Cluster Comput
DOI 10.1007/s10586-014-0404-x

Energy-efficient data replication in cloud computing datacenters

Dejene Boru · Dzmitry Kliazovich · Fabrizio Granelli · Pascal Bouvry · Albert Y. Zomaya

Received: 18 March 2014 / Revised: 4 July 2014 / Accepted: 24 September 2014
© Springer Science+Business Media New York 2015

Abstract Cloud computing is an emerging paradigm that provides computing, communication and storage resources as a service over a network. Communication resources often become a bottleneck in service provisioning for many cloud applications. Therefore, data replication which brings data (e.g., databases) closer to data consumers (e.g., cloud applications) is seen as a promising solution. It allows minimizing network delays and bandwidth usage. In this paper we study data replication in cloud computing data centers. Unlike other approaches available in the literature, we consider both energy efficiency and bandwidth consumption of the system. This is in addition to the improved quality of service QoS obtained as a result of the reduced communication delays. The evaluation results, obtained from both mathematical model and extensive simulations, help to unveil performance and energy efficiency trade-offs as well as guide the design of future data replication solutions.

D. Boru
CREATE-NET, Via alla Cascata 56/D, Trento, Italy
e-mail: dejene.oljira@create-net.org

D. Kliazovich (✉) · P. Bouvry
University of Luxembourg, 6 rue Coudenhove Kalergi, Luxembourg, Luxembourg
e-mail: dzmitry.kliazovich@uni.lu

P. Bouvry
e-mail: pascal.bouvry@uni.lu

F. Granelli
DISI - University of Trento, Via Sommarive 14, Trento, Italy
e-mail: granelli@disi.unitn.it

A. Y. Zomaya
School of Information Technologies, University of Sydney, Darlington, Australia
e-mail: albert.zomaya@sydeny.edu.eu

Keywords Cloud computing · Data replication · Energy efficiency

1 Introduction

Cloud computing is an emerging technology that attracts ICT service providers offering tremendous opportunities for online distribution of services. It offers computing as a utility, sharing resources of scalable data centers [1,2]. End users can benefit from the convenience of accessing data and services globally, from centrally managed backups, high computational capacity and flexible billing strategies [3]. Cloud computing is also ecologically friendly. It benefits from the efficient utilization of servers, data center power planning, large scale virtualization, and optimized software stacks. Nevertheless, electricity consumed by cloud data centers is still in the order of thousands of megawatts [4]. In 2010, datacenters consumed around 1.1–1.5 % of global electricity consumption and between 1.7 and 2.2 % for U.S [5,6]. Pike Research forecasts data center consumption of almost 140 TW h in 2020 [7].

The growth of Internet services at an unprecedented rate requires the development of novel optimization techniques at all levels to cope with escalation in energy consumption, which in place would reduce operational costs and carbon emissions.

In data centers, there is an over provisioning of computing, storage, power distribution and cooling resources to ensure high levels of reliability [8]. Cooling and power distribution systems consume around 45 and 15 % of the total energy respectively, while leaving roughly 40 % to the IT equipment [9]. These 40 % are shared between computing servers and networking equipment. Depending on

Published online: 10 January 2015

the data center load level, the communication network consumes 30–50 % of the total power used by the IT equipment [10].

There are two main approaches for making data center consume less energy: shutting the components down or scaling down their performance. Both approaches are applicable to computing servers [11,12] and network switches [10,13].

The performance of cloud computing applications, such as gaming, voice and video conferencing, online office, storage, backup, social networking, depends largely on the availability and efficiency of high-performance communication resources [14]. For better reliability and high performance low latency service provisioning, data resources can be brought closer (replicated) to the physical infrastructure, where the cloud applications are running. A large number of replication strategies for data centers have been proposed in the literature [8,15–18]. These strategies optimize system bandwidth and data availability between geographically distributed data centers. However, none of them focuses on energy efficiency and replication techniques inside data centers.

To address this gap, we propose a data replication technique for cloud computing data centers which optimizes energy consumption, network bandwidth and communication delay both between geographically distributed data centers as well as inside each datacenter. Specifically, our contributions can be summarized as follows.

- Modeling of energy consumption characteristics of data center IT infrastructures.
- Development of a data replication approach for joint optimization of energy consumption and bandwidth capacity of data centers.
- Optimization of communication delay to provide quality of user experience for cloud applications.
- Performance evaluation of the developed replication strategy through mathematical modeling and using a packet-level cloud computing simulator, GreenCloud [19].
- Analysis of the tradeoff between performance, serviceability, reliability and energy consumption.

The rest of the paper is organized as follows: Sect. 2 highlights relevant related works on energy efficiency and data replication. In Sect. 3 we develop a mathematical model for energy consumption, bandwidth demand and delay of cloud applications. Section 4 provides evaluation of the model outlining theoretical limits for the proposed replication scenarios. Section 5 presents evaluation results obtained through simulations. Section 6 concludes the paper and provides an outline for the future work on the topic.

2 Related works

2.1 Energy efficiency

At the component level, there are two main alternatives for making data center consume less energy: (a) shutting hardware components down or (b) scaling down hardware performance. Both methods are applicable to computing servers and network switches. When applied to the servers, the former method is commonly referred to as dynamic power management (DPM) [11]. DPM results in most of the energy savings. It is the most efficient if combined with the workload consolidation scheduler—the policy which allows maximizing the number of idle servers that can be put into a sleep mode, as the average load of the system often stays below 30 % in cloud computing systems [11]. The second method corresponds to the dynamic voltage and frequency scaling (DVFS) technology [12]. DVFS exploits the relation between power consumption P, supplied voltage V, and operating frequency f :

$$P = V^2 * f.$$

Reducing voltage or frequency reduces the power consumption. The effect of DVFS is limited, as power reduction applies only to the CPU, while system bus, memory, disks as well as peripheral devices continue consuming at their peak rates.

Similar to computing servers, most of the energy-efficient solutions for communication equipment depend on (a) downgrading the operating frequency (or transmission rate) or (b) powering down the entire device or its hardware components in order to conserve energy. Power-aware networks were first studied by Shang at el. [10]. In 2003, the first work that proposed a power-aware interconnection network utilized dynamic voltage scaling (DVS) links [10]. After that, DVS technology was combined with dynamic network shutdown (DNS) to further optimize energy consumption [13].

Another technology which indirectly affects energy consumption is virtualization. Virtualization is widely used in current systems [20] and allows multiple virtual machines (VMs) to share the same physical server. Server resources can be dynamically provisioned to a VM based on the application requirements. Similar to DPM and DVFS power management, virtualization can be applied in both the computing servers and network switches, however, with different objectives. In networking, virtualization enables implementation of logically different addressing and forwarding mechanisms, and may not necessarily have the goal of energy efficiency [21].

2.2 Data replication

Cloud computing enables the deployment of immense IT services which are built on top of geographically distributed platforms and offered globally. For better reliability and performance, resources can be replicated at redundant locations and using redundant infrastructures. To address exponential increase in data traffic [22] and optimization of energy and bandwidth in datacenter systems, several data replication approaches have been proposed.

Maintaining replicas at multiple sites clearly scales up the performance by reducing remote access delay and mitigating single point of failure. However, several infrastructures, such as storage devices and networking devices, are required to maintain data replicas. On top of that, new replicas need to be synchronized and any changes made at one of the sites need to be reflected at other locations. This involves an underlying communication costs both in terms of the energy and network bandwidth. Data center infrastructures consume significant amounts of energy and remain underutilized [23]. Underutilized resources can be exploited without additional costs. Moreover, the cost of electricity differs at different geographical locations [24] making it another parameter to consider in the process of data replication.

In [15], an energy efficient data replication scheme for datacenter storage is proposed. Underutilized storage servers can be turned off to minimize energy consumption, while keeping one of the replica servers for every data object alive to guarantee the availability. In [8], dynamic data replication in cluster of data grids is proposed. This approach creates a policy maker which is responsible for replica management. It periodically collects information from the cluster heads, which significance is determined with a set of weights selected according to the age of the reading. The policy maker further determines the popularity of a file based on the access frequency. To achieve load balancing, the number of replicas for a file is computed in relationship with the access frequency of all other files in the system. This solution follows a centralized design approach, thus exposing it to a single point of failure.

In [16], the authors suggest replication strategy across multiple data centers to minimize power consumption in the backbone network. This approach is based on linear programming and determines optimal points of replication based on the data center traffic demands and popularity of data objects. Since power consumption of aggregation ports is linearly related to the traffic load, an optimization based on the traffic demand can bring significant power savings. This work focuses on replication strategies between different data centers, but not inside data centers.

Another optimization of data replication across data centers is proposed in [17]. The aim is to minimize data access delay by replicating data closer to data consumers. Optimal location of replicas for each data object is determined by periodically processing a log of recent data accesses. Then, replica site is determined by employing a weighted k-means clustering of user locations and deploying replica closer to the centroid of each cluster. The migration from an old site to a new site is performed if the gain in quality of service of migration (communication cost) is higher than a predefined threshold.

A cost-based data replication in cloud datacenter is proposed in [18]. This approach analyzes data storage failures and data loss probability that are in the direct relationship and builds a reliability model. Then, replica creation time points are determined from data storage reliability function.

The approach presented in this paper is different from all replication approaches discussed above by (a) the scope of data replication which is implemented both within a data center as well as between geographically distributed data centers, and (b) the optimization target, which takes into account system energy consumption, network bandwidth and communication delay to define the employed replication strategy.

3 System model

In this section we present a model of geographically distributed cloud computing system which supports replication of data. The model focuses on the performance of cloud applications, utilization of communication resources and energy efficiency.

3.1 Cloud applications

Most of the cloud applications, such as online office and social networking, rely on tight interaction with databases. Data queries can be fulfilled either locally or from a remote location. To ensure data availability and reduce access delays data replication can be used.

Database replication decisions can be based on the data usage patterns and analysis of data popularity [25]. The popularity is measured as a number of access events in a given period of time. Furthermore, the popularity is not constant over time. Typically, a newly created data has the highest demand. Then, the access rate decays over time. For example, a newly posted YouTube video attracts most of the visitors. However, as the time passes it starts to lose popularity and the audience [26].

