Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Minimizing the Monetary Penalty and Energy Cost of Server Migration Service

Yukinobu Fukushima[1] | Tatsuya Suda[2] | Tutomu Murase[3] | Yuya Tarutani[4] | Tokumi Yokohira[4]

[1]The Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan
[2]University Netgroup Inc., CA, USA.
[3]Information Technology Center, Nagoya University, Aichi, Japan
[4]The Graduate School of Interdisciplinary Science and Engineering in Health Systems, Okayama University, Okayama, Japan

Correspondence
Yukinobu Fukushima
Email: fukusima@okayama-u.ac.jp

Summary

Current IaaS (Infrastructure as a Service) cloud service may not satisfy communication QoS (Quality of Service) requirements of delay-sensitive network applications, if there is a significant physical distance between a server of the network application (NetApp server) at a data center and its network application clients (NetApp clients). In order to improve communication QoS of NetApp clients, we propose Server Migration Service (SMS) in this paper. SMS allows NetApp servers to migrate among different locations in the network (1) to optimally locate themselves in relation to NetApp clients and mitigate the QoS degradation caused by location-related factors (i.e., propagation delays on network links) and (2) to optimally distribute traffic load over routers and processing load over (physical) computers and decrease the energy consumption. We develop a mixed-integer programming model that determines when and to which locations NetApp servers migrate to minimize the total operating cost of SMS, i.e., the sum of the monetary penalty incurred due to QoS violation and energy cost incurred due to energy consumption, while preventing NetApp servers from excessively migrating and adversely impacting QoS of the non-SMS service that share the resource of the substrate network with SMS. Simulation results show that the model developed in this paper achieves up to 42% lower total operating cost of SMS compared to the model that only minimizes the monetary penalty of SMS without considering the energy cost of SMS.

KEYWORDS:
IaaS, SMS (server migration service), server locations, energy cost, monetary penalty, mixed-integer programming

1 | INTRODUCTION

IaaS (Infrastructure as a Service) cloud service (e.g., Amazon EC2 [1], Google Compute Engine [2] and Microsoft Azure [3]) has been widely used for hosting network applications (NetApps). In IaaS cloud service, subscribers may operate their virtual machines (VMs) at a data center with modest initial capital investment of purchasing computer hardware and modest operation/management complexity. However, in the current IaaS cloud service, a VM, which runs a server side process (NetApp server) of a NetApp, is always placed at IaaS provider's data center and stays at the data center for its entire life-time. Because of this, if there is a significant physical distance between a NetApp server and its clients (NetApp clients) in the current IaaS cloud service, it is difficult to provide the NetApp with a desired level of communication QoS and support highly interactive NetApps such as augmented reality, online realtime games and autonomous driving. For instance, a multiplayer realtime online games require the round-trip delay between a game server (a NetApp server) and a game client (a NetApp client) to be less than 60 ms[4]. If the NetApp server is on the East Coast of the USA, and if its NetApp client is on the West Coast, the total round-trip delay including the propagation delay, processing delay, queuing delay and packet transmission delay will easily become longer than the required delay.

In order to improve communication QoS (NetApp QoS) for a NetApp in IaaS cloud service, we propose a new service, server migration service (SMS), in this paper. In SMS, micro data centers called work places (WPs) are deployed at various locations in a substrate network, and NetApp servers automatically and dynamically migrate among WPs towards their NetApp clients in order to support NetApp clients that dynamically change in their number and locations and to mitigate the degradation of NetApp QoS due to the physical distances among the NetApp clients. For instance, SMS enables a game server (a NetApp server) on the East Coast of the USA to automatically and dynamically migrate to a WP adjacent to its game client (a NetApp client) on the West Coast, and significantly reducing the round-trip delay and improving communication QoS between the NetApp server and NetApp client.

In SMS, a key research question is when and to which WPs NetApp servers should (automatically and dynamically) migrate in order to maximize the financial gain of the IaaS providers adopting SMS (SMS providers). In answering this question, both positive and negative impacts of server migration need to be considered. The positive impacts of server migration include (1) improvement of NetApp QoS and (2) reduction of power consumption in SMS equipment (i.e., routers and WPs). With SMS, NetApp servers may migrate closer to their NetApp clients and reduce the communication delay between them. With SMS, NetApp servers may also migrate to reduce the traffic load of routers and the processing load of WPs such that power consumption of routers and WPs are reduced. The negative impacts of server migration include possible degradation of communication QoS of non-SMS services that share the substrate network with SMS. Server migration traffic, i.e., traffic generated by migrating NetApp servers, may occupy a significant portion of the resources of the substrate network and may thus degrade communication QoS of non-SMS. Therefore, it is important to determine when and to which WPs NetApp servers migrate considering both positive and negative impacts of server migration.

In this paper, in order to answer the key research question of when and to which WPs NetApp servers should migrate described in the paragraph above, we discuss the business models of SMS, introduce monetary metrics to measure the positive and negative financial impacts of SMS, and build a mixed-integer programming model. In the business models of SMS, an SMS provider receives service subscription fees from the SMS subscribers. An SMS provider pays SMS subscribers the monetary penalty (monetary penalty of SMS), when the agreed-upon NetApp QoS in SMS is violated. It also pays non-SMS subscribers (receiving non-SMS services sharing resources of the substrate network with SMS) the monetary penalty (monetary penalty of non-SMS), when the SMS traffic causes QoS degradation to non-SMS subscribers. An SMS provider also pays the utility company the energy cost of SMS equipment (routers and WPs). Our monetary metrics are based on the business models of SMS and consist of the operating cost of SMS (the sum of the monetary penalty of SMS to all SMS subscribers and the energy cost of SMS of all SMS equipment) and the monetary penalty of non-SMS. Using these monetary metrics, we build a mixed-integer programming model, and SMS providers determine, by solving the mixed-integer programming model, when and to which WPs NetApp servers migrate to minimize the operating cost of SMS (i.e., to maximize the positive impact of server migration), while keeping the monetary penalty of non-SMS (i.e., the negative impact of server migration) below an acceptable level.

The rest of the paper is organized as follows. Section 2 highlights the contributions of this paper. Section 3 explains various models employed in SMS, i.e., a NetApp model, a substrate network model, SMS equipment and its energy cost model, and business models and the monetary metrics. Section 4 describes how NetApp servers migrate and formulates it as a mixed-integer programming model. Section 5 presents numerical examples. Section 6 concludes the paper.

