A Large-scale Power-saving Cloud System with a Distributed-management Scheme

Toshiaki Suzuki, Tomoyuki Iijima,
Isao Shimokawa, and Toshiaki Tarui
Central Research Laboratory
Hitachi, Ltd.
Yokohama, Kanagawa, Japan
{toshiaki.suzuki.cs, tomoyuki.iijima.fg,
isao.shimokawa.sd, toshiaki.tarui.my}@hitachi.com

Tomohiro Baba, Yasushi Kasugai,
and Akihiko Takase
Telecommunications & Network Systems Division
Hitachi, Ltd.
Kawasaki, Kanagawa, Japan
{tomohiro.baba.mn, yasushi.kasugai.rs,
akihiko.takase.wa}@hitachi.com

***Abstract*—A large-scale power-saving cloud system with a distributed management scheme is proposed. The system is composed of multiple data centers (DCs) connected by a wide-area network (WAN). In addition, it includes an inter-DC-management server, multiple DC-management servers, and multiple user VM-management servers. To reduce the power consumption of the DCs and the WAN, virtual machines (VMs) are migrated and data-routing paths are optimized under the condition that quality of service (QoS) is maintained by simultaneously providing necessary CPU resources and network bandwidth for services by the VMs. Aiming to enhance our previously proposed system, the management scheme is based on distributed management instead of central management. In the previous system, one DC-management server gathers all information to determine an appropriate alternative server to which VMs are migrated. On the contrary, in the proposed system, to distribute management load, each user VM-management server sends specifications of a VM to be migrated to other user VM-management servers. The other user VM-management servers then independently return a list of alternative servers that can accommodate the intended VM. After receiving the lists, the user VM-management server selects the most-suitable server. A prototype of the proposed system comprising 1,000 VMs, 400 servers, and four DCs was developed and evaluated. The time for determining reallocation of 1,000 VMs is within five minutes, which is about five times shorter than that taken by the previous system. These results indicate that the proposed system can reduce power consumption for one-week cloud operation by 30%.**

Keywords - power saving, QoS, cloud system, virtual-machine migration, distributed management, resource allocation.

I. INTRODUCTION

This work is an expansion of our previous work presented at ENERGY2013 [1]. In the previous work, a centralized management scheme for a large-scale power-saving cloud system composed of multiple data centers was focused on. In the current work, to enable the system to control a larger number of virtual machines (VMs), this scheme is extended to create a distributed management scheme. In addition, the performance of the extended scheme was evaluated in an environment with 1,000 VMs.

Lately, in conjunction with the increasing number of data centers (DCs) being constructed, the amount of electric power consumed by information and communication technology (ICT) systems has been dramatically rising [2]. As one of the biggest issues concerning ICT systems, including DCs, power-saving measures have therefore been attracting lots of attention [3].

To address power-consumption issues, many standardizations and technical developments aiming to make ICT systems more power efficient are being actively promoted. Although conventional activities have aimed at reducing the electric-power consumption of ICT systems, the respective power consumptions of the "server resource" and "network resource" are controlled separately. Conventionally, power-saving control has therefore been optimized on a resource-to-resource basis. However, total electric-power consumption by a whole large-scale cloud system, comprising multiple DCs and a wide-area network (WAN) connecting them, has not been optimized while service quality provided by the system is maintained. Besides, if power-saving control is conducted separately per resource, it might cause a serious problem for other resources. For example, an excessive aggregation of servers by VM migrations might degrade access quality to a VM since data flows are aggregated to the same routing path; as a result, network-link bandwidth is exceeded, and network congestion occurs.

With these issues in mind, we are aiming to develop an efficient power-saving control scheme for both network and server resources while "quality of service" (QoS) of networks and servers, such as bandwidth and CPU power, is guaranteed by integrated power-consumption management of both network and server resources. In a previous work [1][4], we proposed a power-saving cloud system centrally managed by one control server that gathers all information needed for determining allocation of VMs. This system, however, faces a scalability issue. In the present work, aiming at total power saving covering both WAN resources and DC resources, we propose a large-scale power-saving cloud system managed by cooperation between a WAN management server, user VM-management servers, and an inter-DC-management server.

The rest of this paper is organized as follows. Section II describes related works. Section III explains the requirements concerning a power-saving cloud system. Section IV proposes a large-scale power-saving cloud system with a distributed management scheme. The proposed system simultaneously saves electric power and guarantees access bandwidth to a VM. Sections V and VI respectively describe a prototype system and present the results of

evaluations concerning power saving and determining reallocation of 1,000 VMs. Section VII concludes the paper.

II. Related Work

To tackle power-consumption issues, many schemes have been proposed and standardization activities are ongoing. As for a power-saving scheme for ICT systems, "server-resource virtualization" (that is, saving power consumed by servers by optimizing necessary resources) has been under research and development [5][6]. In addition, power-saving schemes at the node and link levels [7][8] have been proposed. These schemes are useful for reducing the power consumption of our proposed cloud system at the link level. In addition, power-saving schemes [9][10][11][12] for the network level have been proposed. Power-saving schemes at the DC/server level [13][14][15][16] and at the inter-DC level [17][18] have also been proposed.

In the meantime, standardization activities, such as those undertaken by the Energy Management Working Group (EMAN) in the Internet Engineering Task Force (IETF) [19], the Institute of Electrical and Electronics Engineers (IEEE) [20], the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) [21], and the Distributed Management Task Force, Inc. (DMTF) [22], are continuing.

In conventional power-saving schemes like those mentioned above, network and DC/server resources are controlled separately and/or without consideration for the QoS of networks. In the current study, therefore, integrated management for maintaining network QoS and reducing energy consumption of servers is addressed.

III. Requirements of a Power-saving Cloud System

A power-saving cloud system provides various services and resources, such as application software, CPU processing power, and storage, via a network. To create a power-saving cloud system and to reduce its total electric-power consumption during off-peak hours (such as late evening), only the minimum resources required for providing cloud services should be activated.

To control the power consumption of a target system, average loads on physical servers, VMs on those servers, and network nodes should be monitored in real time. In addition, VMs should be appropriately reallocated according to predicted future loads on servers and VMs when the load is under a predefined threshold during off-peak hours. After the appropriate reallocation of VMs, unnecessary physical servers should be turned off. The nodes or ports on the nodes that transmitted data to unnecessary physical servers should also be turned off or switched from active mode to sleep mode. Furthermore, service quality (such as access bandwidth to a VM) should be guaranteed before, as well as after, the power-saving control by VM migration. In addition, a power-saving scheme should be applied to not only small cloud systems comprising a single DC but also large-scale systems comprising multiple DCs.

