rate (alr): Analysis of a mac handshake protocol. In: Proceedings of the 31st IEEE conference on local computer networks, 2006, pp 533–534
31. Gupta M, Grover S, Singh S (2004) A feasibility study for power management in lan switches. In: Proceedings of the 12th IEEE international conference on network protocols, 2004, ICNP 2004, pp 361–371
32. Gupta M, Singh S (2007) Dynamic ethernet link shutdown for energy conservation on ethernet links. In: IEEE international conference on communications, 2007, ICC '07, pp 6156–6161
33. Gunaratne C, Christensen K, Suen S (2006) Ethernet adaptive link rate (alr): analysis of a buffer threshold policy. In: Proceedings of IEEE GLOBECOM
34. Meisner D, Gold BT, Wenisch TF (2009) Powernap: eliminating server idle power. SIGPLAN Not 44:205–216
35. Horvath T, Abdelzaher T, Skadron K, Liu X (2007) Dynamic voltage scaling in multitier web servers with end-to-end delay control. IEEE Trans Comput 56(4):444–458
36. Pouwelse J, Langendoen K, Sips H (2001) Energy priority scheduling for variable voltage processors. In: International symposium on low-power electronics and design, pp 28–33
37. Bianchini R, Rajamony R (2004) Power and energy management for server systems. Computer 37(11):68–76
38. Shang L, Peh LS, Jha N (2003) Dynamic voltage scaling with links for power optimization of interconnection networks. In: Proceedings of the 9th international symposium on high-performance computer architecture, 2003, HPCA-9 2003, pp 91–102
39. Sherwood R, Gibb G, Yap KK, Appenzeller G, Casado M, McKeown N, Parulkar G (2009) Flowvisor: a network virtualization layer. Tech. rep., Deutsche Telekom Inc. R&D Lab, Stanford University, Nicira Networks
40. Voorsluys W, Broberg J, Venugopal S, Buyya R (2009) Cost of virtual machine live migration in clouds: a performance evaluation. In: Proceedings of the 1st international conference on cloud computing, CloudCom '09. Springer-Verlag, Berlin, Heidelberg, pp 254–265
41. Meng X, Pappas V, Zhang L (2010) Improving the scalability of data center networks with traffic-aware virtual machine placement. In: Proceedings of IEEE INFOCOM, 2010, pp 1–9
42. Stage A, Setzer T (2009) Network-aware migration control and scheduling of differentiated virtual machine workloads. In: Proceedings of the 2009 ICSE workshop on software engineering challenges of cloud computing, CLOUD '09. IEEE Computer Society, Washington, DC, pp 9–14
43. Beloglazov A, Buyya R (2010) Energy efficient resource management in virtualized cloud data centers. In: 10th IEEE/ACM international conference on cluster, cloud and grid computing (CCGrid), 2010, pp 826–831
44. Kusic D, Kephart J, Hanson J, Kandasamy N, Jiang G (2008) Power and performance management of virtualized computing environments via lookahead control. In: International conference on autonomic computing, 2008. ICAC '08, pp 3–12
45. Wang G, Ng TSE (2010) The impact of virtualization on network performance of amazon ec2 data center. In: Proceedings of the 29th conference on information communications, INFOCOM'10, pp 1163–1171. IEEE Press, Piscataway. http://dl.acm.org/citation.cfm?id=1833515.1833691
46. Sommers J, Barford P, Duffield N, Ron A (2005) Improving accuracy in end-to-end packet loss measurement. In: Proceedings of the 2005 conference on applications, technologies, architectures, and protocols for computer communications, SIGCOMM '05. ACM, New York, pp 157–168
47. Wang L, Khan S, Dayal J (2011) Thermal aware workload placement with task-temperature profiles in a data center. J Supercomput 61:780–803
48. Abbasi Z, Varsamopoulos G, Gupta SKS (2010) Thermal aware server provisioning and workload distribution for internet data centers. In: Proceedings of the 19th ACM international symposium on high performance distributed computing, HPDC '10. ACM, New York, pp 130–141
49. Banerjee A, Mukherjee T, Varsamopoulos G, Gupta S (2010) Cooling-aware and thermal-aware workload placement for green hpc data centers. In: International conference on green computing, 2010, pp 245–256
50. Parolini L, Sinopoli B, Krogh BH, Wang Z (2011) A cyber-physical systems approach to data center modeling and control for energy efficiency. Proc IEEE 99:1–15

51. Wang Z, Tolia N, Bash C (2010) Opportunities and challenges to unify workload, power, and cooling management in data centers. In: Proceedings of the 5th international workshop on feedback control implementation and design in computing systems and networks, FeBiD '10. ACM, New York, pp 1–6
52. Wang L, von Laszewski G, Huang F, Dayal J, Frulani T, Fox G (2011) Task scheduling with ann-based temperature prediction in a data center: a simulation-based study. Eng Comput (Lond) 27(4):381–391
53. Wang L, Fu C (2010) Research advances in modern cyberinfrastructure. N Gener Comput 28:111–112
54. Wang L, von Laszewski G, Kunze M, Tao J (2010) Cloud computing: a perspective study. N Gener Comput 28(2):137–146
55. Chen G, He W, Liu J, Nath S, Rigas L, Xiao L, Zhao F (2008) Energy-aware server provisioning and load dispatching for connection-intensive internet services. In: Proceedings of the 5th USENIX symposium on networked systems design and implementation, NSDI'08. USENIX Association, Berkeley, pp 337–350