Several studies of HTTP requests [27,28] and social networks [29] suggest using a power law distribution which has a long-tailed gradual decay:

$$a(t) = a_0 t^{-k}, \tag{1}$$

where a_0 is a maximum number of access events recorded after content publication, t is current time, and k is a coef-

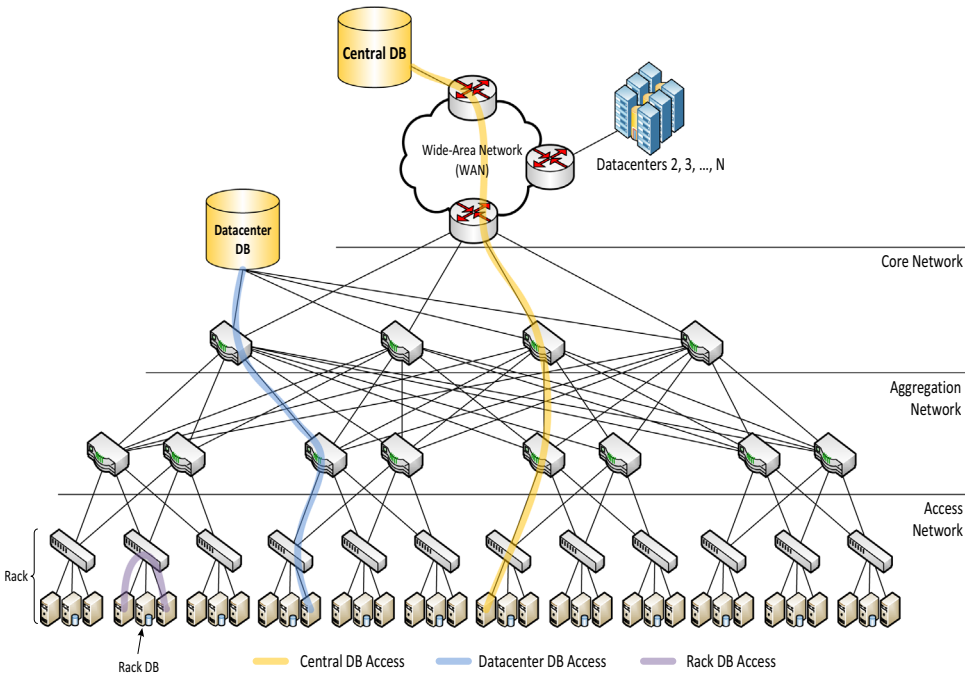


Fig. 1 Cloud computing datacenter

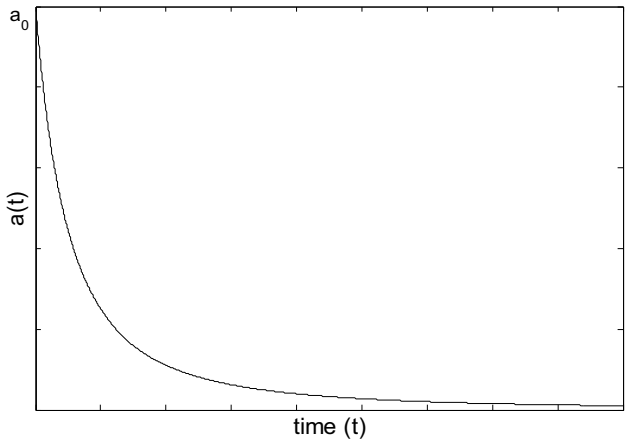


Fig. 2 Data access distribution

ficient typically in the range [30,31], which depends on the type of the content [27]. Figure 2 presents a plot of Eq.(1).

The access rate can be obtained as the first derivative from Eq. (1).

$$r_a(t) = \frac{da}{dt}. \tag{2}$$

Whenever cloud applications access data items, they can modify them with some probability updating the database. Therefore, the update rate can be expressed as a fraction of the access rate.

$$r_u(t) = \rho \cdot r_a(t), \tag{3}$$

where $\rho \epsilon [0, 1]$ controls relation between the access rate $r_a(t)$ and the update rate. For $\rho = 1$, cloud applications modify every data item they access, while for $\rho = 0$ these modifications is never performed.

Figure 3 presents the timeline of a workload execution in data center. It begins with the user request arrival at the datacenter gateway. After being scheduled it is forwarded through the data center network to the selected computing resource for execution. At the server, the workload can request data item if it is needed for its execution. For this, it queries a database and waits for the database reply to arrive. The database querying delay corresponds to the round-trip time and depends on the database location. As soon as the database reply is received, the workload execution is started. At the end of the execution, some workloads will send a modified data item back to the database for the update. As a result, the total delay associated with the workload execution in datacenters can be computed as follows:

$$d_{dc} = d_{req} + 2 \cdot d_{db} + d_{exec} + d_{update}, \tag{4}$$

where d_{req} is a time required for the workload description to arrive at the computing server, d_{db} is a one-way communication delay between the server and the database, d_{exec} is a workload execution time which is defined by the size of the computing work of the workload and computing speed of the server, and d_{update} is the time required to update database.

3.2 Cloud computing system architecture

Large-scale cloud computing systems are composed of geographically distributed across the globe data centers (see Fig. 1). The most widely used data center topology is the three tier fat tree [31], which consists of three layers of net-

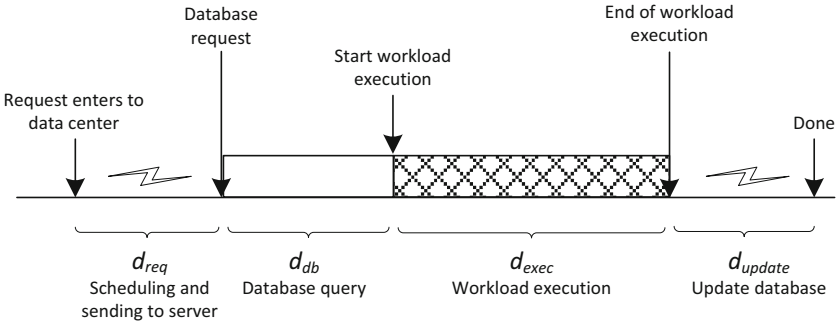


Fig. 3 Workload execution timeline

work switches: core, aggregation and access. The core layer provides packet switching backplane for all the flows going in and out datacenter. The aggregation layer integrates connections and traffic flows from multiple racks. The access layer is where the computing servers, physically attached to the network, are arranged in racks.

Central database (Central DB) is located in the wide-area network and hosts all the data required by the cloud applications. To speed up database access and reduce access latency, each data center hosts a local database, called datacenter database (Datacenter DB), which is used to replicate the most frequently used data items from the Central DB. In addition, each rack hosts at least one server capable of running local rack-level database (Rack DB), which is used for subsequent replication from the Datacenter DB.

All database requests produced by the cloud applications running at the computing servers are first directed to the rack-level database server. Rack DB either replies with the requested data or forwards the request to the Datacenter DB. In a similar fashion, the Datacenter DB either satisfies the request or forwards it up to the Central DB.

When data is queried, the information about requesting server, the rack, and the datacenter is stored. In addition, the statistics showing the number of accesses and updates are maintained for each data item. The access rate (or popularity) is measured as the number of access events per period of time. While accessing data items cloud applications can modify them. These modifications have to be sent back to the database and updated in all the replica sites.

A module called replica manager (RM) is located at the Central DB. It periodically analyzes data access statistics to identify which data items are the most suitable for replication and at which replication sites. The availability of access and update statistics makes it possible to project data center bandwidth usage and energy consumption.

The following subsections present a model of the considered cloud computing system in terms of energy consumption, usage of network bandwidth and communication delays. The objective is to (a) minimize system-level energy consumption, (b) minimize utilization of network bandwidth and (c) minimize communication delays encountered in the data center network.

3.3 Energy consumption of computing servers

The power consumption of a server depends on its CPU utilization. As reported in [32–34], an idle server consumes about two-thirds of its peak power consumption. This is due to the fact that servers must keep memory modules, disks, I/O resources and other peripherals operational even when no computations are performed. Then, the power consumption scales with offered CPU load according to the following equation [34,35]:

$$P_s(l) = P_{fixed} + \frac{(P_{peak} - P_{fixed})}{2}\left(1 + l - e^{-\frac{l}{a}}\right), \quad (5)$$

where P_{fixed} is an idle power consumption, P_{peak} power consumed at the peak load, l is a server load, and a is a utilization level at which the server attains asymptotic, i.e. close to linear power consumption versus the offered load. For most of the CPUs, a ϵ [0.2, 05].

CPU power consumption is proportional to $V^2 f$, where V is voltage and f is an operating frequency [36]. Voltage reduction requires frequency downshift. This implies a cubic relation from f. To account of it, Eq. (5) can be rewritten as follows:

$$P_s(l) = P_{fixed} + \frac{(P_{peak} - P_{fixed})}{2}\left(1 + l^3 - e^{-\frac{l^3}{a}}\right). \quad (6)$$

Eq. (6) forms the basis for DVFS power management which can adjust operating frequency when server is underutilized to conserve operational power consumption [10]. Figure 4 plots server power consumption given in Eqs. (5) and (6).

3.4 Storage requirements of data replication

One of the pitfalls of data replication is the increased usage of storage and networking resources, which can result in higher capital investment and operation costs. To estimate

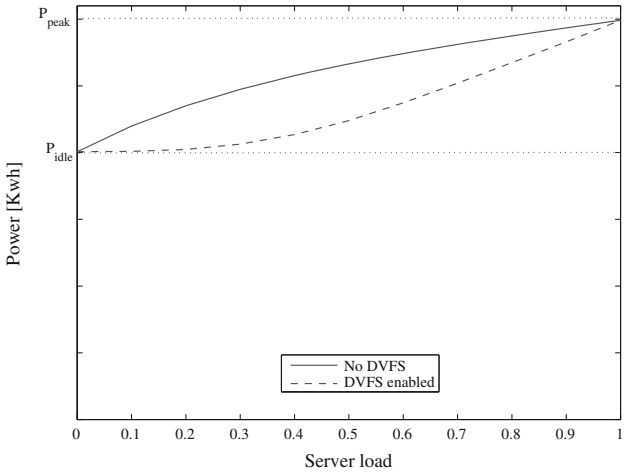


Fig. 4 Server power consumption

the involved storage requirements, assume a uniform storage capacity C $[GB]$ and N data objects stored in the cloud system each of the size S_i $[GB]$. Then, if the replication factor for each data object is r_i, the number of nodes required for maintaining the cloud data is given by:

$$N_s = \left\lfloor \frac{\sum_i^N r_i S_i}{C} \right\rfloor. \tag{7}$$

System storage requirements can be further reduced by applying a number of advanced data replication techniques, such as data deduplication, compression, and erasure-coding [37,38].