To satisfy the above-mentioned requirements, the power-saving control should be executed according to the following procedures, namely, four power-saving policies.

- Policy 1: Power consumption of the DC can be reduced by turning off unnecessary physical servers that are no longer used after an appropriate reallocation of VMs in the DC by VM migration.
- Policy 2: Power consumption of the DC can be reduced by turning off unnecessary physical servers and network nodes that are no longer used after aggregation of running physical servers and data-transmission routes in the DC by VM migration.
- Policy 3: Power consumption of the DC can be reduced by turning off unnecessary physical servers and nodes in the DCs that are no longer used after aggregation of running physical servers and data-transmission routes by VM migration between DCs (based on cooperation between DC management and WAN management).
- Policy 4: Power consumptions of the DC and WAN can be reduced by turning off nodes (or their ports) in the WAN that are no longer used after VM migration between DCs and aggregation of data-transmission routes.

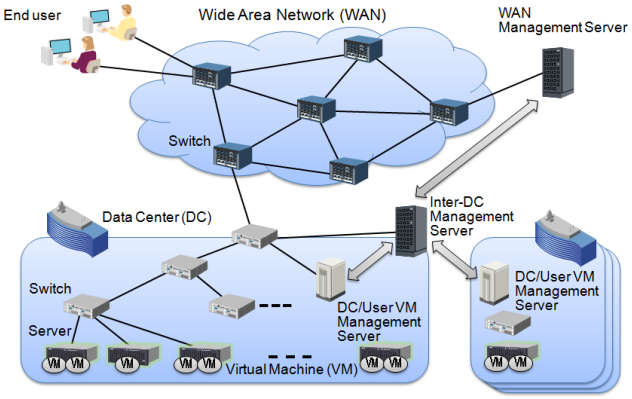

Figure 1. Proposed power-saving cloud system

On the basis of the above four policies, resources can be controlled from the viewpoint of power saving as well as from the viewpoint of service quality. More specifically, power consumption of the system should be reduced by aggregation of both server resources and network resources while service quality of a network path between an end user and the VM providing application services is maintained.

IV. Proposed Power-saving Cloud System

A large-scale power-saving cloud system with a distributed-management scheme is proposed as follows. Specifically, a distributed power-saving management structure and its procedures are described.

A. System architecture

The architecture of the proposed power-saving cloud system is shown schematically in Figure 1. The system is composed of multiple DCs connected by a WAN. More specifically, the DC consists of multiple switches (SWs) for transmitting data, servers for providing various services, a

DC/user-management server for controlling user resources in the DC, and an inter-DC-management server for controlling multiple DC-management servers. The WAN consists of multiple SWs and a WAN-management server for monitoring and controlling resources in the WAN.

In the power-saving cloud system, the DC-management server monitors the loads of networks in the DC in real time. On the other hand, the user VM-management server monitors the loads of servers and VMs in the DC in real time. In addition, the DC- and user VM-management servers predict future loads of the resources, namely, servers, VMs, and SWs, by using statistical analysis (for example, by an autoregressive model [23]) based on the past history of loads. Specifically, the loads of these resources for eight hours are predicted according to their seven-day history of every five minutes. Besides, to reduce power consumption on the DC side, the management servers determine and control reallocation of resources such as VMs and routing paths.

To reduce power consumption of the WAN, the WAN-management server monitors loads of SWs in the WAN. It then gathers statistical-monitoring data and predicts future loads on each SW. Electric power consumed by the WAN is saved by optimizing data-routing paths and turning off SWs (or their ports) that are no longer used.

In summary, power consumption of the whole system is reduced by reallocating VMs between the DCs appropriately on the basis of cooperation between multiple DCs and user VM-management servers and the WAN management server.

B. Distributed management structure

The management structure of the proposed large-scale power-saving cloud system is shown in Figure 2. In this structure, one inter-DC-management server controls multiple DC-management servers. Each DC has one DC-management server that monitors network conditions of each DC and predicts future loads of the network. Multiple user VM-management servers monitor loads of multiple user servers and predict future loads of the servers and VMs. The user is assigned resources such as VMs and networks, and those resources are separated by a VLAN from other users' resources.

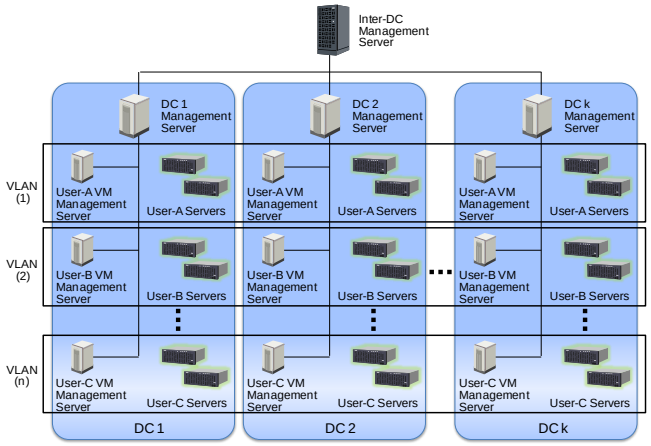

Figure 2. Distributed management structure for multiple DCs

C. Overview of power-saving scheme by VM migration

The process steps of a typical power-saving scheme based on VM migration by multiple management servers are shown schematically in Figure 3. In the proposed system, the inter-DC management server activates power-saving control according to the loads on the physical servers and VMs [step (1)]. The user VM-management server determines the order of VM migration [step (2)]. It then obtains "congestion potential" via the inter-DC-management server and the WAN-management server [step (3)]. To migrate VMs between DCs, the user VM-management server sends resource sizes, which are needed by the VM to be migrated, to other user VM-management servers in other DCs [step (4)]. These other user VM-management servers then send lists of servers that can accommodate the VM to be migrated [step (5)]. In addition, the original user VM-management server receives predicted future loads (e.g., SWs) on the DC network [step (6)]. To move the VM according to the lists received from the outside servers, the predicted loads of the DCs, and the effectiveness of the power saving, the user VM-management server determines one alternative server to which the VM is migrated [step (7)]. It then triggers an actual VM migration [step (8)]. The VM-migration result (i.e., notification of VM-migration completion) is transmitted from the user VM-management server to the inter-DC-management server [step (9)].