2 | CONTRIBUTIONS OF THIS PAPER

In this section, we first explain three major contributions of this paper in subsection 2.1. We then focus on one of the contributions, financial considerations, and discuss in depth how this paper differs from the existing studies in subsections 2.2 and 2.3.

2.1 | Contributions of This Paper

Contributions of this paper include the following.

- A new service to augment the current IaaS
  - This paper proposes SMS. SMS is a new service and augments the current IaaS [1,2,3]. SMS augments IaaS by (1) providing finer granularity of locations for NetApp servers to run on and (2) allowing NetApp servers to automatically and dynamically migrate as the NetApp clients change their locations.
- Financial considerations
  - This paper is among the first to address the monetary issues associated with QoS degradation in IaaS caused by migrating NetApp servers. In particular, this paper: (1) adopts monetary metrics (i.e., monetary penalty instead of throughput and delay, and energy cost of routers and computers instead of traffic load and processing load placed on routers and computers [1]) and (2) considers the negative financial impact of the traffic that migrating servers generate (the Server Migration traffic).
- Theoretical foundation
  - This paper lays a theoretical foundation, through providing a mixed-integer programming model [2] and obtaining the theoretical lower bound on the operating cost of SMS. SMS providers may use the theoretical foundation presented in this paper to set their pricing policies for the services they provide.

In the following subsection, we discuss in detail the financial considerations aspect of the contributions of this paper.

2.2 | Financial Considerations: Detailed Discussions

As discussed in the previous subsection, one of the contributions of this paper is that this paper is among the first to address the monetary issues associated with migrating NetApp servers through (1) adopting monetary metrics and (2) considering the negative financial impact of the traffic that migrating servers generate. In the following, we discuss in detail how this paper is original in these two aspects and different from the existing studies that address various research issues related to migrating servers, a key aspect of SMS. The existing studies we discuss in this subsection are taken from various research domains, including Cloud systems, edge computing systems [7] and fog computing systems [8], content delivery network (CDN), mobile agent systems, mobile crowdsensing, big data security and database migration. We first discuss how this paper is original compared to the existing studies in the research domains of Cloud systems, edge and fog computing systems. We then discuss how this paper is original compared to the existing studies in the research domains of CDN, mobile agent systems, mobile crowdsensing, big data security and database migration.

Table 1 shows that this paper is original in considering (1) monetary metrics and (2) the financial negative impact of the traffic that migrating servers generate and is different from the existing studies in the research domains of Cloud systems, edge computing systems [7] and fog computing systems [8]. Table 1 shows the research domain in the first column, provides a brief description of research issues addressed in the second column, and lists reference numbers of the corresponding existing studies. The fourth column of Table 1 shows if the monetary metrics are adopted or not in the corresponding existing studies, and the fifth column shows if the negative financial impact of the server migration traffic is considered or not.

As Table 1 shows, this paper is original in simultaneously considering both monetary metrics and negative financial impact of Server Migration traffic, unlike the existing studies in Cloud systems and edge and fog computing systems. Such existing studies often consider neither monetary metrics nor negative financial impact of Server Migration traffic and sometime only consider either monetary metrics or negative financial impact of Server Migration traffic, but not both. The following provides detailed discussions of the originality of this paper in these two aspects through illustrating how it is different from the existing studies in these aspects.

- This paper addresses both monetary metrics and negative financial impact of the Server Migration traffic; Some existing studies [9,10,11,21] in a cloud system, edge computing systems and fog computing systems only adopt monetary metrics without addressing negative financial impact of the Server Migration traffic. (Papers [9,10,11] in "Cloud systems" and Paper [21] in "Edge and fog computing systems" in Table 1.)

[1] In our previous paper [5,6], we only addressed the monetary penalty. In this paper, we address both the monetary penalty and energy cost of SMS.

[2] Our mixed-integer programming model belongs to a class of computationally NP-hard problems. A class of NP-hard problem is still solvable depending on the size of the problem and the values of the parameters.

- Paper[9] focuses on VM placement in a geo-distributed cloud system. It adopts two types of monetary metrics, the energy cost of running VMs on (physical) computers at data centers and the communication link cost of renting communication links from the wide area network provider, and develops a heuristic algorithm that achieves optimal trade-off between these two types of monetary metrics. Paper[9], however, only studies where to initially place servers, and unlike this paper, neither allows servers to migrate nor addresses the negative financial impact of the Server Migration traffic.
- Papers[10,11] focus on VM placement in a single data center in a cloud system. Papers[10,11] adopt two types of monetary metrics, the energy cost of physical computers at the data center and the monetary penalty of paying service subscription fees back to the subscribers when SLA (service level agreement) is violated, and develop algorithms to determine to which (physical) computers servers migrate within the single data center to minimize the sum of the energy cost and monetary penalty. Papers[10,11], however, only study a single data center, not networked data centers and do not address the negative financial impact of the Server Migration traffic.
- Paper[21] focuses on task scheduling in a fog computing system. It develops a framework called FogPlan for QoS-aware and dynamic fog service provisioning (QDFSP) in order to minimize the resource cost, while satisfying QoS requirement of delay-sensitive applications. It formulates QDFSP problem as an integer nonlinear programming model and proposes two greedy algorithms for solving the integer nonlinear programming model. Paper[21], however, does not address the negative financial impact of the Server Migration traffic.

- This paper addresses both monetary metrics and negative financial impact of the Server Migration traffic; Only one existing study[12] in a cloud system addresses the negative impact of the Server Migration traffic. Paper[12], however, does so in a traditional manner without adopting monetary metrics and without considering negative financial impact of the Server Migration traffic. (Paper[12] in "Cloud systems" in Table 1.) There are no existing studies in edge and fog computing systems addressing the negative (financial or non-financial) impact of the Server Migration traffic.
  - Paper[12] focuses on VM placement in a single data center in a cloud system. It adopts (non-monetary) negative impact of the Server Migration traffic in determining to which (physical) computers servers migrate within a single data center. It quantifies the negative impact of the Server Migration traffic as the product of the following three parameters: the memory size of the server, the page dirty ratio of the server (i.e., the ratio of the server memory pages that are modified per unit time to the total number of the server memory pages), and NetApp's sensitivity against server migration (i.e., how much NetApp's QoS degrades due to NetApp server migration). Paper[12], however, only considers a single data center, not networked data centers, and does not associate any monetary metric with the

TABLE 1 Novelty of this paper: Comparison with the existing studies.