3.5 Energy consumption of network switches

Network switches are hardware devices which consist of the port transceivers, line cards, and switch chassis [39]. All these components contribute to the switch energy consumption. Several studies characterize energy consumption of network switches [40–42]. According to [43] and [44], the power consumption of switch chassis and line cards remain constant over time, while the consumption of network ports can scale with the volume of the forwarded traffic as follows:

$$P_{switch} = P_{chassis} + n_c * P_{linecard} + \sum_{r=1}^{R} n_p^r * P_p^r * u_p^r, \tag{8}$$

where $P_{chassis}$ is a power related to switch chassis, $P_{linecard}$ is the power consumed by a single line card, n_c is number of line cards plugged into switch, P_p^r is a power drawn by a port running at rate r, n_p^r is number of ports operating at rate r and $u_p^r \in [0, 1]$ is a port utilization which can be defined as follows:

$$u_p = \frac{1}{T} \int_t^{t+T} \frac{B_p(t)}{C_p} dt = \frac{1}{T * C_p} \int_t^{t+T} B_p(t)\, dt, \tag{9}$$

where $B_p(t)$ is an instantaneous throughput at the port's link at the time t, C_p is the link capacity, and T is a measurement interval.

Each port can be configured to operate at different rates and its power consumption scales exponentially with the increase in the transmission rate [45–47]. Downgrading port rate is especially useful as almost all of the links are never fully utilized for a long duration. Typical link utilization is only 3–5 % [48]. Network packets arrive in bursts [49], while between bursts links remain idle. When idle, there is no need to keep links operating at the peak rates. Instead, link rates can be adapted to satisfy long-term traffic demands using IEEE 802.3az Energy Efficient Ethernet (EEE) [50] standard.

3.6 Bandwidth model

In this section we analyze network capacity of data centers and bandwidth requirements of cloud applications that access database for different replication strategies.

An availability of per-server bandwidth is one of the core requirements affecting design of modern data centers. The most widely used three-tier fat tree topology (see Fig. 1) imposes strict limits on the number of hosted core, aggregation, and access switches as well as the number of servers per rack [30]. For example, a rack switch serving 48 servers each connected with 1 Gb/s link has only two 10 Gb/s links in the uplink. As a result, its uplink bandwidth appears to be oversubscribed by a factor of $48 \cdot 1G \,/\, 20\,G = 2.4$, which also limits the per server available bandwidth to 416 Mb/s. Another bandwidth multiplexing occurs at the aggregation layer. An aggregation switch offers 12 ports to the access layer and is connected to all the core layer switches. For the three tier architecture with 8-way Equal Cost Multipath Routing (ECMP) [30], the oversubscription ratio at the aggregation layer is 1.5. This further reduces the per server bandwidth down to 277 Mbps for fully loaded connections.

According to the model of cloud applications described in Sect. 3.1, all communications inside data center can be broadly categorized to the uplink and downlink types. The uplink flows are those directed from the computing servers towards the core switches. Conversely, the downlink flows are those from the core switches to the computing servers.

In the uplink, network bandwidth is used for propagating database requests and when applications need to update modified data items:

$$B_{ul} = N_{serv} \left(R_a S_{req} + R_u S_{data} \right), \tag{10}$$

where N_{serv} is the number of computing servers, S_{req} is the size of data request, and S_{data} is the size of the updated data item. R_a and R_u are data access and update rates respectively.

In the downlink, the bandwidth is used for sending job descriptions to computing servers for execution, receiving database objects and propagating data item updates between

data replicas:

$$B_{dl} = N_{serv} \cdot R_a \cdot \left(S_{job} + S_{data}\right) + B_{rep}, \tag{11}$$

where S_{job} is the size of the job description, S_{data} is the size of the requested data object in bits, and B_{rep} is the bandwidth required to update all the replicas.

B_{rep} is different on different segments of the downlink. For the wide-area network it corresponds to the update between Central DB and Datacenter DBs

$$B_{rep.wan} = N_{serv} \cdot N_{dc} \cdot R_u \cdot S_{data}, \tag{12}$$

while for the network inside data center it corresponds to the update between Datacenter DBs and Rack DBs

$$B_{rep.dc} = N_{serv} \cdot N_{rack} \cdot R_u \cdot S_{data}, \tag{13}$$

where N_{dc} is the number of Datacenter DBs and N_{rack} is the number of Rack DBs in each data center.

Now, having computed the bandwidth required by running applications and their data base interactions, we can obtain residual bandwidth by subtracting it from the network capacity. It will be different for every tier of the data center network due to bandwidth oversubscription involved.

For a three-tier data center with N_{serv} servers, N_{acc} access, N_{agg} aggregation and N_{core} core switches, the corresponding network capacities at each tier can be obtained as follows:

$$BC_{access} = N_{serv} \cdot C_{access}, \tag{14}$$

$$BC_{agg} = 2 \cdot N_{access} \cdot C_{agg}, \tag{15}$$

$$BC_{core} = N_{agg} \cdot N_{core} \cdot C_{core}, \tag{16}$$

where C_{access}, C_{agg} and C_{core} are the capacities at the access, aggregation and core tiers respectively. Commonly, C_{access} is equal to 1 Gb/s, while C_{agg} and C_{core} correspond to 10 Gb/s links in modern datacenters.

The uplink capacity is always limited due to over subscription at lower layers. Therefore, the residual bandwidth in the downlink R^l_{dl} and in the uplink R^l_{ul} available at each tier of the network can be obtained as follows:

$$\begin{aligned} R^l_{dl} &= BC^l_{dl} - B_{dl}, \\ R^l_{ul} &= BC^{l+1}_{ul} - B_{ul}, \end{aligned} \tag{17}$$

where $l \in (access, agg, core)$ is an index indicating a tier level. The expression $l+1$ refers to the tier located above the tier l.

At any moment of time the residual bandwidth left not in use in the data center can be computed as follows:

$$R_{dl} = \min\left(R^{core}_{dl}, R^{agg}_{dl}, R^{access}_{dl}\right), \tag{18}$$

$$R_{up} = \min\left(R^{core}_{ul}, R^{agg}_{ul}, R^{access}_{ul}\right). \tag{19}$$

3.7 Database access and energy consumption

Having the model of energy consumption for computing servers (Sect. 3.3) and network switches (Sect. 3.4), we can obtain total energy consumption of data center IT equipment as follows:

$$E_{dc} = \sum_{s=1}^{S} E_s + \sum_{k=1}^{K} E^{core}_k + \sum_{l=1}^{L} E^{agg}_l + \sum_{m=1}^{M} E^{access}_m, \tag{20}$$

where E_s is the energy consumed by a computing server s, while E^{core}_k, E^{agg}_l, E^{access}_m are the energy consumptions of k core, l aggregation, and m access switches respectively.

Taking into account the model of cloud applications (Sect. 3.1), the load of individual servers becomes proportional to the workload execution and database query delays and can be obtained as follows:

$$E_s = P_s(l) \cdot (2 \cdot d_{db} + d_{exec}) \cdot R_a \cdot T, \tag{21}$$

where $P_s(l)$ is a power consumed by the server executing a workload obtained according to Eq. (5), d_{db} is the time required to query and receive a data item from the database, d_{exec} is the workload execution time, R_a is an average database access rate, and T is a total time of the workload execution. The delay d_{db} depends on the database location and employed replication strategy. If data query is satisfied from replica databases, d_{db} becomes smaller, as propagation delay inside datacenter is in the order of micro seconds. The delay associated with the database update is not included as it becomes a job of the network to deliver the update after computing server becomes available for executing other tasks.

For network switches, energy consumption depends on the amount of traversing traffic and utilization of network ports (see Eq. (8)). Port utilization and traffic volumes are proportional to the size of job descriptions, data requests, data traffic, and data updates. Equations (10) and (11) allow computing traffic requirements in the uplink and the downlink respectively, while Eqs. (14), (15), and (16) define bandwidth capacity for each segment (access, aggregation, and core) of the network. Based on the aforementioned and by adapting Eq. (8), the energy consumption of the access switches can be computed as follows:

$$E_{access} = \left(P^{access}_{fixed} + \frac{N_{serv}}{N_{access}} \cdot P^{access}_p \cdot \frac{B_{dl}}{BC_{access}} + 2 \cdot P^{agg}_p \cdot \frac{B_{ul}}{BC_{access}} \cdot \frac{N_{access}}{N_{serv}}\right) \cdot T, \tag{22}$$

where P_{fixed} corresponds to the power consumption of the switch chassis and line cards, N_{serv}/N_{access} is the number of servers per rack, P^{access}_p and B_{dl}/BC_{access} are power consumption and port utilization of an access link, while P^{agg}_p

and B_{ul}/BC_{access} are power consumption and port utilization of an aggregation network link.

Similarly, the energy consumption of the aggregation and core switches can be computed as follows:

$$E_{agg} = \left(P_{fixed}^{agg} + 2 \cdot \frac{N_{access}}{N_{agg}} \cdot P_p^{agg} \cdot \frac{B_{dl}}{BC_{agg}} + N_{core} \cdot P_p^{core} \cdot \frac{B_{ul}}{BC_{core}} \right) \cdot T, \quad (23)$$

$$E_{core} = \left(P_{fixed}^{core} + N_{agg} \cdot P_p^{core} \cdot \frac{B_{dl}}{BC_{core}} \right) \cdot T, \quad (24)$$

where $2 \cdot N_{access}/N_{agg}$ is the number of aggregation switch links connected to racks, while P_p^{core} and B_{ul}/BC_{core} are the power consumption and port utilization of a core network link.

4 Model evaluation

In this section we perform evaluation of the system model developed in Sect. 3. The main performance indicators are: data center energy consumption, available network bandwidth and communication delay.

4.1 Setup scenario

Considering three tier data center architecture presented in Fig. 1, we assume a uniform distribution of jobs among the computing servers as well as traffic in the data center network. Both computing servers and network switches implement DVFS [12] and DPM [11] power management techniques. With DVFS, servers can scale power consumption of their CPUs with the offered computing load. Similarly, power consumption of communication ports can be adjusted in network switches based on the load of the forwarded traffic. The DPM technology allows enabling a sleep mode in idle servers and switches.

Table 1 summarizes data center setup parameters. The topology is comprised of 1,024 servers arranged into 32 racks interconnected by 4 core and 8 aggregation switches. The network links interconnecting the core and aggregation switches as well as the aggregation and access switches are 10 Gb/s. The bandwidth of the access links connecting computing servers to the top-of-rack switches is 1 Gb/s. The propagation delay of all these links is set to 3.3 μs. There is only one entry point to the datacenter through a gateway switch, which is connected to all the core layer switches with 100 Gb/s, 50 ms links.

Table 2 presents the power consumption profiles of data center servers and network switches. The server peak energy consumption of 301 W is composed of 130 W allocated for a peak CPU consumption [51] and 171 W consumed by other devices like memory, disks, peripheral slots, mother board, fan, and power supply unit [34]. As the only component which scales with the load is the CPU power, the minimum consumption of an idle server is bound and corresponds to 198 W.