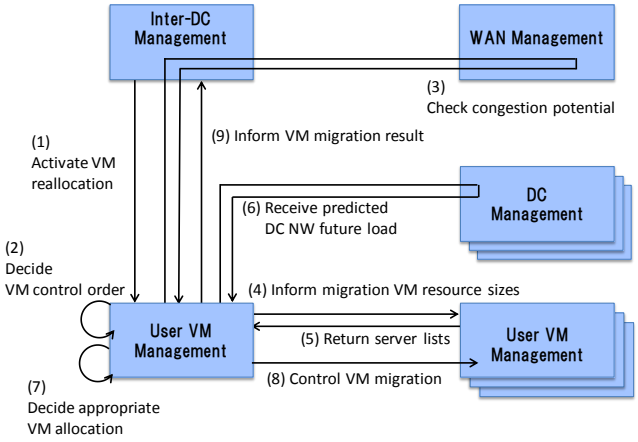

Figure 3. Process steps of reallocation of VM resources

D. Detailed VM-resource reallocation

The nine above-mentioned steps for resource reallocation are explained in detail as follows:

1) VM-reallocation trigger by inter-DC-management server: The inter-DC-management server starts (or stops) optimizing reallocation of VMs to each user VM-management server when the loads on servers and VMs are low (such as in the late evening).

2) Determination of VM-reallocation order by user VM-managementt server: The user VM-management server determines the order for reallocating running VMs in each virtual local-area network (VLAN). The reallocation order is determined according to (i) decending order of idle power,

(ii) ascending order of number of running VMs on a server, (iii) ascending order of assigned CPU resources, and (iv) ascending order of assigned memory resources.

3) Checking of congestion potential for the WAN by user VM-management server: To maintain access quality to a VM after the VM is migrated to another DC, the user VM-management server receives the congestion potential concerning the WAN from the WAN-management server via the inter-DC-management server. The congestion potential is evaluated on the basis of the history of the monitored data and predicted future loads in the case of fluctuation of bandwidth for each port of the switches. If network congestion is possible in the future, data-routing paths including the congestion point are not used for VM migration. More specifically, the IP address of the VM to reallocate, the identifier of the source DC, and the identifier of the VLAN to which the VM belongs are transmitted from the user VM-management server to the inter-DC-management server. A list of alternative DCs that can accommodate the migrated VM and the above-mentioned information from the user VM-management server are then transmitted from the inter-DC-management server to the WAN-management server. The congestion potential for the routing path between the user and an altenative DC is sent from the WAN-management server to the user VM-management server.

4) Informing other user VM-management servers about resource sizes of migrated VMs: The user VM-management server predicts future loads on the CPU and consumption of the bandwidth resource by the intended VM. It then sends the required sizes of resources by the VM to other user VM-management servers. In our previous system [1], the user VM-management server receives information concerning all servers from other user VM-management servers in other DCs. However, in that case, the user VM-management server has to choose one target server by itself from huge lists of alternative servers. However, scalability of this procedure is an issue. In the case of the proposed system, the user VM-management server therefore informs other user VM-management servers about sizes of required resouces to accommodate a VM. The other user VM-management servers then search for altenative servers.

5) Returning server lists of alternative servers: The user VM-management servers in other DCs receive the sizes of required resources to accommodate the VM to be migrated and search for altenative servers according to the required resources. The user VM-management servers in other DCs return lists of alternative servers that can accommodate the VM to be migrated.

6) Receiving predicted future load of a DC network: The DC-management server in each DC monitors the conditions of the network in the DC and predicts future loads of the network. The user VM-management server receives the predicted future loads of networks in each DC.

7) Determination of target server: The user VM-management server detemines an appropriate VM reallocation by considering all alternative DCs. Specifically, all servers that can provide enough resources to run the intended VM in the future and maintain access qualtiy to the VM at the same time are selected as alternative servers for the reallocation of the VM. The most-effective server for power saving is then selected as the final target server for the VM migration.

The user VM-management server receives lists of altenative servers that can accommodate the intended VM [steps (4) and (5)]. It also receives future loads of networks of other DCs [step (6)]. On the other hand, it predicts future loads on the CPU and consumption of the bandwidth resource by the intended VM [step (4)]. It temporarily determines several target servers to which the intended VM is reallocated by comparing the received available future resources for all alternative servers in other DCs and the amount of necessary resources for the intended VM.

The user VM-management server determines whether switches on the routing path between the entrance of the DC and an alternative server in another DC can provide enough bandwidth for the intended VM after the VM migration according to the infromation from step (3). On the basis of the monitored information from the WAN-management server, it checks the congestion potential for the routing path between the WAN edge connecting the DC and another WAN edge connecting an end user. To determine the most-appropriate alternative server, the user VM-management server checks all the above-mentioned evaluation points, i.e., CPU load, network congestion, and bandwidth. The most-appropriate server that can meet the requirements stated in Section III and has the most-effective power-saving advantage is then selected by the user VM-management server as the target server for the VM migration.

8) VM migration by user VM-management server: The VM migration is executed according to the trigger by the user VM-management server. As for VM-migration methods, various technologies have been developed [5], [6] and can be used for an alternative VM-migration scheme by combining them with the proposed power-saving cloud system. After executing the VM migration, the user VM-management server updates stored topology information. In addition, to predict future load, when the VM has been migrated to a server in another DC, the history of the VM's resources (such as CPU load) is moved to another user VM-management server.

9) Information about VM-migration completion sent from user VM-management server to inter-DC-management server: After all VM migrations have been executed, the user VM-management server informs the inter-DC-management server that all VM reallocations are complete. In addition, the histories of migration from the source servers to destination servers are transmitted from the user VM-management server to the inter-DC-management server, which receives the migration histories and stores them. These histories are used when the migrated VMs are returned to the original allocated servers when CPU load increases.

E. Procedure for selecting alternative servers

The procedure for selecting alternative servers is depicted in Figure 4. The user VM-management server that is instructed to start VM reallocation by the inter-DC-management server takes the main role of a server for determining the allocation of VMs for each user. On the other hand, the user VM-management servers that belong to other DCs take the role of sub-servers.

The main user VM-management server makes a list that includes all servers that provide application services in order of power-consumption efficiency. In addition, it monitors loads of the VMs, servers, and their network ports. It then predicts future loads of those resources and manages the list with predicted loads of those resources. After receiving a request for reallocation of VMs, it sends predicted loads of the VM to be migrated to other user VM-management servers in parallel [step (1) in Figure 4].