Domain of research	Issues addressed	Papers	Monetary metrics	Negative financial impact of server migration traffic
SMS	Dynamically migrating servers	This paper	✓	✓
Cloud systems	Determining which physical computers/data centers to place VMs	9–11	✓	
		12		✓
		13–20		
Edge/Fog computing systems	Scheduling tasks to perform at edge computers (21–23), Allocation of CPU resources to tasks running on edge computers (24, 25), Determining which network node to place edge computers (26), Modeling power consumption of edge computers (27), Designing a distributed storage system (28)	21	✓	
		22–28		

negative impact of the Server Migration traffic. Consequently, cloud service providers may not use the contributions of paper [12] in order to determine to which (physical) computers servers migrate in optimizing their monetary benefits.

- This paper addresses both monetary metrics and negative financial impact of the Server Migration traffic; Many existing studies [13,14,15,16,17,18,19,20,22,23,24,25,26,27,28] in cloud systems, edge computing systems and fog computing systems address migrating servers considering neither the monetary metrics nor the negative financial impact of the Server Migration traffic. (Papers [13,14,15,16,17,18,19,20] in "Cloud systems" and Papers [22,23,24,25,26,27,28] in "Edge and fog computing systems" in Table 1.)

  - Paper [13] focuses on VM placement in a single data center in a cloud system. It adopts (non-monetary) energy consumption, defined as a function of the total CPU utilization of (physical) computers at the data center, and develops heuristic algorithms that determine to which (physical) computers servers migrate in order to minimize the total energy consumption at the data center, while satisfying SLA placed on servers (e.g., response time of servers). Paper [14] also focuses on VM placement in a single data center in a cloud system. It develops an algorithm that determines how many copies of the server to make and to which (physical) computer these copies of the server to place in order to minimize the total energy consumption of the data center, while satisfying SLA of servers. Paper [15] also focuses on VM placement in a single data center in a cloud system. It proposes an adaptive algorithm to reduce the number of VMs needed by an application, while satisfying the application's SLA and develops an optimal off-line algorithm. Paper [16] also focuses on VM placement in a single data center in a cloud system. It proposes an integer programming model that jointly minimizes the energy consumptions of physical servers and switches. Unlike this paper, papers [13,14,15,16] do not associate any monetary metric with the total energy consumption and application's QoS, and consequently, cloud service providers may not use the contributions of papers [13,14,15,16] in order to determine to which (physical) computers servers migrate in optimizing the monetary benefits. In addition, papers [13,14,15,16] neither address the negative financial impact of the Server Migration traffic nor consider servers migrating between multiple data centers.
  - Paper [17] focuses on VM placement between multiple data centers in a cloud system. It develops a mixed-integer programming model that determines to which data centers servers are placed in order to minimize the combined energy consumption of the data centers and of the network. Paper [18] also focuses on VM placement between multiple data centers in a cloud system. It proposes an energy and carbon-efficient VM placement algorithm to minimize the total carbon footprint. Paper [19] also focuses on VM placement between multiple data centers in a cloud system. It proposes 1) a data center selection algorithm for minimizing the maximum distance (i.e., latency) between the selected data centers, 2) a rack and physical computer selection algorithm and 3) a VM placement algorithm for reducing the inter-rack communications. Paper [20] focuses on the integrated control of VM placement and traffic engineering for a geo-distributed cloud system. It develops a mixed integer programming model for minimizing the average link delay. Papers [17,18,19,20], however, do not adopt the monetary metrics nor address the negative financial impact of the Server Migration traffic.
  - Paper [22] focuses on task scheduling in a mobile edge computing system. It proposes a heuristic algorithm for minimizing a combination of service response time and energy consumption. Paper [23] focuses on task scheduling in a fog computing system. It proposes a heuristic algorithm based on the moth-flame optimization algorithm in order to minimize task execution time, while satisfying the QoS requirements of applications. Paper [24] focuses on resource allocation for mobile augmented reality (MAR) applications in a multi-access edge computing system. It proposes an orchestration scheme to determine the frame resolution of images and to determine computing capacity at MAR devices in order to minimize a combination of energy consumption, latency and accuracy loss of MAR devices, while keeping the maximum latency constraint and the minimum accuracy constraint. Paper [25] focuses on resource allocation in an edge computing system. It develops an algorithm to dynamically adjust the number of container-based microservices in order to minimize the system cost, while satisfying the system stability. Paper [26] focuses on (physical) edge server deployment in an edge computing system. It proposes a greedy algorithm based on the minimum dominating set problem in order to minimize the number of edge servers, while satisfying QoS requirement of applications. Paper [27] focuses on power management of edge servers in an edge computing system. It proposes a linear power model to measure the energy consumption of edge servers. Paper [28] focuses on distributed storage system in an edge computing system. It proposes a distributed storage system architecture called EdgeKV, which

provides fault-tolerance, reliability, consistency, latency guarantee, and privacy. Papers[22,23,24,25,26,27,28], however, consider neither the monetary metrics nor the negative financial impact of Server Migration traffic.

As discussed earlier at the beginning of this subsection, this paper is original in (1) adopting monetary metrics and (2) considering the negative financial impact of the traffic that migrating servers generate and different from the existing studies that address research issues related to migrating servers, in CDN, mobile agent systems, mobile crowdsensing, big data security and database migration. Such existing studies address research issues such as load balancing of distributed servers in CDN[29,30], migration of mobile agents in mobile agent systems[31,32], participant selection in mobile crowdsensing[33], controlling the spread of virus in big data[34] and migration of databases in database migration[35]. This paper considers both monetary metrics and negative financial impact of the Server Migration traffic, while Paper[33] only considers monetary metrics without addressing negative financial impact of Server Migration traffic and Papers[29,30,31,32,34,35] consider neither the monetary metrics nor negative financial impact of Server Migration traffic.

2.3 | A Preliminary Version of This Paper

An earlier and preliminary version of this paper has appeared as a conference paper[36]. The current version of this paper significantly extends its earlier version as described below and presents a full and complete picture of our research on the subject of server migration in SMS.