Table 1 Datacenter topology

Parameter	Value
Gateway nodes	1
Core switches	4
Aggregation switches	8
Access (rack) switches	32
Computing servers	1,024
Gateway link	100 Gb/s, 50 ms
Core network link	10 Gb/s, 3.3 μs
Aggregation network link	10 Gb/s, 3.3 μs
Access network link	1 Gb/s, 3.3 μs

Table 2 Power Consumption of Datacenter Hardware

Parameter	Power Consumption [W]		
	Chassis	Line cards	Port
Gateway, core, aggregation switches	1558	1212	27
Access switches	146	–	0.42
Computing server	301		

Energy consumption of network switches is almost constant for different transmission rates as 85–97 % of the power is consumed by switches' chassis and line cards, and only a small portion of 3–15 % is consumed by the port transceivers. The values for power consumption are derived from [52] and [53].

4.2 Bandwidth consumption

The bandwidth consumption is typically low in the uplink. The uplink is used for sending database queries and database update requests. The update requests can be large in size. However, they are sent only at the fraction of the access rate. In the downlink, the required bandwidth is mainly determined by the size of the data items and the data access rate (see Sect. 3.5 for details).

Figure 5 presents the downlink system bandwidth requirements with no database updates. Being proportional to both the size of a data item and the update rate, the bandwidth consumption grows fast and easily overcomes corresponding capacities of the core, aggregation and access segments of the datacenter network requiring replication. Having only

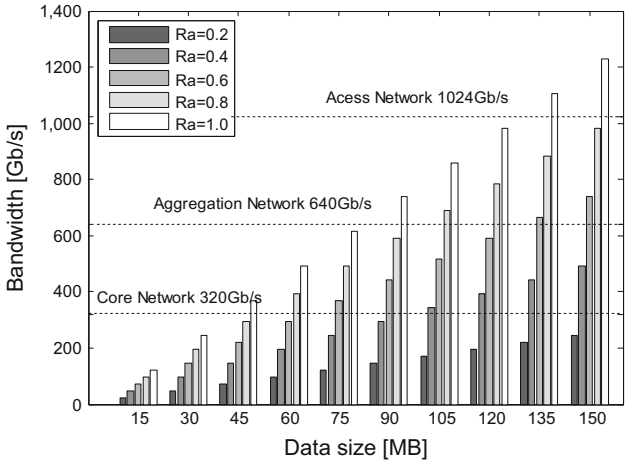


Fig. 5 Downlink bandwidth demand

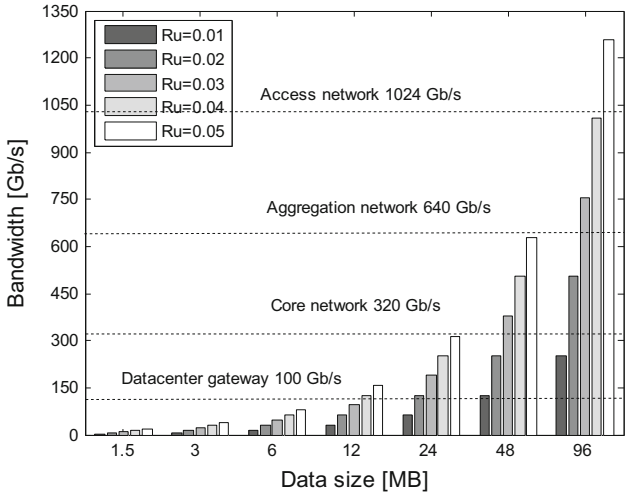


Fig. 6 Bandwidth required for updating replicas

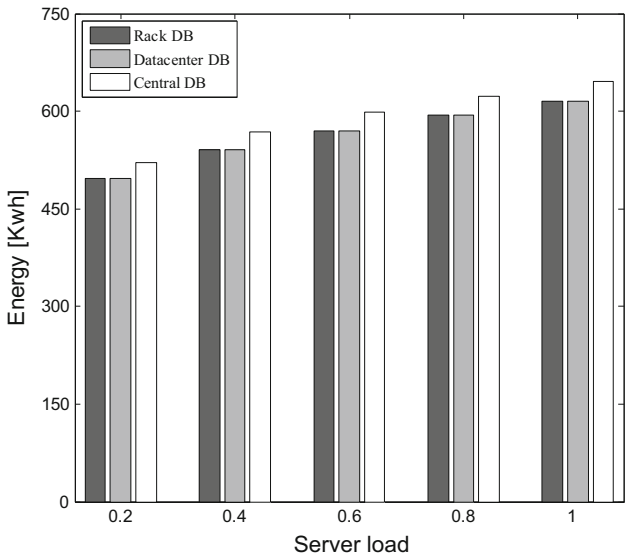


Fig. 7 Energy consumption of computing servers

100 Gb/s at the gateway link would trigger replication even for the small data items of less than 12 MB (or 8 Ethernet packets) for the access rate of 1 Hz requiring data replication from Central DB to the Datacenter DB in order to avoid the bottleneck. The bandwidth provided by the core network of 320 Gb/s will be exceeded with data items larger than 40 MB for the access rate of 1 Hz. Similarly, the bandwidth of the aggregation network of 640 Gb/s will be exceeded after 78 MB and will require additional data replication from Datacenter DB to Rack DBs. Finally, data size larger than 125 MB will cause congestion in the access segment of the network clearly indicating the limits.

Figure 6 shows bandwidth required for propagating replica updates in the downlink from Central DB to Datacenter DB and from Datacenter DB to Rack DBs. The charts reveal that even if replication is used and the data access is localized the burden on network bandwidth becomes considerable when the size of data access is large and frequency of data updates is high. In particular, for the case of Datacenter DB replication high data update rates can exceed the capacity of the gateway link. When updating data on Rack DBs, the bandwidth is consumed at both the core and aggregation layers.

4.3 Energy consumption

According to the model presented in Sect. 3, the energy consumed by IT equipment is composed of the energy consumed by the computing servers as well as core, aggregation, and access switches. Energy consumption of the computing servers is presented in Fig. 7. The servers execute cloud applications which perform a certain amount of computing job and make a single database query for successful completion. The obtained energy consumption increases with the increase in server load. This is due to the fact that energy is consumed during both phases, while doing computing work as well as while waiting for database data to arrive. The minimum querying time corresponds to the round-trip communication delay between the computing server and the database (see Fig. 3 for details). However, in real systems communication delays are larger and are the subject to queuing delays on congested links and protocol-related procedures which often delay transmissions while waiting for previously transmitted data to be acknowledged.

Unlike in the case of computing servers, the energy consumption of network switches is less sensitive to variation in the amount of forwarded traffic. It is mainly due to the fact that only port level power consumption scales with the traffic load under DVFS power saving, while other hardware components, such as switch chassis and line cards, remain always active. Figure 8 reports the obtained energy consumption levels of network equipment. The result suggests that devising

power saving modes that shut down entire hardware components of a switch would allow substantial savings. However, it has to be noted that applying such kind of approaches will affect network connectivity and may result in system performance degradation as datacenter load cannot be accurately predicted.

Figure 9 reports the tradeoff between datacenter energy consumption, which includes the consumption of both the servers and network switches, and the downlink residual bandwidth. For all replication scenarios, the core layer reaches saturation earlier, being the smallest of the datacenter network segments with the capacity of 320 GB/s. Generally, residual bandwidth decreases for all network segments with increase of the load. The only exception is the gateway link, which available bandwidth remains constant for Datacenter DB and Rack DB replication scenarios since data queries are processed at the replica databases and only data updates are routed from Central DB to Datacenter DB. Due to the core network saturation, the maximum size of the data segment in Central DB and Datacenter DB scenarios is equal to 39 MB, while for the Rack DB replication the size of 120 MB can be achieved as the downlink traffic becomes restricted to the access segment of the network. It is important to note that the relative variation of the consumed energy is much smaller than the drop in available network bandwidth. As a result, the benefit of Rack DB replication is two-fold: on one hand network traffic can be restricted to the access network, which has lower nominal power consumption and higher network capacity, while on the other, data access becomes localized improving performance of cloud applications.

5 Replication algorithm

To ensure energy efficiency and performance of cloud applications, we propose a replication algorithm which takes into account power consumption and bandwidth required for data access. Every data object is available at the Central DB, and based on the historical observations of the access frequency data can be replicated to the Datacenter DB and Rack DBs. For every access to meta data information, which includes data object ID, datacenter ID, and rack ID, along with the number of requests, is maintained.

A module called RM located at the Central DB periodically analyzes this meta data information to identify which data objects need to be replicated and where. RM computes access and update rates in previous intervals and makes an estimate for their future values. With these parameters, it becomes possible to compute energy consumption and bandwidth demand in the upcoming intervals using models presented in Sects. 3.5 and 3.6. In addition, congestion levels in the datacenter network are monitored to determine the most suitable candidates for replication.

Replication Algorithm

Notations:

- E_{cen} , $E_{dc,}$ E_{ra} ; B_{cen} , $B_{dc,}$ B_{ra} : Energy & bandwidth consumption in Central, Datacenter, and Rack replication cases respectively.
- $L_{u,j}$, $L_{th,j,}$: Gateway link utilization and threshold value at datacenter j.
- R_a^i, R_{th}^i: Average access rate and threshold value of data object O_i.
- dc_j, ra_k^j: Datacenter j and rack k in datacenter j respectively.
- $O_i \rightarrow dc_j$: Replication of data ata datacenter j
- $O_i \rightarrow ra_k^j$: Replication of data object i at rack k in the datacenter j.