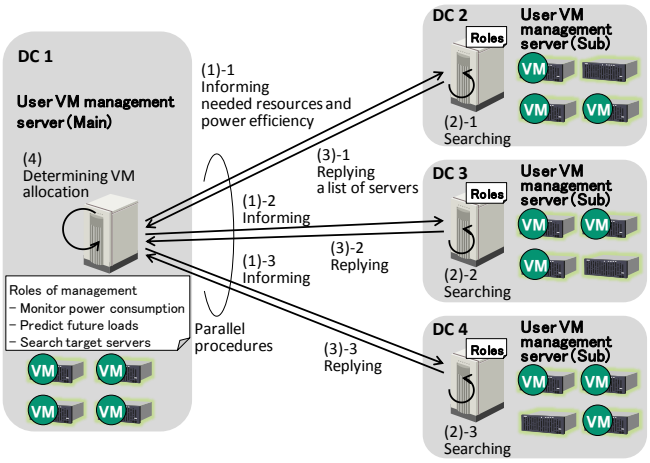

Figure 4. Procedure for selecting alternative servers

The other user VM-management servers acting as sub servers also make a list of servers to provide application services and manage it on the basis of the order of power-consumption efficiency. In addition, the sub user VM-management servers monitor loads of resources, such as the servers and their network ports, and predict future loads of those resources. When they receive requests to find suitable servers to accommodate the VM to be migrated from other DCs, each sub user VM-management server searches for possible servers by comparing the resources to be consumed by the migrated VM and remaining resources of each server [step (2)]. After finding possible servers, each sub user VM-management server sends a list of possible servers in parallel. In the process of finding the possible servers, when the other user VM-management servers find several alternative servers that can accommodates the migrated VM, they stop the search process and send the search result to the main user VM-management server that sent the request [step (3)].

When the main user VM-management server receives the lists of alternative servers from sub user VM-management servers, it selects several alternative servers in order of power-consumption efficiency. It then checks whether the alternative servers can accommodate the VM to be migrated and whether there is enough network bandwidth from a user to the DC to which the alternative server belongs. In addition, it determines one target server to which the VM is to be migrated [step (4)].

F. Scale-out function

When the loads of VMs' resources (such as CPU and consumed bandwidth) decrease, since the loads of resources used by each VM are very low, those VMs are aggregated to other servers to make "unnecessary servers". In addition, unnecessary servers are shut down or turned to sleep mode, and power consumption is decreased. On the contrary, when the loads of VMs or used bandwidth increase, the VMs should be distributed to multiple servers or the traffic should be detoured to other routes. When the VMs are again migrated, if the original serves are not on, they are turned on. To do that, the original server position where the VM is executed is memorized, and VMs are again migrated to the original servers when their loads increase.

G. Power-consumption model

A power-consumption model for the proposed cloud system is defined as follows. The amount of power (P_{All}) consumed by the cloud system is given by formula (1), where P_{IT} means power consumption of IT equipment, and P_{NET} means power consumption of network nodes. Formula (2) indicates P_{IT} is calculated by summing the power consumption of each server (P_{SV}) since the proposed system includes multiple servers as IT equipment. Here, i (i = 1, 2, 3, . . , N) means the number of the server. In addition, n means CPU load (%) on the server. P_{SV} is given by formula (3) [24][25]. $P_{idle(i)}$ means the power consumption of the ith server during idle time, and $P_{max(i)}$ means power consumption under maximum load. Formula (4) gives P_{NET} of a network calculated by summing the power consumption of each node. Here, k (k = 1, 2, 3, . . , M) means the number of the node. In addition, m means load (%) on a node in terms of bandwidth. The power consumption of the node (P_{NODE}) is given by formula (5) [7]. $P_{idle(k)}$ means power consumption by the kth node during idle time, and $P_{max(k)}$ means power consumption under maximum load. Here, P_{SV} and P_{NODE} are assumed to fit a linear function, as shown in Figure 5. The relations between power consumption and CPU load and between power consumption and traffic are independently evaluated in advance. According to that evaluation, the relation between power consumption and load (traffic) fits a linear function well (as shown in Figure 5).

$$P_{All} = P_{IT} + P_{NET} \quad (1)$$

$$P_{IT} = \Sigma_i P_{SV(i)}[n] \quad (2)$$

$$P_{SV(i)}[n] = P_{idle(i)} + (P_{max(i)} - P_{idle(i)})(n/100) \quad (3)$$

$$P_{NET} = \Sigma_k P_{NODE(k)}[m] \quad (4)$$

$$P_{NODE(k)}[m] = P_{idle(k)} + (P_{max(k)} - P_{idle(k)})(m/100) \quad (5)$$

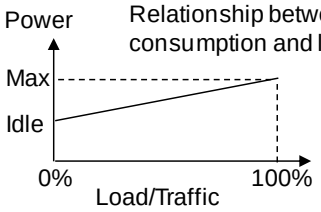

Figure 5. Assumed power consumption based on load/traffic

V. EVALUATION OF POWER SAVING

The performance of the proposed system with the centralized management scheme was evaluated by using an assumed CPU load model of a VM and consumed bandwidth by the VM per day.

A. Evaluation system

The evaluation system is shown schematically in Figure 6, and the number of pieces of ICT equipment is listed in Table I. In this system, switches, servers, and VMs in the DCs are emulated by open-source software, while switches in the WAN and management servers are real hardware. The performance (i.e., power-consumption reduction) of the power-saving control scheme for the DCs and WAN was evaluated as explained in detail in the following sections.

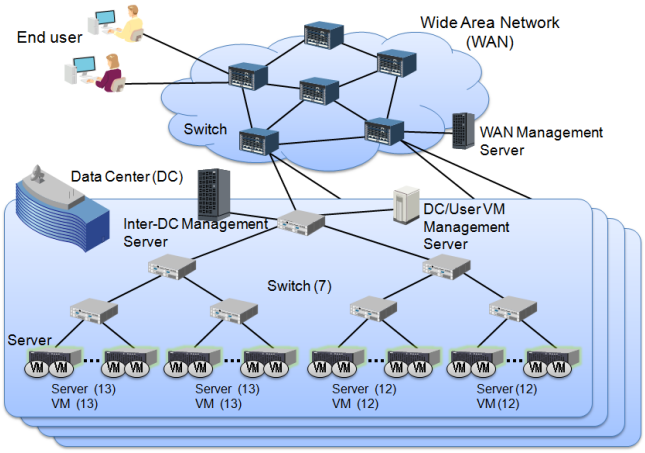

Figure 6. Evaluation system

TABLE I. NUMBER OF PIECES OF ICT EQUIPMENT

	Item	Number of pieces of ICT equipment in four DCs
1	WAN-management server	1
2	SWs in WAN	6
3	DCs	4
4	Inter-DC-management server	1
5	DC/user VM-management servers	4
6	SWs in DCs	28
7	Servers in DCs	200
8	VMs on servers	200

B. Evaluation of power-saving control for DCs

The effectiveness of applying the power-saving control scheme for DCs per day was evaluated. First, a CPU-load model of a VM in the DC is assumed. The bandwidth consumed by the VM for one day is also assumed. The power consumed by DCs for one day is then evaluated on the basis of these assumptions.