- The current paper significantly extends its earlier version and (1) adds new investigation of a non-SMS business model, and (2) presents a mathematical model that is more general than that presented in the earlier version of this paper.
  - The non-SMS business model in the current paper enables us to consider the negative financial impact of migrating servers on the financial benefits of providing non-SMS service. Non-SMS business model was not discussed in the earlier version of this paper.
  - A mathematical model in the current paper considers various financial cost functions, making the model more general. They include the monetary penalty functions and QoS functions to take into account the monetary impact of Server-Client communication traffic, Server-Server communication traffic and non-SMS service traffic, as well as the energy cost functions and energy consumption functions to take into account the monetary impact of energy used by routers and WPs. The earlier version of this paper only presents a limited mathematical model and does not consider the monetary penalty functions and QoS functions for non-SMS service traffic. The limited mathematical model in the earlier version of the paper prohibited us from considering the negative financial impact of Server Migration traffic.
- The current paper adds a completely new set of numerical results based on the new mathematical model described in the current paper. The new numerical results include those using a new network (JPN25) and establish the effectiveness of the proposed SMS and mathematical model for practical networks.

3 | SERVER MIGRATION SERVICE (SMS)

In this section, we first explain a basic concept of SMS (Server Migration Service) in subsection 3.1. The next 3 subsections explain key components of SMS and their models assumed in this paper; network applications (NetApps) and the model for the NetApps in subsection 3.2, a substrate network that supports SMS and its model in subsection 3.3, and network equipment deployed in a substrate network to support SMS (SMS equipment) and its energy cost model in subsection 3.4. We also explain the business models that SMS providers use (SMS business models) in subsection 3.5 and discuss both positive and negative impacts of SMS in subsection 3.6.

3.1 | Outline of SMS

In order to augment the current IaaS and improve NetApp QoS, we propose SMS in this paper. Figure 1 depicts overview of SMS. SMS is a new class of service; it provides computing resources over a network (similarly to IaaS), and it does so, while maintaining NetApp QoS and minimizing the cost of providing SMS through allowing NetApp servers to migrate (unlike IaaS).

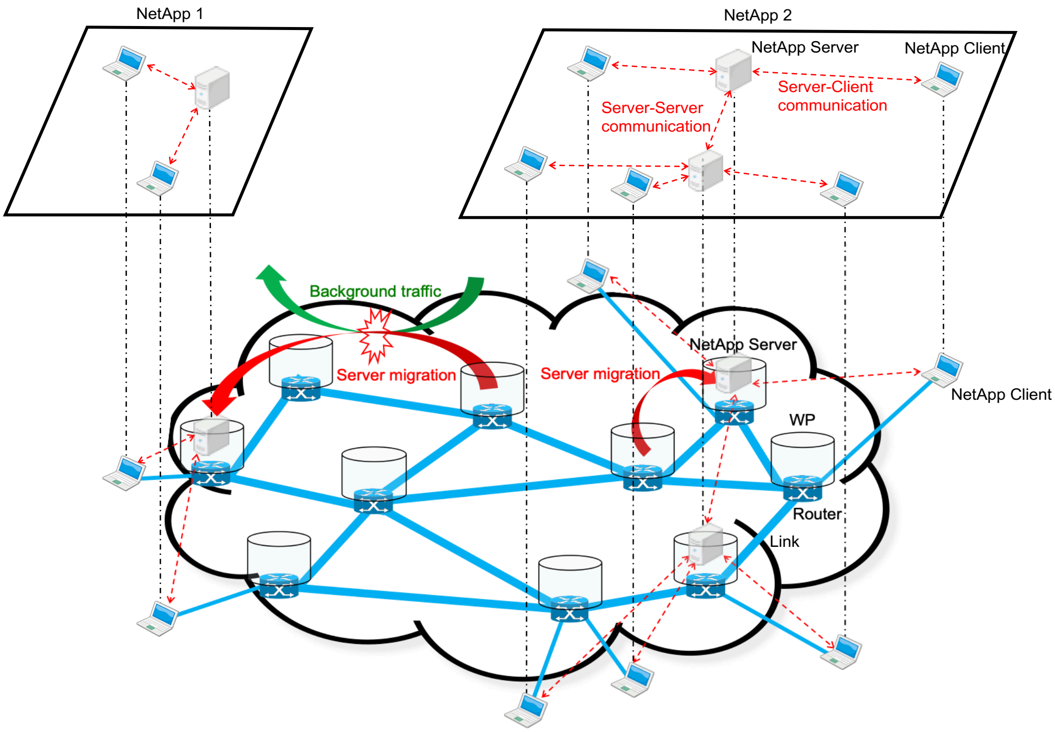


FIGURE 1 Overview of server migration service

SMS also differs from the existing IaaS in that (1) SMS provides finer granularity of locations for NetApp servers to run on and (2) SMS allows NetApp servers to automatically and dynamically migrate as the NetApp clients change their locations.

Similarly to IaaS, SMS provides its subscribers (NetApp providers) with computing resources (e.g., VMs) on demand over the substrate network upon which SMS is provided. In SMS, a WP (work place) is a micro data center owned by an SMS provider. Unlike data centers in IaaS, a WP hosts only a handful of (physical) computers and is significantly smaller in scale than IaaS data centers. An SMS provider deploys a number of WPs over the substrate network. A NetApp provider (an SMS subscriber) runs NetApp servers on VMs on (physical) computers at SMS provider's WP, and provide network applications to NetApp clients running on client terminals.

3.2 | NetApp (Network Application) Model in SMS

Figure 2 depicts the NetApp model assumed in SMS. It is the same model assumed for a NetApp in our previous paper[6]. A NetApp consists of one or more NetApp servers S_i ($i = 1, 2, \cdots, n$) and one or more NetApp clients C_i^j ($i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m_i$), where n denotes the total number of NetApp servers, and m_i denotes the total number of NetApp clients that NetApp server S_i serves, and C_i^j denotes the j-th NetApp client that the i-th NetApp server serves. A VM[3] on a (physical) computer at a WP runs a NetApp server, while a client terminal, for instance a smartphone, a tablet and a laptop runs a NetApp client.

- A NetApp client communicates with its NetApp server and requests/receives a service. A NetApp server communicates with its NetApp clients and provides a service. The Server-Client communication refers to such communication between a NetApp server and a NetApp client.
- A NetApp server communicates with other NetApp servers of the same network application (NetApp) and provides the service that a NetApp client requests. The Server-Server communication refers to such communication between NetApp servers.