1. INPUTS:
   Parameters to compute energy & bandwidth as in Sec III-E and Sec III-F
2. OUTPUT:
   $O_i \rightarrow R_{j||k}$
3. For all O_i
4. If $R_a^i \geq R_{th}^i$
5. If $O_i \in dc_j$
6. Compute $E_{dc_j}, E_{ra_k^j}$ && $B_{dc_j}, B_{ra_k^j}$ $\forall j, \forall k$
7. If $E_{dc_j} > E_{ra_k^i}$, $B_{dc_j} > B_{ra_k^i}$
8. If $\left(E_{ra_k^j}, B_{ra_k^j}\right)_{min}$ $\forall j, \forall k$
9. $O_i \rightarrow ra_k^j$
10. Else If $L_{u,j} \geq L_{th,j}$&&$R_{a,k}^i \geq R_{a,r\neq k}^i \forall i, \forall j, \forall k$
11. $O_i \rightarrow ra_k^j$
12. End If
13. Else
14. Compute E_{cen}, B_{cen} && E_{dc_j}, B_{dc_j} $\forall j$
15. IF $E_{cen} > E_{dc_j}$ && $B_{cen} > E_{dc_j}$
    IF (E_{dc_j}, B_{dc_j})$_{min}$
16. $O_i \rightarrow dc_j$
17. End If
18. End If
19. End For

6 Simulation results

For performance evaluation purposes we developed the GreenCloud simulator [19] and extended it with the required data replication functionality. GreenCloud is a cloud computing simulator which captures data center communication processes at the packet level. It is based on the widely known Ns2 platform [54] for TCP/IP network simulation. In addition, GreenCloud offers fine-grained modeling of the energy consumed by computing hardware, network switches and communication links. To achieve consistency with modeling results presented in previous sections, the simulation scenario

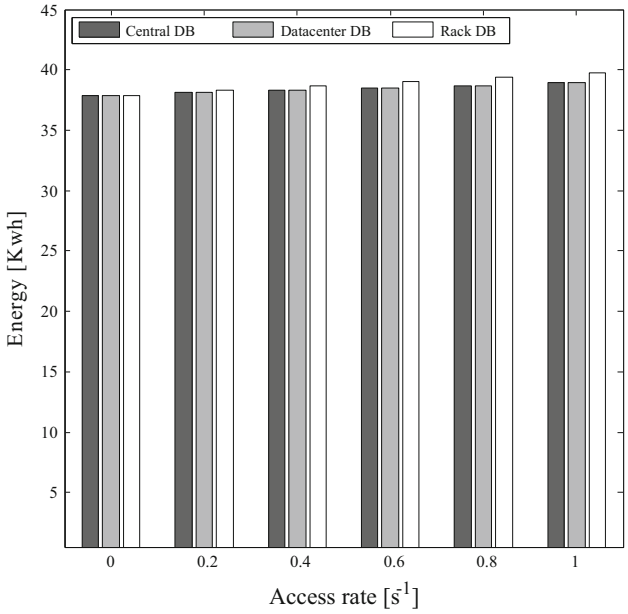


Fig. 8 Energy consumption of network switches

was selected accordingly using topology setup presented in Table 1 and energy consumption profiles from Table 2.

The workload generation events are exponentially distributed in time to mimic typical process of user arrival. As soon as a scheduling decision is taken for a newly arrived workload, it is sent over the data center network to the selected server for execution. The workload execution and data querying follow the timeline diagram presented in Fig. 3. The size of a workload description and database queries is limited to 1,500 bytes and fits into a single Ethernet packet. The sizes of data items, data access and update rates, as well as the replication threshold vary in different simulation runs. The duration of each simulation is 60 minutes. DNS power saving scheme is enabled both for servers and switches in all simulation scenarios.

The main metrics selected for performance evaluation are (a) energy consumption at the component and system levels, (b) network bandwidth and (c) communication delay.

The following subsections report the effect of data size and the update rate on energy consumption, network bandwidth and access delay characteristics of the system.

6.1 Effect of data size

Figure 10 presents the measurements of energy consumption of computing servers for data item size varied from 10 to 40 MB. Each server accesses one data item every 300 ms and makes no updates of the database. Two trends can be observed in the obtained results. The first trend is that energy consumption increases with the increase in the data size. The second is that energy consumption decreases as data become available closer to the computing server locations. The reason is that communication delay is included

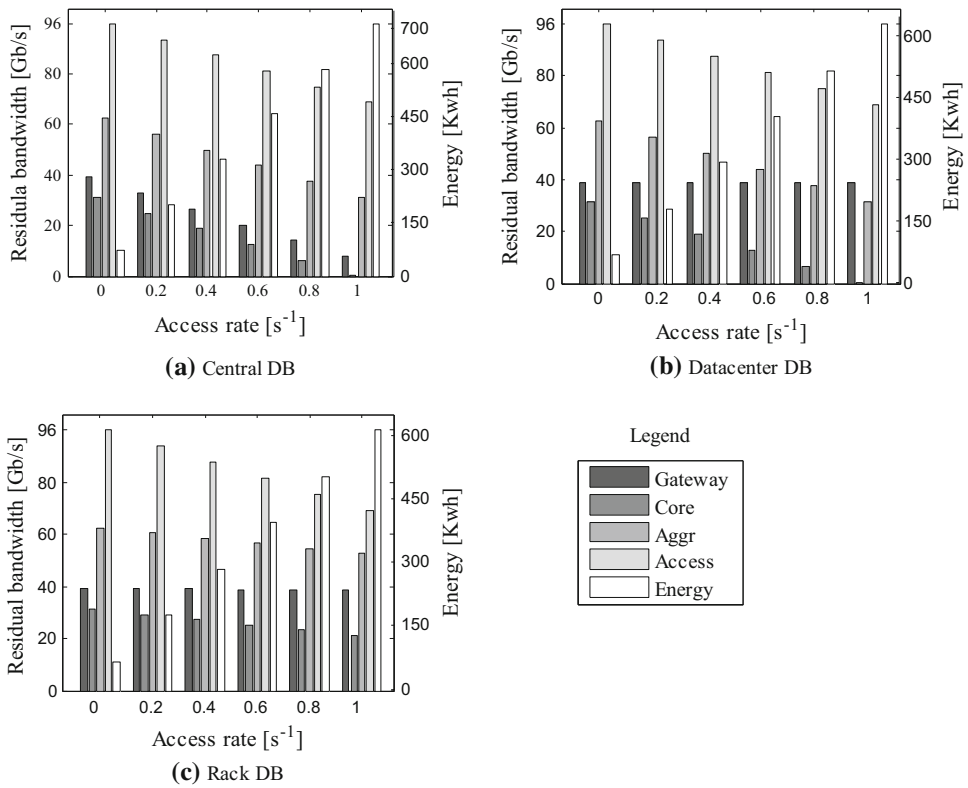


Fig. 9 Energy and residual bandwidth for **a** Central DB, **b** Datacenter DB, and **c** Rack DB replication scenarios

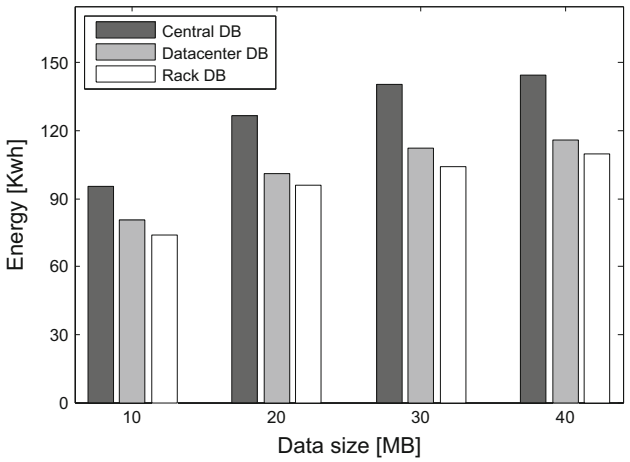


Fig. 10 Energy consumption of servers

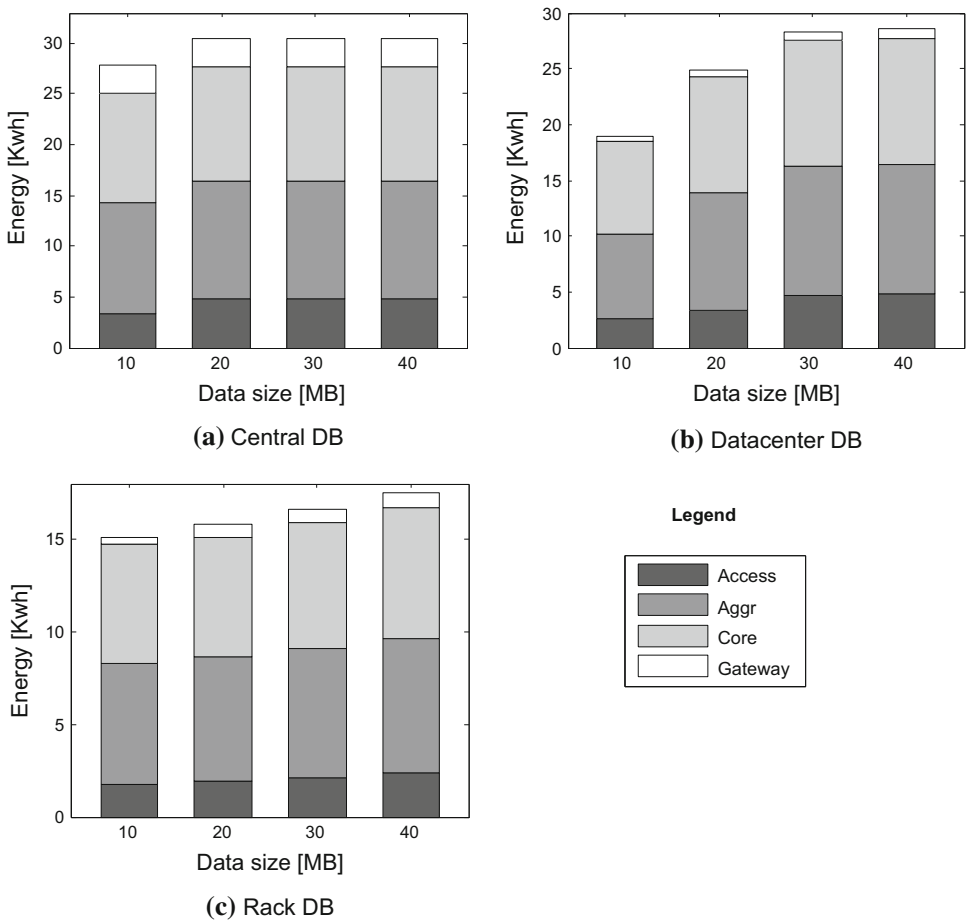
into the execution time of the cloud application (see Fig. 3), which prevents servers to enter into the sleep mode. These delays become larger with the increase in data item size, but can be reduced by shortening round-trip times to the database.

Energy consumption of network switches scales similarly with the increase in data size (see Fig. 11). The consumption with no replication (Central DB) is higher than for other replication scenarios as all the layers become actively involved into traffic forwarding. For the rack replication (Rack DB), the consumption can be reduced considerably, as data traffic mostly remains constrained in the access part of the network.

Figure 12 shows the downlink bandwidth requirement for different data sizes. Bandwidth demand remains high for large data sizes for all replication scenarios. The bandwidth slightly varies with the time, which is the effect of the exponential arrival of the incoming tasks. It should be noted that for a Central DB (Fig. 12a) and Datacenter DB (Fig. 12b) replication scenarios, the reported bandwidth is consumed at all the segments (gateway, core, aggregation and access) of the network, while for the Rack DB replication (Fig. 12c), the access is localized and the reported bandwidth is consumed only in the access part of the network.