1) Workload model for a VM per day

The assumed loads on the CPU as well as the incoming data flow to and the outgoing data flow from a VM are schematically shown in Figure 7. As depicted in the figure, the peak load is set only one time (around noon), and the loads during business hours are high, while the loads during the night (namely, those of the CPU, incoming flow, and outgoing flow) are low. The effectiveness of the power-saving control scheme is evaluated by comparing two cases: either executing appropriate VM reallocations or not.

The topology of the DC is shown in the lower part of Figure 6. The specifications and number of pieces of each apparatus in the DC are listed in Table II.

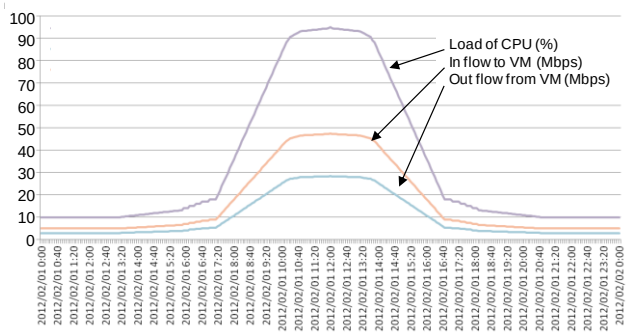

Figure 7. Load model of VM (CPU and in/out data flow)

TABLE II. SPECIFICATIONS AND NUMBER OF PIECES OF EACH APPARATUS IN ONE DC

	Apparatus	Idle power	Max. power	Number
1	Server (Model 1)	120 W	170 W	17
2	Server (Model 2)	110 W	150 W	17
3	Server (Model 3)	177 W	251 W	16
4	VM	—	—	50
5	Switch	350 W	450 W	7

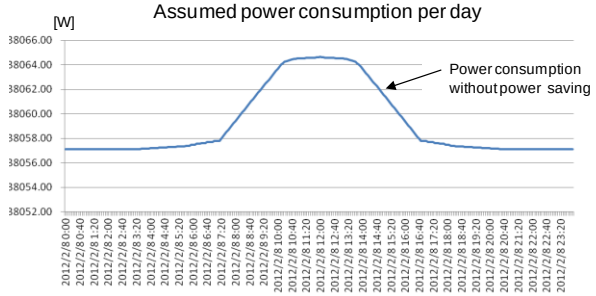

Figure 8. Electric-power consumption of a DC per day

2) Electric-power consumption of a DC per day

Electric-power consumption of a DC for one day (under the assumed loads for each server shown in Figure 7) is shown in Figure 8. The effectiveness of the power-saving control scheme under the following three conditions was evaluated. In the first condition, the VM is reallocated when the CPU loads are less than 75%. In the second and third conditions, reallocations are executed under CPU loads of 50% and 25%, respectively. On the other hand, when the load on the CPU is over these thresholds, reallocated VMs are returned to the original locations to maintain service quality.

3) Energy consumption of DCs per day

The evaluated fluctuations of power consumption of all DCs for the three above-mentioned conditions concerning power-saving control (CPU loads of 75%, 50%, and 25%) are shown in Figure 9. The result in the case of no VM reallocation is also shown in the figure for comparison. The figure verifies the effectiveness of the power-saving control scheme under the three conditions.

In addition, the results for VM reallocation to maintain VM access quality and energy consumption per day are listed in Table III. The number of VMs is shown in the upper row, while the number of servers (SVs) is shown in parentheses in the lower row. According to the table, some VMs are migrated between DCs (since the number of VMs in the DC changes after appropriate VM reallocations). In addition, the number of running servers is dramatically reduced after the VM migration. The CPU resource for a server is assumed to be enough for six VMs with CPU loads of 50%. The reductions in energy consumption under CPU loads of 25%, 50%, and 75% are 45.2%, 45.7%, and 47.6%, respectively.

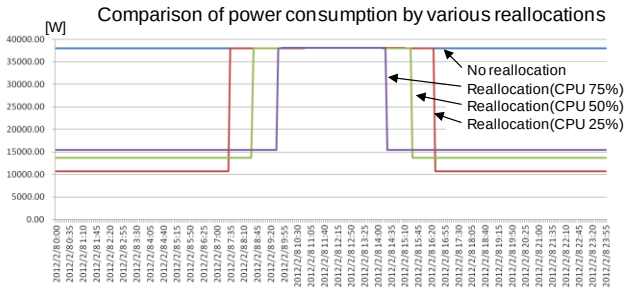

Figure 9. Electric-power consumption of a DC per day

TABLE III. ELECTRIC-ENERGY CONSUMPTION OF DCS PER DAY

	Optimization timing	DC1 VM (SV)	DC2 VM (SV)	DC3 VM (SV)	DC4 VM (SV)	Energy consumption per day (kWh)
1	No optimization	50 (50)	50 (50)	50 (50)	50 (50)	913.421
2	CPU load: 25%	60 (5)	48 (4)	48 (4)	44 (4)	500.388
3	CPU load: 50%	54 (9)	48 (8)	48 (8)	50 (9)	496.395
4	CPU load: 75%	52 (13)	48 (12)	52 (13)	48 (12)	478.323

C. Evaluation of power-saving control for a WAN

Power-saving control for a wide-area network (WAN) for one day was evaluated. In particular, the effectiveness of the power-saving scheme (based on bandwidth control by link aggregation) was evaluated. The topology of the evaluated WAN is shown in Figure 6. The specifications of the switches in the WAN are the same as those listed in Table II.

Power-saving control by appropriate data routing (including link-aggregation control) was executed after appropriate VM reallocation between DCs. The fluctuation of power consumption of the WAN is shown in Figure 10. Energy consumptions under the three types of control are compared in Table IV. According to these results, the reductions in energy consumption achieved by the power-saving control scheme under CPU loads of 25%, 50%, and 75% are 10.4%, 12.0%, and 13.7%, respectively.