[3] We assume that size of VMs are on the order of megabytes or gigabytes. We do not consider the case where VMs include big data whose size are on the order of terabytes or petabytes and server migration time/cost are enormous.

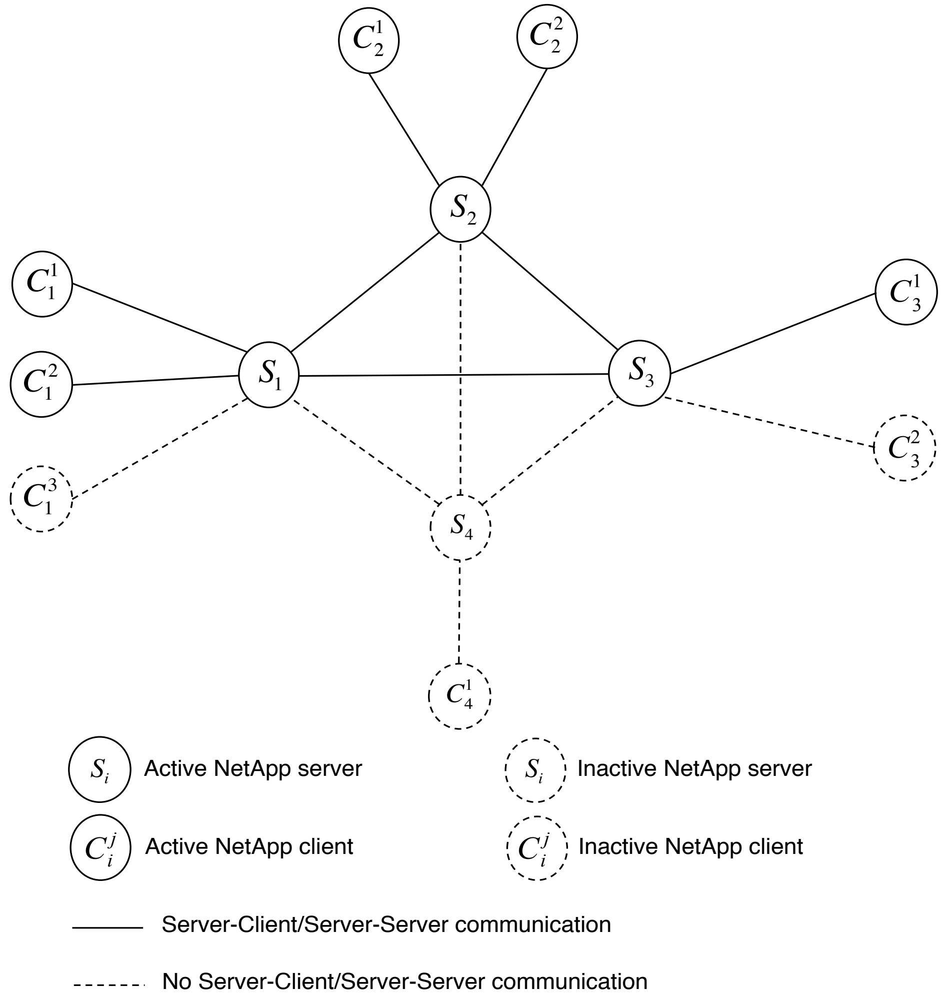


FIGURE 2 Network application model

A NetApp server and a NetApp client is either active (i.e., awaken and ready to communicate) or inactive (i.e., not running and not ready to communicate). We assume the following.

- A NetApp client dynamically changes from active to inactive and from inactive to active.
- A NetApp server is always active, when one or more of its NetApp clients are active.
- Server-Client communication occurs only when a NetApp server and a NetApp client are both active. In Figure 2, NetApp clients C_1^1, C_1^2, C_2^1, C_2^2, and C_3^1 are active, and these active NetApp clients communicate with their corresponding NetApp servers. NetApp clients C_1^3, C_3^2, and C_4^1 are inactive, and there is no Server-Client communication between these inactive NetApp clients and their corresponding NetApp servers.
- Server-Server communication occurs only when both NetApp servers are active. In Figure 2, NetApp servers S_1, S_2, and S_3 are active, and these active NetApp servers communicate only between themselves. NetApp server S_4 is inactive, and there is no Server-Server communication between the inactive NetApp server S_4 and active NetApp servers S_1, S_2, and S_3.

We assume that a WP has a limited computing resources and runs a small number of active NetApp servers. The capacity of a WP refers to the total amount of the computing resources that the WP has.

There are three traffic types generated by a NetApp: (1) the Server-Client traffic (that Server-Client communication generates), (2) the Server-Server traffic (that Server-Server communication generates), and (3) the Server Migration traffic (that NetApp server migration generates).

- The type of a NetApp determines the volume of the Server-Client traffic and that of the Server-Server traffic. For instance, a real-time application of an online multiplayer game typically generates a large volume (around 10 Mbps) of the Server-Client traffic [37], as a NetApp server in such an application sends screen images to its NetApp clients on a real-time basis. On the other hand, an application of a strategy game typically generates a small volume (around 2 to 3 kbps) of the Server-Client traffic [38].
- The Server Migration traffic typically contains significant volume of traffic. Its traffic volume may reach several hundreds of Mbps [39], as a VM that runs a NetApp server may also migrate with the NetApp server in order to run the NetApp server at a new WP.

3.3 | Substrate Network Model

SMS uses resources of the substrate network upon which SMS is provided. The substrate network spans over a wide geographical area, and WPs are distributed over the substrate network. In this paper, we use the substrate network model described below. It is the same model assumed in our previous paper [6].

SMS shares the resources of the substrate network with the non-SMS service and competes with the non-SMS service for the resources of the substrate network. We assume that both SMS and non-SMS service are provided by a single provider, who owns resources of the substrate network (single provider model). Namely, a single provider acts as the SMS provider and provides SMS to its SMS subscribers using the resources of the substrate network, and at the same time, it acts as the non-SMS service provider and provides non-SMS service to its non-SMS service subscribers sharing the same resources of the substrate network as those used to provide SMS.

When SMS competes with the non-SMS service for the resources of the substrate network, we assume that the SMS traffic (that SMS generates) has higher priority over the non-SMS service traffic (that non-SMS service generates). The SMS traffic has higher priority, because of the interactive and real-time nature of the applications that SMS supports. It is notable that, because the SMS traffic has higher priority over the non-SMS service traffic, the SMS traffic may degrade QoS of the non-SMS service, while the non-SMS service traffic does not degrade QoS of SMS.