56. Le K, Bianchini R, Martonosi M, Nguyen T (2009) Cost-and energy-aware load distribution across data centers. In: Proceedings of HotPower, Citeseer
57. Valancius V, Laoutaris N, Massoulié L, Diot C, Rodriguez P (2009) Greening the internet with nano data centers. In: Proceedings of the 5th international conference on emerging networking experiments and technologies, CoNEXT '09. ACM, New York, pp 37–48
58. Vrbsky S, Lei M, Smith K, Byrd J (2010) Data replication and power consumption in data grids. In: IEEE 2nd international conference on cloud computing technology and science (CloudCom), 2010, pp 288–295
59. Berral JL, In Goiri, Nou R, Julià F, Guitart J, Gavaldà R, Torres J (2010) Towards energy-aware scheduling in data centers using machine learning. In: Proceedings of the 1st international conference on energy-efficient computing and networking, e-Energy '10. ACM, New York, pp 215–224
60. Joanna Koloodziej SUK, Xhafa F (2011) Genetic algorithms for energy-aware scheduling in computational grids. In: 6th IEEE international conference on P2P, parallel, grid, cloud, and, internet computing (3PGCIC)
61. Khan S, Min-Allah N (2011) A goal programming based energy efficient resource allocation in data centers. J Supercomput, pp 1–18. doi:10.1007/s11227-011-0611-7
62. Khan S, Ahmad I (2009) A cooperative game theoretical technique for joint optimization of energy consumption and response time in computational grids. IEEE Trans Parallel Distributed Sys 20(3): 346–360
63. Buyya R, Beloglazov A, Abawajy JH (2010) Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. CoRR abs/1006.0308. arXiv:1006.0308v1
64. Cisco (2007) Cisco data center infrastructure 2.5 design guide. Cisco press
65. Dean J, Ghemawat S (2008) Mapreduce: simplified data processing on large clusters. Commun ACM 51:107–113
66. Guo C, Wu H, Tan K, Shi L, Zhang Y, Lu S (2008) Dcell: a scalable and fault-tolerant network structure for data centers. In: Proceedings of the ACM SIGCOMM 2008 conference on data communication, SIGCOMM '08. ACM, New York, pp 75–86
67. Guo C, Lu G, Li D, Wu H, Zhang X, Shi Y, Tian C, Zhang Y, Lu S (2009) Bcube: a high performance, server-centric network architecture for modular data centers. SIGCOMM Comput Commun Rev 39(4):63–74
68. Greenberg A, Lahiri P, Maltz DA, Patel P, Sengupta S (2008) Towards a next generation data center architecture: scalability and commoditization. In: Proceedings of the ACM workshop on programmable routers for extensible services of tomorrow, PRESTO '08. ACM, New York, pp 57–62
69. Al-Fares M, Loukissas A, Vahdat A (2008) A scalable, commodity data center network architecture. SIGCOMM Comput Commun Rev 38:63–74
70. Niranjan Mysore R, Pamboris A, Farrington N, Huang N, Miri P, Radhakrishnan S, Subramanya V, Vahdat A (2009) Portland: a scalable fault-tolerant layer 2 data center network fabric. SIGCOMM Comput Commun Rev 39:39–50
71. Bilal K, Khan Samee U, Kolodziej J, Zhang L, Madani SA, Min-Allah N, Wang L, Chen D (2010) A comparative study of data center network architectures. In: 26th EUROPEAN conference on modelling and simulation, ECMS 2012
72. Popa L, Ratnasamy S, Iannaccone G, Krishnamurthy A, Stoica I (2010) A cost comparison of datacenter network architectures. In: Proceedings of the 6th international conference, Co-NEXT '10, pp 16:1–16:12. ACM, New York

73. Chiaraviglio L, Matta I (2010) Greencoop: cooperative green routing with energy-efficient servers. In: Proceedings of the 1st international conference on energy-efficient computing and networking, e-Energy ’10. ACM, New York, pp 191–194
74. Al-fares M, Radhakrishnan S, Raghavan B, Huang N, Vahdat A (2010) Hedera: Dynamic flow scheduling for data center networks. In: Proceedings of networked systems design and implementation (NSDI) symposium
75. Raiciu C, Barre S, Pluntke C, Greenhalgh A, Wischik D, Handley M (2011) Improving datacenter performance and robustness with multipath tcp. SIGCOMM Comput Commun Rev 41:266–277
76. Alizadeh M, Greenberg A, Maltz DA, Padhye J, Patel P, Prabhakar B, Sengupta S, Sridharan M (2010) Data center tcp (dctcp). SIGCOMM Comput Commun Rev 40:63–74
77. Wu H, Feng Z, Guo C, Zhang Y (2010) Ictcp: Incast congestion control for tcp in data center networks. In: Proceedings of the 6th international conference, Co-NEXT ’10, pp 13:1–13:12. ACM, New York
78. Shang Y, Li D, Xu M (2010) Energy-aware routing in data center network. In: Proceedings of the 1st ACM SIGCOMM workshop on Green networking, Green Networking ’10. ACM, New York, pp 1–8
79. Gamal AE, Nair R, Prabhakar B, Uysal-biyikoglu E, Zahedi S (2002) Energy-efficient scheduling of packet transmissions over wireless networks. In: Proceedings of the INFOCOM conference, pp 1773–1783
80. Heller B, Erickson D, McKeown N, Griffith R, Ganichev I, Whyte S, Zarifis K, Moon D, Shenker S, Stuart S (2010) Ripcord: a modular platform for data center networking. SIGCOMM Comput Commun Rev 41:457–458
81. Lim SH, Sharma B, Nam G, Kim EK, Das C (2009) Mdcsim: A multi-tier data center simulation, platform. In: IEEE international conference on cluster computing and workshops, 2009. CLUSTER ’09, pp 1–9
82. Buyya R, Ranjan R, Calheiros R (2009) Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: challenges and opportunities. In: International Conference on high performance computing simulation, 2009. HPCS ’09, pp 1–11