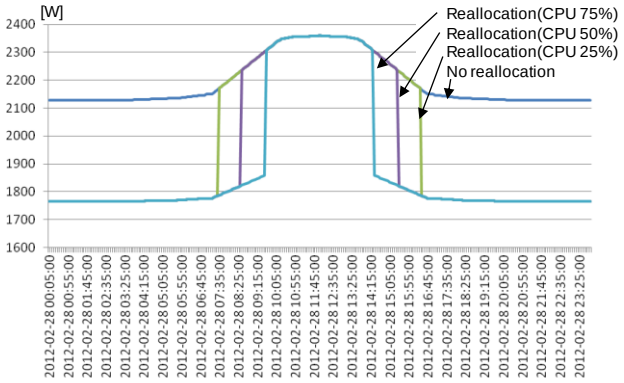

Figure 10. Electric power consumption of a WAN per day

TABLE IV. ELECTRIC-ENERGY CONSUMPTION OF ONE WAN PER DAY

	Optimization timing	Electric-energy consumption per day (kWh)	Reduction (%)
1	No optimization	52.470	—
2	25% CPU load	46.989	10.4
3	50% CPU load	46.194	12.0
4	75% CPU load	45.265	13.7

D. Power-saving effect for entire cloud system

The effectiveness of the power-saving control scheme for the entire proposed cloud system is shown in Table V. According to the table, the reductions of energy consumption achieved by the power-saving control scheme under CPU loads of 25%, 50%, and 75% are 43.3%, 43.8%, and 45.8%, respectively. In other words, energy consumption is reduced by approximately 40%. In addition, the highest reduction is accomplished under CPU load of 75%.

TABLE V. ELECTRIC-ENERGY CONSUMPTION OF ENTIRE CLOUD SYSTEM PER DAY

	Optimization timing	Optimization term	Electric-energy consumption (kWh)	Reduction (%)
1	No optimization	—	965.891	—
2	CPU load: 25%	15h00m	547.377	43.3
3	CPU load: 50%	17h00m	542.589	43.8
4	CPU load: 75%	19h10m	523.588	45.8

E. Discussion of power-saving effect

According to the results of this evaluation of a large-scale power-saving cloud system composed of multiple DCs and a WAN, energy consumption of the entire system is reduced by about 40% by the proposed power-saving control scheme. With regard to the power saving for the DCs only, energy consumption is reduced by over 45%. On the other hand, energy consumption of the WAN is reduced by only about 10%. The reason that the reduction of energy consumption of the DCs is high is the effectiveness of turning off unnecessary servers after appropriate VM reallocation. On the other hand, the reason that the reduction of the energy consumption of the WAN is low is that unnecessary switches were not turned off (since turning off unnecessary links is only possible under the assumed evaluation conditions). In the evaluation, migrations of management servers are not considered since they are not removable. Although energy consumed by them should be considered, their energy consumption is a bit small since the number of turned-off servers is much larger than that of management servers. In addition, the function for controlling network and server resources in a coordinated manner can be evaluated regardless of their energy consumptions. With regard to power saving for the entire proposed system, the reductions in energy consumption achieved by the power-saving control scheme under CPU loads lower than 25%, 50%, and 75% are 43.3%, 43.8%, and 45.8%, respectively. On the other hand, the times taken for the resource optimization under the three above conditions are 15 hours, 17 hours, and 19 hours and 10 minutes, respectively. When the power-saving control is executed under a CPU load of 75%, the time taken for the optimization is the longest, and reduction in energy consumption is the highest. These results verify the effectiveness of the proposed power-saving control scheme.

VI. EVALUATION OF LARGE CLOUD WITH 1,000 VMS

As explained in the previous section, the power-saving efficiency of the proposed scheme for a cloud system with 200 VMs was evaluated. In addition, as explained in this section, the times for determining VM reallocation and power-saving efficiency of a cloud system with 1,000 VMs for both the previously proposed system and the presently proposed system were evaluated.

A. Evaluation of time for determination of VM reallocation

1) Structure of evaluation system

The evaluation system includes 1,000 VMs and 400 servers in four DCs (as shown in Table VI). In addition, a WAN-management server, an inter-DC-management server, four DC/user VM-management servers are included. As for this system, switches, servers, and VMs in the DCs are emulated by open-source software, while switches in the WAN and management servers are real apparatuses.

TABLE VI. NUMBER OF PIECES OF ICT EQUIPMENT

	Item	Number of pieces of ICT equipment in 4 DCs
1	WAN-management server	1
2	SWs in WAN	3
3	DCs	4
4	Inter-DC-management server	1
5	DC/user VM-management servers	4
6	SWs in DCs	92
7	Servers in DCs	400
8	VMs on servers	1,000

2) Evaluation parameters and conditions

The main user VM-management server receives a list of multiple alternative servers from the sub user VM-management servers. The number of alternative servers is set as a parameter of the evaluation. As evaluation conditions, 768 VMs are migrated from original servers to other servers, and 232 VMs are not migrated. In addition, after the reallocation, the number of active servers is reduced from 400 to 84 (as listed in Table VII).

TABLE VII. NUMBER OF PIECES OF ICT EQUIPMENT AFTER VM REALLOCATION

	Number of DCs	Active SWs	Active servers	Number of VMs
1	DC 1	23	21	252
2	DC 2	23	21	252
3	DC 3	23	21	252
4	DC 4	23	21	244
	Total	92	84	1,000

3) Evaluation results

The evaluation results are shown in Figure 11. In the figure, the time for determining reallocation of 1,000 VMs is depicted. The horizontal axis shows the number of alternative servers per DC. Specifically, the number “1” means that only one alternative server is on the list sent from each other user VM-management server when the main user VM-management server requests an alternative server to migrate a VM to. On the other hand, the number “10” means that 10 alternative servers are on the list sent from each other user VM-management server. The vertical axis shows the time to determine reallocation of 1,000 VMs.

According to the results shown in Figure 11, in the case of our previous scheme with centralized management, over

twenty-one minutes (1274 seconds) are needed to determine reallocation. On the other hand, in the case of the proposed distributed management scheme, the determination is done within about five minutes (250 seconds) when the main user VM-management server receives a list with one alternative server from each other user VM-management server. The determination time is about five times shorter than that taken by the previous system. Even if the main user VM-management server receives lists with five alternative servers, seven minutes (303 seconds) are enough to determine reallocation of 1,000 VMs.

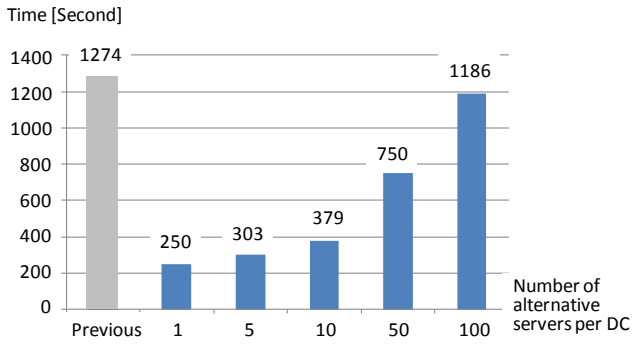

Figure 11. Time for determining VM reallocation

4) Discussion on performance of determining VM reallocation

In the previous centalized management scheme, the user VM-management server gathers information about future loads on all servers and SWs, and it determines one target server to migrate a VM to while guaranteeing QoS such as bandwidth to access a VM. On the other hand, in the proposed distributed management scheme, the user VM-management server sends requests for a list of alternative servers to other management servers in parallel. The load of the user VM-management server in the proposed system is therefore lower than that of the user VM-management server in the previous centalized-management-based system.