3.4 | SMS Equipment and its Energy Cost Model in SMS

The substrate network in SMS deploys a number of routers and a number of WPs. These routers and WPs are collectively referred to as the SMS equipment in this paper. For the simplicity of discussions, we assume without losing a generality that a WP only consists of a single (physical) computer in this paper. An SMS provider runs SMS equipment in order to provide SMS and pays the utility company energy cost of running SMS equipment based on how much electricity its SMS equipment consumes.

In this paper, we assume the following energy consumption model and the energy cost model. The energy consumption model determines relationship between the traffic load of a router and the energy consumption of the router and relationship between the processing load of a WP and the energy consumption of the WP. The energy cost model determines relationship between the energy consumption of a router and the energy cost of the router and relationship between the energy consumption of a WP and the energy cost of the WP.

For a router, we assume the following energy consumption model and the energy cost model.

- The energy cost of a router depends on the energy consumption of the router. The utility company uses its proprietary information and determines the energy cost of a router based on its energy consumption.
- The energy consumption of a router is equal to the sum of the energy consumption of each line-card in the router.
- The energy consumption of a line-card depends on the traffic load of the input link connected to the line-card.
  - When a line-card is inactive (i.e., when there is no traffic load on a line-card), the line-card goes into a sleep-mode and consumes no or little energy.
  - When a line-card is active (i.e., when there is non-zero traffic load on a line-card), how much energy the line-card consumes depends on the type of the line-card: an energy-efficient line-card adjusts its energy consumption based on its traffic load, while a non energy-efficient line-card consumes a constant amount of energy regardless of its traffic load.
- The traffic load of input link e connected to line-card e (the traffic load of link e), u_e, is defined as follows:

$$u_e = t_e^{all} / B_e \tag{1}$$

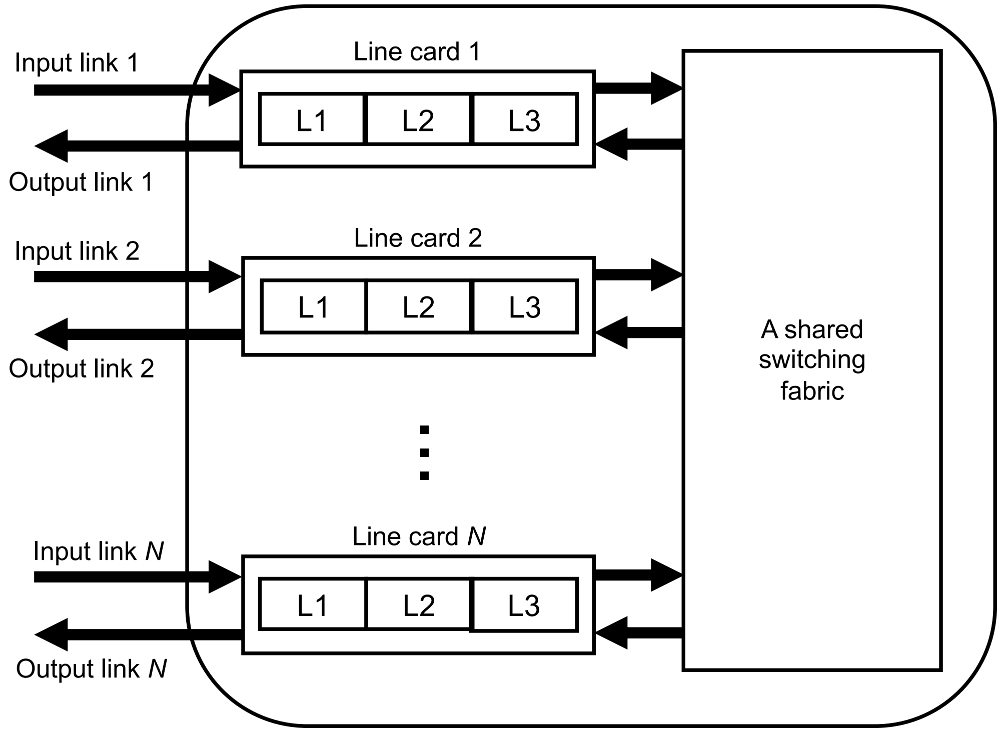


FIGURE 3 An example router architecture (a distributed architecture)

where t_e^{all} is the sum of the transmission rates of all traffic that traverses input link e, and B_e is the capacity of input link e.

The energy consumption model and the energy cost model described above for a router are reasonable and justified below.

- The assumption of the energy consumption of a router being determined by its line-cards and becoming the sum of the energy consumption of each line card is justified by the following. Many existing routers use a distributed architecture[40]. In this architecture (Figure 3), a shared switching fabric connects multiple line-cards, each of which has a module for processing layers 1-3 functionality, namely, line-cards conduct most of the processing required at the router. In this router architecture, line-cards thus dominate the energy consumption of the router. For instance, for a 320 Gbps router of this architecture, line-cards accounts for approximately 91% of the energy consumption of the router[40].
- The assumption of the line-card energy consumption depending on the traffic load of the input link connected to the line-card is justified, because a module for processing layers 1-3 functionality in a line-card is mainly devoted to processing the incoming traffic load (packets) on the input link.[4]
- The assumption of the energy-efficient line-card adjusting its energy consumption based on its traffic load is justified, because the energy-efficient line-card has rate adaptation capability, i.e., the line-card adjusts its energy consumption based on its traffic load by dynamically adjusting the voltage and frequency of its module for processing layers 1-3 functionality[41].

For a WP, we assume the following energy consumption model and the energy cost model.

- The energy cost of a WP depends on the energy consumption of the WP. The utility company uses its proprietary information and determines the energy cost of a WP based on its energy consumption.
- The energy consumption of a WP depends on the processing load of the WP.
  - When a WP is inactive (i.e., when there is no processing load on a WP), the WP goes into a sleep-mode and consumes no or little energy.

[4]The module for processing layers 1-3 functionality in a line-card mostly performs processing of incoming packets. Upon receiving a packet on the input link, this module processes the packet and determines to which output link it should be forwarded, while, upon receiving a packet from the shared switching fabric, it simply forwards the packet to its output link.