Figure 13 reports data access delays measured as an average time elapsed from the moment of sending data request and having the requested data arrived. As expected, access delay becomes smaller for replicas located closer to servers and for all the replication scenarios an increase in the size of data objects increases data access delay.

Fig. 11 Energy consumption of network switches



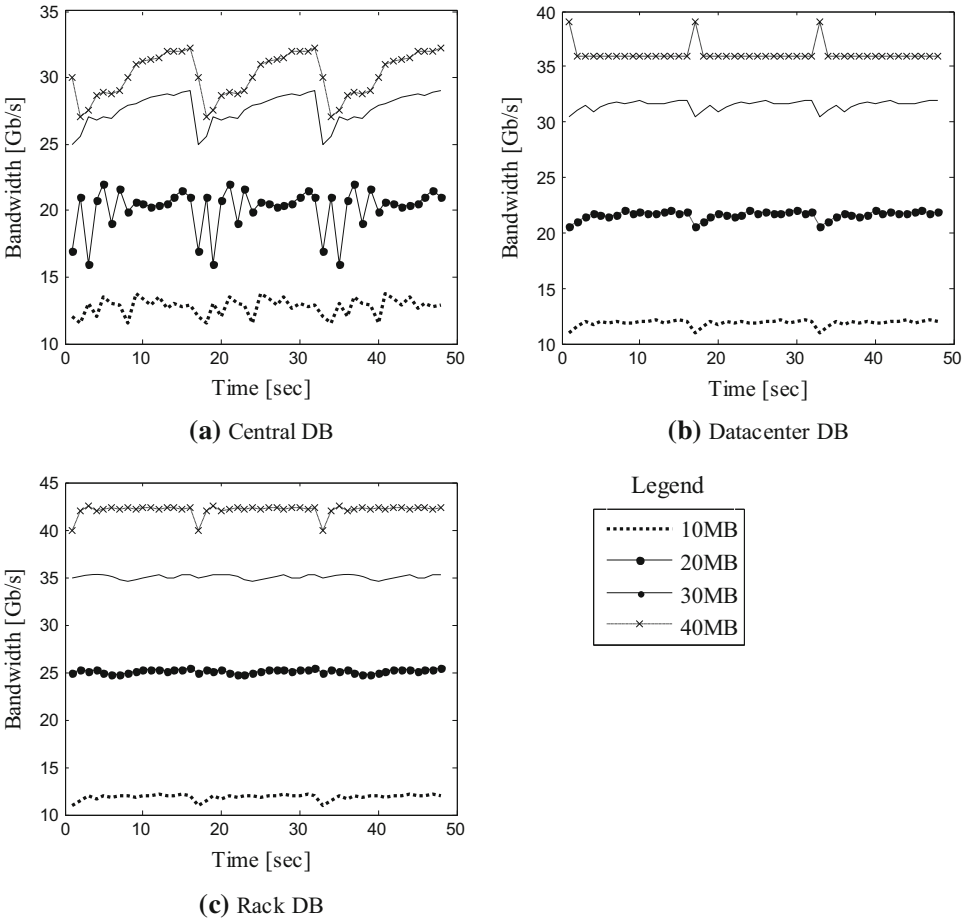


Fig. 12 Bandwidth consumption for **a** Central DB, **b** Datacenter DB, and **c** Rack DB replication scenarios

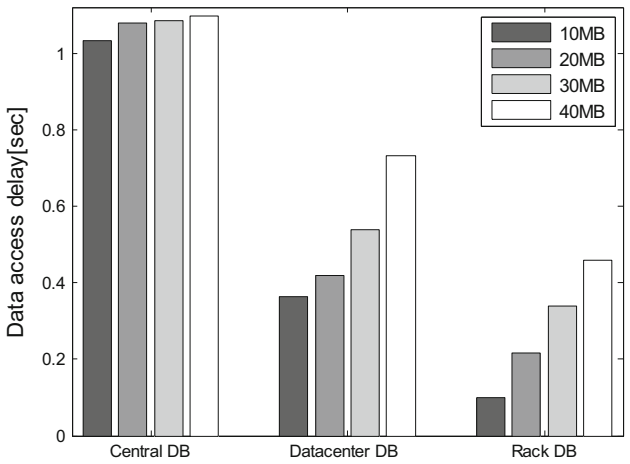


Fig. 13 Data access delay

The simulation results presented above show that for cloud applications, that perform database updates rarely, replicating data closer to the computing servers always helps to save energy, conserve bandwidth and minimize communication delays speeding up execution.

6.2 Effect of data update rate

To better understand the impact of database updates on energy consumption of the system, we kept the size of the data item of 6 MB and access rate of 0.3 Hz fixed while varying the number of updates requested by the cloud applications in the interval $[0,\ R_a]$. For the update rate equal to the access rate, cloud applications modify every accessed data item and send updates to the database. As reported in Fig. 14, update rate variations do not affect energy consumption of computing servers as the role of the servers is just to send modified data item to the Central DB at the end of the workload execution.

Figure 15 presents energy consumption of network switches for fixed data size of 6 MB and access rate of 0.3 Hz and different update rates. As expected, energy consumption grows with the increase in the update rate due to longer awake periods. Switches at all layers are involved into forwarding database update traffic. In the uplink, they forward replica updates sent from the servers to the Central DB. While in the downlink, database updates from Central DB to Datacenter DB and from Datacenter DB to Rack DBs are propagated.

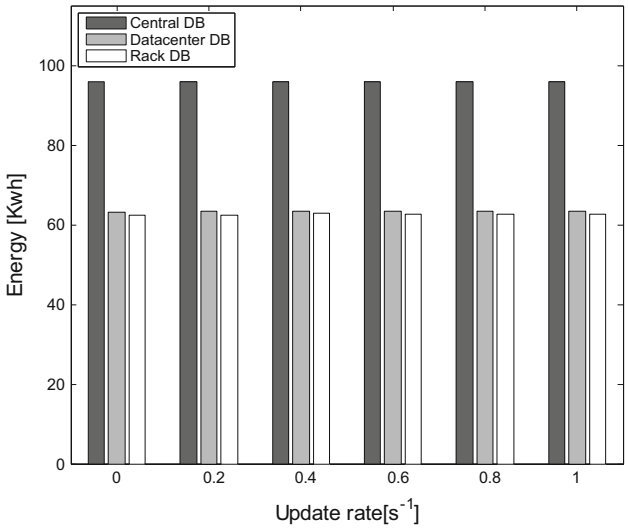


Fig. 14 Energy consumption of servers

In the case of Datacenter DB replication (see Fig. 15b), only the gateway and core switches are involved into update traffic forwarding. While in the case of Rack DB replication (see Fig. 15c), both core and aggregation networks carry database updates for 32 Rack DBs. The access switches serve both kinds of traffic, for data and for database updates, which justifies their higher energy consumption.

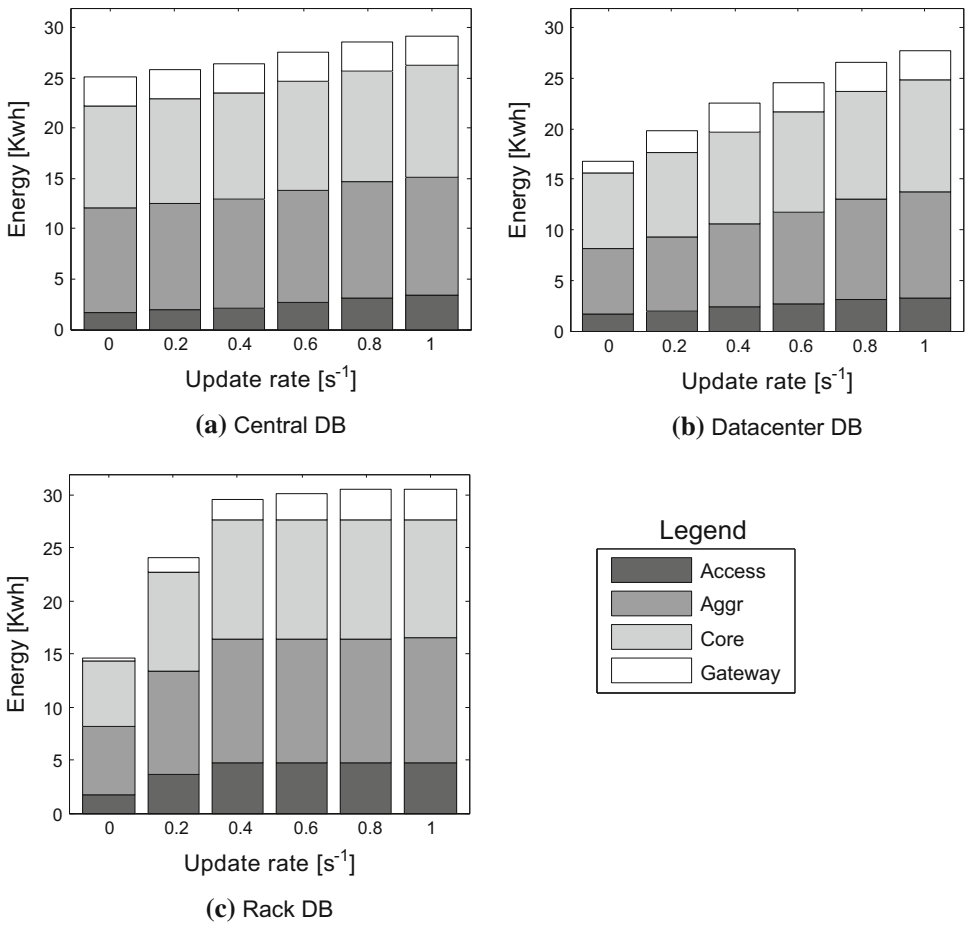
The effect of data update on network bandwidth is shown in Fig. 16. It reports the downlink bandwidth consumption at the core and aggregation layers for the scenario when each accessed data item needs to be modified and updated.

When data is accessed from either Central DB or Datacenter DB, replica update traffic remains at a very low level. However, as soon as Rack DB replication is enabled, the bandwidth usage increases to over 35 Gb/s, as data updates begin to propagate from Datacenter DB to multiple Rack DBs. This underlines a tradeoff between energy and bandwidth consumptions when replicating data closer to the computing servers.

7 Performance comarsion with existing systems

There are many different data replication strategies widely employed in the production environments. Their performance can be measured using a set of metrics. The following are the most commonly used metrics for evaluation of data replication strategies.