According to the evaluation results presented above, the proposed system has better performance in determining reallocation of 1,000 VMs. To move 1,000 VMs to other servers, it might take one hour (including the determination time for the reallocation). However, if 1,000 VMs are reallocated within one hour, it might be possible to reduce power consumption by about 30% because the proposed distibuted-control-based system easily changes the mode from normal operation to power-saving operation if the migration of 1,000 VMs is done within one hour. As shown by these evaluation results, the proposed system had enough time to reduce power consumption.

B. Evaluation of power saving for one week

The power-saving performance of the proposed system for one week was evaluated by assuming CPU and trafic loads based on a real-world server providing business applications.

1) Structure and conditions concerning evaluation system

The system used for evaluating the power-saving scheme is the same as that used in the evaluation mentioned above. However, as shown in Figures 12 to 15, models of CPU load and traffic are defined for one day in detail. In this evaluation, power-saving efficiencies for one day and one week were evaluated. Specifically, the power-saving efficiency on one weekday and one weekend day were evaluated. The power-saving efficiency for one week was also evaluated as total power saving over seven days. Specifically, total one-week power consumption was calculated by adding the power consumptions for five weekdays and one weekend.

2) Load modes and power-saving operation

Models of load and power consumption are shown in Figs. 12 and 13, which show a load model for a VM's CPU and a traffic model for VMs for one weekday, respectively. A load model for a VM's CPU and a traffic model for VMs for one weekend day are shown in Figs. 14 and 15, respectively. As shown in Figs. 12 and 13, high-load situations occur two times. Therefore, on the weekday, power-saving operation is done twice; however, on the weekend, the power-saving operation is done once.

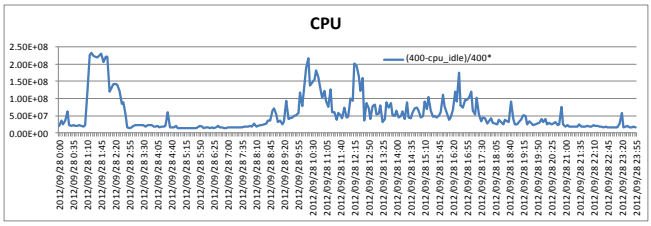

Figure 12. VM's CPU load model for a weekday

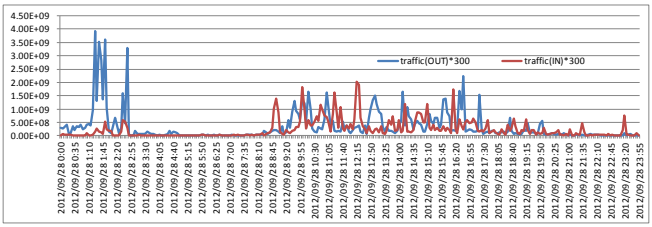

Figure 13. VMs' traffic models for a weekday

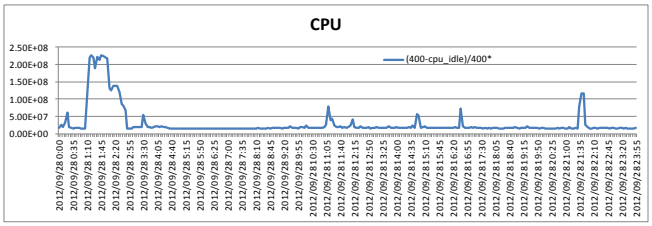

Figure 14. VM's CPU load model for a weekend day

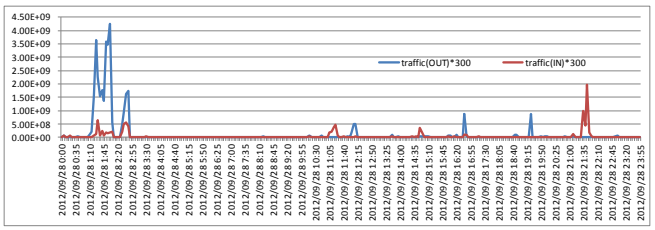

Figure 15. VMs' traffic models for a weekend day

3) Results of evaluation of power-saving control scheme

The effectiveness of the power-saving control scheme for one day and one week is shown in Table VIII. The reduction of energy consumption for one weekday is 24.3%, while the reduction for one weekend day is 46.6%. The energy consumption for one week is calculated by summing the energy consumptions for five weekdays and one weekend. In addition, the reduction for one week is 30.6%. As shown in Table VIII, the proposed system can reduce electric-energy consumption by 30%.

TABLE VIII. ELECTRIC-ENERGY CONSUMPTION OF A SYSTEM WITH 1,000 VMS

	Day	Without energy saving (kWh)	With energy saving (kWh)	Reduction (%)
1	1 weekday	2,076.707	1,572.646	24.3
2	1 weekend day	2,072.799	1,107.081	46.6
3	5 weekdays	10,383.535	7,863.230	24.3
4	1 weekend	4,145.598	2,214.162	46.6
5	1 week	14,529.133	10,077.392	30.6

C. Limitation of prototype system

The above-described evaluations verified a coordinated control function between networks' and servers' resources that maintains network QoS, such as network bandwidth consumed by a VM before it is migrated. In the evaluation, the QoS was controlled according to predicted future network traffic so that network congestion does not occur. From that perspective, a prototype system can guarantee the QoS. However, unpredicted network traffic occurs rarely. To guarantee the QoS of the network traffic with high priority even if unpredicted network traffic occurs, the system should create a transmission path with reserverd network bandwidth.

There is a tradeoff between perfect optimization of energy consumption and guaranteeing network QoS. To provide the perfect optimization, future traffic must be predicted perfectly. A level of the optimization depends on usage efficiency of network bandwidths. To increase the level of optimization, fewer network nodes are used. In that case, the risk of violation of the QoS increases. It therefore seems that there is an important balance between optimization of energy consumption and maintaining network QoS.

To evaluate energy saving and all communication overheads by management servers accurately, it is better to use a power meter for all network nodes and severs, since indirectly calculated power consumption was used in the present evaluation. In the proposed system, the effect of the management servers on communication overheads was not evaluated and remains as future work.

VII. CONCLUSION AND FUTURE WORK

A large-scale power-saving cloud system with a distributed-management scheme is proposed. The system is composed of multiple DCs connected by a WAN. It also includes an inter-DC-management server, multiple DC-management servers, and multiple user VM-management servers. In the proposed system, VMs are reallocated to reduce power consumption under condition of guaranteeing necessary CPU resources and network bandwidth for providing cloud services. Power saving covering the entire system is executed by cooperation between user VM-management servers and an inter-DC-management server.