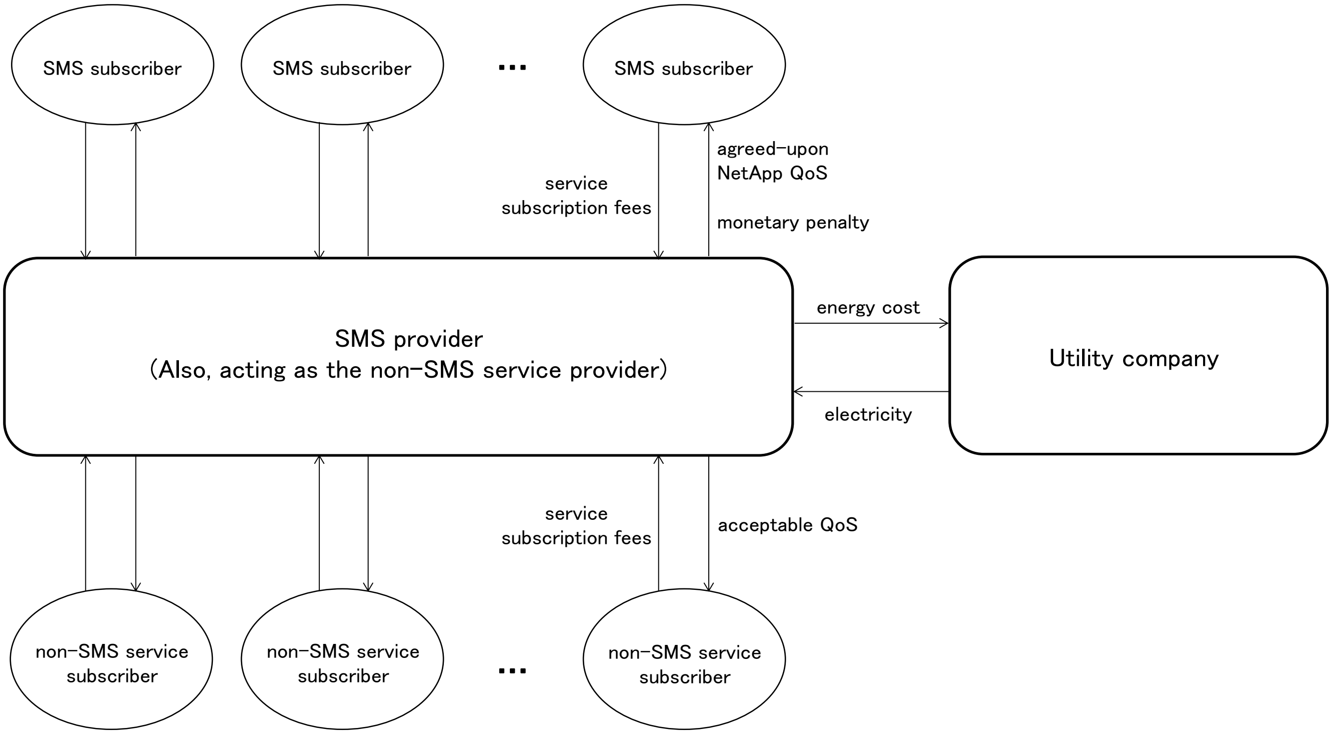


FIGURE 4 SMS business models

 – When a WP is active (i.e., when there is non-zero processing load on a WP), how much energy the WP consumes depends on the type of the WP: an energy-efficient WP adjusts its energy consumption based on its processing load, while a non energy-efficient WP consumes a constant amount of energy regardless of its processing load.

- The processing load of WP l, u_l, is defined as follows:

$$u_l = v_l^{all}/C_l \tag{2}$$

where v_l^{all} is the sum of the amount of computing resources required by each active NetApp server running on WP l, and C_l is the capacity of WP l (i.e., the total amount of computing resources available at WP l).

The assumption of an energy-efficient WP adjusting its energy consumption based on its processing load is justified because, similarly to an energy-efficient line-card of a router, the energy-efficient WP has rate adaptation capability, i.e., the WP adapts its energy consumption based on its processing load by dynamically adjusting the voltage and frequency of its CPU[42]

3.5 | SMS Business Models

SMS provider and SMS subscribers

Figure 4 includes the business model between an SMS provider and SMS subscribers. An SMS provider and SMS subscribers agree on the business model described below.

- An SMS provider provides SMS to its subscribers and tries to provide the agreed-upon NetApp QoS in Server-Client and Server-Server communications.
- In return, the SMS provider receives service subscription fees from the SMS subscribers.
- When the SMS provider violates the agreed-upon NetApp QoS, it pays SMS subscribers the monetary penalty. This monetary penalty of SMS is proportional to the degree of NetApp QoS violation; when the degree is higher, the monetary penalty is larger.

SMS provider and utility company

Figure 4 includes the business model between an SMS provider and the utility company. An SMS provider and the utility company agree on the business model described below.

- The energy cost of SMS depends on the energy consumption of the SMS equipment (routers and WPs). The utility company uses its proprietary energy cost model (described in subsection 3.4) and determines the energy cost of SMS from the energy consumption of the SMS equipment.
- The SMS provider pays the utility company the energy cost.

SMS provider's total net income

With the business models described above, SMS provider's total revenue, total operating cost and net income become the following:

- The total revenue that an SMS provider acquires from SMS is the sum of service subscription fees from its SMS subscribers.
- The total operating cost of SMS that an SMS provider incurs is the sum of the monetary penalties to SMS subscribers (the total monetary penalty of SMS) and the sum of the energy costs to the utility company to run SMS equipment (the total energy cost of SMS).
- The total net income that an SMS provider acquires from SMS is the total revenue from SMS minus the total operating cost of SMS.

We assume that the number of SMS subscribers is constant (i.e., the total revenue from SMS is constant). As a result, we can maximize the total net income from SMS by minimizing the total operating cost of SMS.

SMS provider and non-SMS service subscribers

As explained in subsection 3.3, SMS share the resources of the substrate network with the non-SMS service and competes with the non-SMS service for the resources of the substrate network. Thus, the SMS traffic may degrade QoS of the non-SMS service. When this happens, some non-SMS service subscribers may move to a different non-SMS service provider seeking better QoS. As a result, the non-SMS service provider loses some of its revenue from non-SMS service and decreases its net income. This loss in the revenue and in the net income is referred to as the monetary penalty of the non-SMS service.