Fig. 15 Energy consumption of network switches



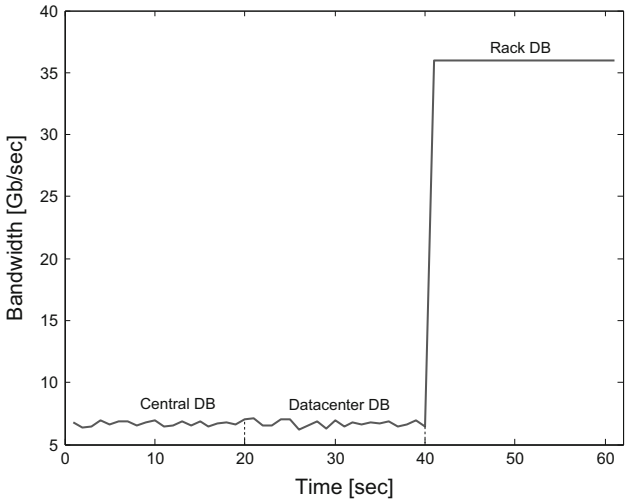


Fig. 16 Downlink bandwidth with high update rate

7.1 Availablity

One of the main goals of replication is to assure data availability. An unexpected failure in storage infrastructure or datacenter blackout could cause unavailability. To resist these effects, copies of data objects are maintained on redundant infrastructures and in several geographically distributed datacenters. Therefore, the availability is usually measured by a probability of failures in data center components or storage infrastructure.

In the proposed replication approach, we assume that every data object is permanently stored at the Central DB and in addition, depending on the access pattern, it is replicated in Datacenter DB and Rack DBs. Any failures in Datacenter DBs can be recovered from Central DB and vice versa. Moreover, unlike in several other methods [55,56], the proposed approach implements a dynamic replication to only maintain optimal number of replicas to ensure both availability and the QoS of cloud applications.

7.2 Response time

Another important reason for data replication is in the reduced data access response time for cloud applications. Bringing and maintaining the data closer to the servers where applications are executed significantly decrease access time for this data and greatly improves overall system performance. However, on the other side, the number and location of replicas should be selected carefully as excessive replication may increase the associated costs and traffic load in the data center network required for replica updates.

The proposed replication approach takes into account the tradeoff between data size, data access and update rates, available network bandwidth and properties of the implemented data center topology to make optimal decisions. First, data objects are replicated in Rack DBs closer to computing nodes and hence response time is reduced. Second, data objects that are frequently accessed are replicated which reduced total number of replicas. Maintaining optimal number of replicated data objects minimizes network load required to keep all replicas up to date and network response time. The obtained simulation results (see Fig. 13) indicate that, the proposed replication always keep the response time within the boundaries required by environment [57,58].

7.3 Datacenter congestion

Network congestion can cause significant degradation of data center performance and one of the most sensitive to congestion points is data center gateway which is a bottleneck handling both the incoming and outgoing traffic. To overcome this problem the proposed replication technique monitors datacenter gateway traffic load to induce replication of data objects with higher access frequency. Data objects are replicated either at the Datacenter DB and/or Rack DBs.

The residual bandwidth at datacenter gateway can be used to indicate congestion. The simulation results (see Fig. 9) confirm that replication is an effective tool to control traffic load at the data center gateway.

8 Conclusions and future work

This paper reviews the topic of data replication in geographically distributed cloud computing data centers and proposes a novel replication solution which in addition to traditional performance metrics, such as availability of network bandwidth, optimizes energy efficiency of the system. In addition, optimization of communication delays leads to improvements in quality of user experience of cloud applications. It extends a preliminary version of this work which has been published in [59].

The evaluation of the proposed replication solution is based on the developed mathematical model and simulations using GreenCloud, the simulator focusing on energy efficiency and communication processes in cloud computing data centers [19]. The obtained results confirm that replicating data closer to data consumers, i.e., cloud applications, can reduce energy consumption, bandwidth usage and communication delays substantially.

Future work on the topic will be focused developing a testbed implementation of the proposed solution.

Acknowledgments The authors would like to acknowledge the funding from National Research Fund, Luxembourg in the framework of ECO-CLOUD project (C12/IS/3977641).

References

1. Buyya, R., Yeo, C.S., Venugopal, S.: Market-oriented cloud computing: vision, hype, and reality for delivering IT services as computing utilities. In: IEEE International Conference on High Performance Computing and Communications (HPCC), pp. 5–13. Dalian, China, (2008).
2. Hussain, H., et al.: A survey on resource allocation in high performance distributed computing systems. Parallel Comput. **39**(11), 709–736 (2013)
3. Hayes, B.: Cloud computing. Mag. Commun. ACM **51**(7), 9–11 (2008)
4. Katz, R.H.: Tech titans building boom. IEEE Spectr. **46**(2), 40–54 (2009)
5. Koomey, J.: Worldwide electricity used in data centers. Environ. Res. Lett. **3**(3), 034008 (2008)
6. Koomey, J.G.: Growth in Data center electricity uses 2005 to 2010. Analytics Press, Oakland (2011)
7. "Cloud Computing Energy Efficiency, Strategic and Tactical Assessment of Energy Savings and Carbon Emissions Reduction Opportunities for Data Centers Utilizing SaaS, IaaS, and PaaS", Pike Research, (2010).
8. Chang, R.S., Chang, H.P., Wang, Y.T.: A dynamic weighted data replication strategy in data grids. In: IEEE/ACS International Conference on Computer Systems and Applications, pp. 414–421. (2008).
9. Brown, R., et al.: Report to congress on server and data center energy efficiency: public law 109–431. Lawrence Berkeley National Laboratory, Berkeley (2008)
10. Shang, L., Peh, L.S., Jha, N.K.: Dynamic voltage scaling with links for power optimization of interconnection networks. In: Ninth International Symposium on High-Performance Computer Architecture (HPCA), pp. 91–102. (2003).
11. Wang, Shengquan, Liu, Jun, Chen, Jian-Jia, Liu, Xue: Powersleep: a smart power-saving scheme with sleep for servers under response time constraint. IEEE J. Emerg. Sel. Top. Circuits Syst. **1**(3), 289–298 (2011)
12. Horvath, T., Abdelzaher, T., Skadron, K., Liu, X.: Dynamic voltage scaling in multitier web servers with end-to-end delay control. IEEE Trans. Comput. **56**(4), 444–458 (2007)
13. Kim, J.S., Taylor, M.B., Miller, J., Wentzlaff, D.: Energy characterization of a tiled architecture processor with on-chip networks. In: International Symposium on Low Power Electronics and Design, pp. 424–427. (2003).
14. Kliazovich, D., Pecero, J.E., Tchernykh, A., Bouvry, P., Khan, S.U., Zomaya, A.Y.: CA-DAG: communication-aware directed acyclic graphs for modeling cloud computing applications. In: IEEE International Conference on Cloud Computing (CLOUD), Santa Clara, USA, (2013).
15. Lin, B., Li, S., Liao, X., Wu, Q., Yang, S.: eStor: energy efficient and resilient data center storage. In: 2011 International Conference on Cloud and Service Computing (CSC), pp. 366–371. (2011).
16. Dong, X., El-Gorashi, T., Elmirghani, J.M.H.: Green IP over WDM networks with data centers. J. Lightwave Technol. **29**(12), 1861–1880 (2011)
17. Ping, F., Li, X., McConnell, C., Vabbalareddy, R., Hwang, J.H.: Towards optimal data replication across data centers. In: International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 66–71. (2011).
18. Li, W., Yang, Y., Yuan, D.: A novel cost-effective dynamic data replication strategy for reliability in cloud data centres. In: International Conference on Dependable, Autonomic and Secure Computing (DASC), pp. 496–502. (2011).
19. Kliazovich, D., Bouvry, P., Khan, S.U.: GreenCloud: a packet-level simulator of energy-aware cloud computing data centers. J. supercomput. **62**(3), 1263–1283 (2012)
20. Chernicoff, D.: The shortcut guide to data center energy efficiency. Realtimepublishers, New York (2009)
21. Sherwood, R., Gibby, G., Yapy, K.-K., Appenzellery, G., Casado, M., McKeowny, N., Parulkary, G.: Flowvisor: a network virtualization layer". Technical report. Deutsche Telekom Inc. R&D Lab, Stanford University, NiciraNetworks (2009)
22. Cisco, Cisco Visual Networking Index: Forecast and Methodology, 2011–2016, May, White paper http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf (2012)
23. Rasmussen, N.: Determining total cost of ownership for data center and network room infrastructure. APC White paper No.6, March (2005).
24. Electricity Information 2012, International Energy Agency, (2012).
25. Madi M.K., Hassan, S.: Dynamic replication algorithm in data grid: survey. International Conference on Network Applications, Protocols, and Services, pp. 1–7. (2008).
26. Cheng, X., Dale, C., Liu, J.: Statistics and social network of YouTube videos. In: 16th International Workshop on Quality of Service, pp. 229–238. (2008).
27. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. **51**, 661–703 (2009)
28. Adamic, L.A., Huberman, B.A.: Zipf, power-laws, and pareto - a ranking tutorial. Glottometrics **3**, 143–150 (2012)
29. Asur, S., Huberman, B.A., Szabo, G., Wang, C.: Trends in social media: persistence and decay. In: 5th International AAAI Conference on Weblogs and Social Media (2011).
30. Cisco Systems, Cisco Data Center Infrastructure 2.5 Design Guide Inc, Nov (2011).
31. Al-Fares, M., Loukissas, A., Vahdat, A.: A scalable, commodity data center network architecture. ACM SIGCOMM **38**, 63–74 (2008)
32. Bellosa, F.: The benefits of event driven energy accounting in power-sensitive systems. In: ACM SIGOPS European Workshop: beyond the PC: new challenges for the operating system, pp. 37–42. (2000).
33. Pelley, S., Meisner, D., Wenisch, T.F., VanGilder, J.W.: Understanding and abstracting total data center power. In: Workshop on Energy Efficient Design (WEED), (2009).
34. Fan, X., Weber, W.-D., Barroso, L.A.: Power provisioning for a warehouse-sized computer. In: ACM International Symposium on Computer Architecture, pp. 13–23. San Diego, (2007).
35. Server Power and Performance characteristics, available on http://www.spec.org/power_ssj2008/
36. Chen, Y., Das, A., Qin, W.: Managing server energy and operational costs in hosting centers. In: ACM SIGMETRICS international conference on Measurement and modeling of computer systems, pp. 303–314. (2005).
37. Upadhyay, A., Balihalli, P.R., Ivaturi, S., Rao, S.: Deduplication and compression techniques in cloud design. Systems Conference (SysCon), 2012 IEEE International, vol., no., pp. 1–6. 19–22 March (2012).
38. Rashmi, K.V., Shah, N.B., Gu, D., Kuang, H., Borthakur D., Ramchandran K.: A Solution to the network challenges of data recovery in erasure-coded distributed storage systems: a study on the facebook warehouse cluster. In: Proceedings of 5^{th} USENIX conf. on Hot Topics in Storage and File Systems, Berkely, CA, USA (2013).
39. Ganesh, A., Katz, R.H.: Greening the switch. In: Conference on Power aware computing and systems, pp. 7. (2008).