The management scheme is based on distributed management instead of the central management of a previous system. In the proposed system, to distribute management load, each user VM-management server sends specifications of the intended VM to other user VM-management servers. The other user VM-management servers then independently return a list of alternative servers that can accommodate the intended VM. After receiving the lists, the user VM-management server selects the most suitable server to which the VM is migrated to and thus reduce electric-energy consumption of the entire system.

A prototype system, composed of 1,000 VMs, 400 servers, and four DCs, was developed and evaluated. The time for determining reallocation of 1,000 VMs is within five minutes, which is about five times shorter than that taken by the previous system. In addition, the evaluation results verify proper reallocation of VMs between DCs and the possibility of energy saving by approximately 30% for one-week cloud operation (under the conditions assumed in this evaluation).

For further study, the proposed power-saving cloud system will be evaluated by considering multiple real-world servers providing a diversity of business applications. In addition, it will be evaluated using power meters in consideration of the effect of multiple management servers on communication overheads. Besides, it will be implemented and evaluated by using real-world DC resources. As a result, it is expected that the proposed system will be enhanced from laboratory quality so that it can be applied as a real cloud system.

ACKNOWLEDGMENTS

Part of this research was supported by the MIC (The Japanese Ministry of Internal Affairs and Communications) projects "Research and Development on Signaling Technology of Network Configurations for Sustainable Environment" and "Research and Development on

Power-saving Communication Technology – Realization of the Eco-Internet".

REFERENCES

[1] T. Suzuki et al., "A Large-scale power-saving cloud system composed of multiple data centers," The Third International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies (ENERGY 2013), pp. 127-133, Mar. 2013.

[2] C. L. Belady, Microsoft Corporation, "Projecting annual new datacenter construction market size," Mar. 2011, http://cdn.globalfoundationservices.com/documents/Projecting_Annual_New_Data_Center_Construction_PDF.pdf [retrieved: May 2014].

[3] GreenTouch, "Our mission," http://www.greentouch.org/index.php?page=about-us/ [retrieved: May 2014].

[4] T. Suzuki et al., "Power-saving ICT platform that guarantees network bandwidth for cloud-service systems," TS-A4: Cloud Computing Technical Session, World Telecommunications Congress, Mar. 2012.

[5] VMware, Inc., "VMware distributed power management," http://www.vmware.com/resources/techresources/1080 [retrieved: May 2014].

[6] Xen Homepage, http://www.xen.org/ [retrieved: May 2014].

[7] M. Yamada, T. Yazaki, N. Matsuyama, and T. Hayashi, "Power efficient approach and performance control for routers," Proc. of IEEE International Conference on Communications (ICC Workshops 2009), pp. 1-5, June 2009.

[8] Y. Fukuda, T. Ikenaga, H. Tamura, M. Uchida, K. Kawahara, and Y. Oie, "Performance evaluation of power saving scheme with dynamic transmission capacity control," Proc. of IEEE Globecom Workshops 2009, pp. 1-5, Nov./Dec. 2009.

[9] J. Baliga, R. Ayre, K. Hinton, W. V. Sorin, and R. S. Tucker, "Energy consumption in optical IP networks," Journal of Lightwave Techology, vol. 27, no. 13, pp. 2391-2403, July 2009.

[10] C. Lange, D. Kosiankowski, R. Weidmann, and A. Gladisch, "Energy consumption of telecommunication networks and related improvement options," IEEE Journal of selected topics in quantum electronics, vol. 17, no. 2, pp. 285-295, March/April 2011.

[11] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy efficiency in telecom optical networks," IEEE Communications surveys & tutorials, vol. 12, no. 4, pp. 441-458, Fourth quarter 2010.

[12] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: A servey of existing approaches and trends in energy-aware fixed nework infrastructures," IEEE Communications surveys & tutorials, vol. 13, no. 2, pp. 223-244, Second quarter 2011.

[13] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, "Green cloud computing: Balancing energy in processing, storage, and transport," Proc. of IEEE, vol. 99, no. 1, Jan. 2011.

[14] S. Yang, L. Chen, H. Tseng, H. Chung, and H. Lin, "Designing automatic power saving on virtualization environment," IEEE International Conference on Communication Technology (ICCT 2010), pp. 966-970, Nov. 2010.

[15] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely, "Data centers power reduction: A two time scale approach for delay tolerant workloads," Proc. of IEEE Infocom 2012, pp. 1431-1439, Mar. 2012.

[16] T. Imada, M. Sato, and H. Kimura, "Power and QoS performance characteristics of virtualized servers," Proc. of IEEE/ACM International Conference on Grid Computing, pp. 232-240, Oct. 2009.

[17] B. Kantarci, L. Foschini, A. Corradi, and H.T. Mouftah, "Inter-and-intra data center VM-placement for energy-efficient large-Scale cloud systems," Proc. of IEEE Globecom Workshops on Management and Security technologies for Cloud Computing 2012, pp. 708-713, Dec. 2012

[18] A.Q. Lawey, T.E.H.El-Gorashi, and M.H. Elmirghani, "Distributed energy efficient clouds over core networks," Journal of Lightwave Techology, vol. 32, no. 7, pp. 1261-1281, April 2014.

[19] The Internet Engineering Task Force (IETF), "Energy management working group (EMAN WG) charter," http://datatracker.ietf.org/wg/eman/charter/ [retrieved: May 2014].

[20] Institute of Electrical and Electronics Engineers (IEEE), "IEEE802.1," http://www.ieee802.org/1/ [retrieved: May 2014].

[21] International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Study Group 13 (SG13), "Question 21/13 – Future networks," http://www.itu.int/ITU-T/studygroups/com13/sg13-q21.html [retrieved: May 2014].

[22] Distributed Management Task Force, Inc. (DMTF), "Cloud management initiative," http://www.dmtf.org/standards/cloud [retrieved: May 2014].

[23] International Business Machines Corp. Homepage, http://www-01.ibm.com/software/analytics/spss/products/statistics/forecasting/ [retrieved: May 2014].

[24] T. Mukherjee, G. Varsamopoulos, S. K. S. Gupta, and S. Rungta, "Measurement-based power profiling of data center equipment", 2007 IEEE International Conference on Cluster Computing, pp.476-477, Sept. 2007.

[25] L.A. Barroso and U. Hölzle, "The case for energy-proportional computing," IEEE Computer, vol. 40, no. 12, pp.33-37, Dec. 2007.