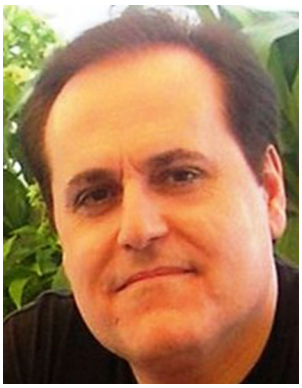
40. Ricciardi, S., Careglio, D., Fiore, U., Palmieri, F., Santos-Boada, G., Sole-Pareta, J.: Analyzing local strategies for energy-efficient networking, In: IFIP International Networking Workshops (SUNSET), pp. 291–300. Springer, Berlin (2011).
41. Sivaraman, V., Vishwanath, A., Zhao, Z., Russell, C.: Profiling per-packet and per-byte energy consumption in the NetFPGA Gigabit router. In: IEEE Conference on Computer Communications (INFOCOM) Workshops, pp. 331–336. (2011).
42. Kharitonov, D.: Time-domain approach to energy efficiency: high-performance network element design. IEEE GLOBECOM Workshops, pp. 1–5. (2009).
43. Mahadevan, P., Sharma, P., Banerjee, S., Ranganathan, P.: A power benchmarking framework for network devices. In: 8th International IFIP-TC 6 Networking Conference, pp. 795–808. Aachen, Germany (2009).
44. Reviriego, P., Sivaraman, V., Zhao, Z., Maestro, J.A., Vishwanath, A., Sanchez-Macian, A., Russell, C.: An energy consumption model for energy efficient ethernet switches. In: International Conference on High Performance Computing and Simulation (HPCS), pp. 98–104. (2012).
45. Sohan, R., Rice, A., Moore, A.W., Mansley, K.: Characterizing 10 Gbps network interface energy consumption. In: IEEE 35th Conference on Local Computer Networks (LCN), pp. 268–271. (2010).
46. Agarwal, Y., Hodges, S., Chandra, R., Scott, J., Bahl, P., Gupta, R.: Somniloquy: augmenting network interfaces to reduce PC energy usage. In: 6th USENIX Symposium on Networked Systems Design and Implementation, pp. 365–380. Berkeley, USENIX Association (2009).
47. Christensen, K., Nordman, B.: Improving the Energy Efficiency of Networks: A Focus on Ethernet and End Devices. Cisco Corporation, San Jose (2006)
48. Odlyzko, A.: Data networks are lightly utilized, and will stay that way", Technical Report Center for Discrete Mathematics & Theoretical Computer Science ACM, (1999).
49. Leland, W., Taqqu, M., Willinger, W., Wilson, D.: On the selfsimilar nature of ethernet traffic. IEEE Trans. Netw. **2**(1), 1–15 (1994)
50. Reviriego, P., Christensen, K., Rabanillo, J., Maestro, J.A.: An initial evaluation of energy efficient ethernet. IEEE Commun. Lett. **5**(15), 578–580 (2011)
51. Intel Inc., Intel Xeon Processor 5000 Sequence, available at: http://www.intel.com/p/en_US/products/server/processor/ xeon5000 (2010)
52. Farrington, N., Rubow, E., Vahdat, A.: Data center switch architecture in the age of merchant silicon. In: 17th IEEE symposium on high performance interconnects (HOTI '09), pp. 93–102. (2009).
53. Mahadevan, P., Sharma, P., Banerjee, S., Ranganathan, P.: Energy aware network operations. In: IEEE INFOCOM workshops, pp. 1–6. (2009).
54. The Network Simulator Ns2 http://www.isi.edu/nsnam/ns/
55. Ghemawat, S., Gobioff, H., Leung, S.T.: The Google file system. ACM SIGOPS Oper. Syst. Rev. **37**(5), 29–43 (2003)
56. Shvachko, K., Hairong, K., Radia, S., Chansler, R.: The Hadoop distributed file system. In: Proceedings the 26th Symposium on Mass Storage Systems and Technologies, Incline Village, NV, USA, May 3–7, pp. 1–10. (2010).
57. Wang, Q., Kanemasa, Y., Li, J., Jayasinghe, D., Kawaba, M.: Response time reliability in cloud environments: an empirical study of n-tier applications at high resource utilization. In: Reliable Distributed Systems (SRDS), 2012 IEEE 31st Symposium on, vol., no., pp. 378,383, 8–11 Oct. (2012).
58. You, Xindong: Zhou, Li, Huang, Jie, Zhang, Jinli, Jiang, Congfeng, Wan, Jian: An energy-effective adaptive replication strategy in cloud storage system. Int. J. Appl. Math. Inf. Sci. **7**(6), 2409–2419 (2013)
59. Boru, D., Kliazovich, D., Granelli, F., Bouvry, P., Zomaya, A.Y.: Energy-efficient data replication in cloud computing datacenters. In: IEEE Globecom 2013 International Workshop on Cloud Computing Systems, Networks, and Applications (GC13 WS - CCSNA). Atlanta, GA, USA (2013).

Dejene Boru received his M.Sc from University of Trento (Italy), in 2013 and B.Sc in Electrical and Computer Engineering from Addis Ababa University (Ethiopia) in 2008. He was a Graduate Assistant at Dire Dawa University (Ethiopia) from Nov. 2008 to Aug. 2010. He is working as a researcher at CREATE-NET(Italy) since Jan. 2013. His research interests are energy efficient cloud computing, wireless indoor localizations and applications, and software defined networking.

Dzmitry Kliazovich is a Research Fellow at the Faculty of Science, Technology, and Communication of the University of Luxembourg. He holds an award-winning Ph.D. in Information and Telecommunication Technologies from the University of Trento (Italy). He was a visiting scholar at the Computer Science Department of the University of California at Los Angeles (UCLA) in 2005 and at the School of Information Technologies of the University of Sydney in 2014. Dr. Kliazovich is a holder of a large number of scientific awards, mainly from the IEEE Communications Society, European Research Consortium for Informatics and Mathematics (ERCIM) and Italian Ministry of Education. His work on energy-efficient scheduling in cloud computing environments received Best Paper Award at the IEEE/ACM International Conference on Green Computing and Communications (GreenCom) in 2010. He served as general and TPC chair in a number of high-ranked international conferences, including the IEEE International Conference on Cloud Networking (CLOUDNET 2014). Dr. Kliazovich is the author of more than 90 research papers and Editorial Board Member of the IEEE Communications Surveys and Tutorials. He is a coordinator and principle investigator of Energy-Efficient Cloud Computing and Communications initiative funded by the National Research Fund of Luxembourg. His main research activities are in the field of energy efficient communications, cloud computing, and next-generation networking.

Fabrizio Granelli is IEEE ComSoc Distinguished Lecturer for the period 2012–15, and Associate Professor at the Dept. of Information Engineering and Computer Science (DISI) of the University of Trento (Italy). From 2008, he is deputy head of the academic council in Information Engineering. He received the «Laurea» (M.Sc.) degree in Electronic Engineering and the Ph.D. in Telecommunications Engineering from the University of Genoa, Italy, in 1997 and 2001, respectively. In August 2004, August 2010 and April 2013, he was visiting professor at the State University of Campinas (Brasil). He is author or co-author of more than 140 papers with topics related to networking, with focus on performance modeling, wireless communications and networks, cognitive radios and networks, green networking and smart grid communications. Dr. Granelli was guest-editor of ACM Journal on Mobile Networks and Applications, ACM Transactions on Modeling and Computer Simulation, and Hindawi Journal of Computer Systems, Networks and Communications. He is Founder and General Vice-Chair of the First International Conference on Wireless Internet (WICON'05) and General Chair of the 11th and 15th IEEE Workshop on Computer-Aided Modeling, Analysis, and Design of Communication Links and Networks (CAMAD'06 and IEEE CAMAD'10). He is TPC Co-Chair of IEEE GLOBECOM Symposium on "Communications QoS, Reliability and Performance Modeling" in the years 2007, 2008, 2009 and 2012. He was officer (Secretary 2005–2006, Vice-Chair 2007–2008, Chair 2009–2010) of the IEEE ComSoc Technical Committee on Communication Systems Integration and Modeling (CSIM), and Associate Editor of IEEE Communications Letters (2007–2011).

Pascal Bouvry is Professor at the University of Luxembourg, elected head of the Interdisciplinary Laboratory for Intelligent and Adaptive Systems (http://ilias.uni.lu), director of studies for the certificate "Smart ICT for business innovation" and Faculty member of the SnT (http://snt.uni.lu). His group (http://pcog.uni.lu) is also managing the HPC facility of the University of Luxembourg. He obtained his Ph.D. degree ('94) in Computer Science at the University of Grenoble (INPG), France. His research at the IMAG laboratory focussed on Mapping and scheduling task graphs onto Distributed Memory Parallel Computers. Pascal Bouvry has academic and industry experience on 3 continents working as manager and executive in major companies such as FICS, Spacebel, SDC, Lat45, Metasolv. Professor Bouvry co-authored over 200 publications in international peer-reviewed venues. He also organized numerous top conferences worldwide and is member of several editorial boards, including IEEE Cloud Computing Magazine. Current interests of Professor Bouvry encompass cloud and grid computing, optimization issues, and related complex problems.



Albert Y. Zomaya is currently the Chair Professor of High Performance Computing & Networking in the School of Information Technologies, The University of Sydney. He is also the Director of the Centre for Distributed and High Performance Computing which was established in late 2009. Professor Zomaya published more than 500 scientific papers and articles and is author, co-author or editor of more than 20 books. He is the Editor in Chief of the IEEE Transactions on Computers and Springer's Scalable Computing and serves as an associate editor for 22 leading journals, such as, the ACM Computing Surveys and Journal of Parallel and Distributed Computing. Professor Zomaya is the recipient of the IEEE Technical Committee on Parallel Processing Outstanding Service Award (2011), the IEEE Technical Committee on Scalable Computing Medal for Excellence in Scalable Computing (2011), and the IEEE Computer Society Technical Achievement Award (2014). He is a Chartered Engineer, a Fellow of AAAS, IEEE, IET (UK). Professor Zomaya's research interests are in the areas of parallel and distributed computing and complex systems.