Figure 4 includes the business model between a non-SMS service provider and non-SMS service subscribers. A non-SMS service provider and non-SMS service subscribers agree on the business model described below.

- A non-SMS service provider provides non-SMS service to its subscribers.
- In return, the non-SMS service provider receives service subscription fees from the non-SMS service subscribers.
- When QoS of the non-SMS service degrades due to the SMS traffic, the non-SMS service provider incurs the monetary penalty of the non-SMS service. This monetary penalty is proportional to the degree of QoS degradation of the non-SMS service; when the degree is higher, the monetary penalty is larger.

3.6 | NetApp Server Migration in SMS

Server migration in SMS

NetApp servers in SMS automatically and dynamically migrate among WPs in the network (1) to optimally locate themselves in relation to NetApp clients and minimize the total monetary penalty of SMS due to NetApp QoS degradation caused by location-related factors (i.e., propagation delays on network links) and (2) to optimally distribute traffic load over routers (SMS equipment) and processing load over WPs (SMS equipment) and minimize the total energy cost of SMS due to the energy consumption of SMS equipment.

Positive impact of NetApp server migration on the monetary penalty of SMS

When NetApp servers migrate, they may decrease QoS degradation in Server-Client and Server-Server communications, and consequently, they may decrease the total monetary penalty of SMS.

- When a NetApp client dynamically changes its location and moves away from its NetApp server, the propagation delay from the NetApp client to its NetApp server increases, and, as a result, QoS in Server-Client communication degrades. In such cases, the NetApp server may migrate towards the NetApp client and prevent QoS from degrading.
- When some NetApp servers dynamically migrate and move away from the other NetApp servers, the propagation delays from the migrating NetApp servers to the other NetApp servers increase, and, as a result, QoS in Server-Server communication degrades. In such cases, the other NetApp servers may migrate towards the migrating NetApp servers and prevent QoS from degrading.

It is notable that, when NetApp servers migrate and improve QoS in Server-Client communication, they may not necessary improve QoS in Server-Server communication, and vice versa. It is important to balance QoS in both Server-Client communication and Server-Server communication and determine when and to which WPs NetApp servers migrate.

Positive impact of NetApp server migration on the energy cost of SMS

When NetApp servers migrate, they may decrease the total energy consumption of routers and WPs (SMS equipment), and consequently, they may decrease the total energy cost of SMS as explained below.

NetApp servers migrate and decrease the total energy consumption of routers in the following manner. A major factor that determines the total energy consumption of routers is the distribution of the traffic load over routers, i.e., how the traffic load that NetApps generate (i.e., Server-Client, Server-Server, and Server Migration traffics) is distributed over different routers in the network. Thus, the total energy consumption of routers improves, when NetApp servers strategically migrate and optimally adjust the distribution of the traffic load over routers. For instance,

- Under some circumstances, distributing traffic load over a larger number of routers each supporting less traffic load may result in a smaller total energy consumption of routers. When the energy consumption of a router exponentially increases with the traffic load it supports (for instance, because the energy consumption of cooling a router non-linearly increases with the traffic load it supports), the energy consumption of each router supporting high traffic load dominates the total energy consumption of routers. Under such circumstances, it would decrease the total energy consumption of routers, when NetApp servers migrate to WPs such that the traffic load generated by NetApps is shifted from routers supporting higher traffic load to routers supporting smaller traffic load.
- Under some circumstances, distributing traffic load over a smaller number of routers each supporting larger traffic load may result in a smaller total energy consumption of routers. When the energy consumption of a router is constant and independent of the traffic load it supports (for instance, because all routers are non energy-efficient), A major factor that determines the total energy consumption of routers is how many routers are active and supporting traffic load, not how much traffic load routers support. Under such circumstances, it would decrease the total energy consumption of routers, when NetApp servers migrate to WPs such that the traffic load generated by NetApps is shifted from routers supporting smaller traffic load to routers supporting higher traffic load. This reduces the total number of currently active routers by utilizing each active router to its full capacity.

The same discussion can be applied to the total energy consumption of WPs.

Negative impact of NetApp server migration on the monetary penalty of non-SMS

When NetApp servers in SMS migrate among WPs, the SMS provider may decrease the total net income from non-SMS service. This is because SMS shares the resources of the substrate network with the non-SMS service. When NetApp servers migrate, they may generate a large volume of traffic (e.g., several hundred Mbps[39]) and occupy a significant portion of the shared resources of the substrate network. In addition, after migrating to new WPs, NetApp servers generate new Server-Client and Server-Server traffics, which may also occupy a significant portion of the shared resources of the substrate network. These penalize QoS of the non-SMS service, and as a result, decrease the total net income of the non-SMS service provider from non-SMS service. It is important that the SMS provider determines when and to which WPs NetApp servers migrate to minimize the total operating cost of SMS, while preventing NetApp servers from excessively migrating and adversely impacting QoS of the non-SMS service that share the resources of the substrate network with SMS.

4 | MIXED-INTEGER PROGRAMMING MODEL FOR ENERGY-AWARE NETAPP SERVER MIGRATION

In this section, we develop a new mixed-integer programming model, considering the total monetary penalty and total energy cost of SMS. We first explain the assumptions employed in the mixed-integer programming model in subsection 4.1. We then explain the objective function and constraints used in the mixed-integer programming model and present a mathematical formulation of the mixed-integer programming model in subsection 4.2.

Our mixed-integer programming model is based on the integer programming model developed without considering the total energy cost in our previous paper[5,6]. Using the newly developed mixed-integer programming model, we determine when and to which WPs NetApp servers migrate in order to minimize the *total operating cost* of SMS (i.e., a sum of the total monetary penalty and total energy cost of SMS), while keeping the monetary penalty of the non-SMS service at an acceptable level.

4.1 | Model Assumptions

A mixed-integer programming model we develop employs the following assumptions.

- Time is slotted.
- Changes in the active/inactive state and locations of NetApp clients occur only at slot boundaries.
- Transmission rates of all types of SMS traffic (i.e., Server-Client, Server-Server, and Server Migration traffics) and the non-SMS service traffic remain constant during a slot.
- A NetApp server starts its migration at the beginning of a slot and completes its migration at the end of the same slot. During the slot, the NetApp server continues to perform its service at the current WP, and it starts performing its service at the new WP in the slot following its migration.
- We consider the time period of length T slots (slot 1, slot 2, $\cdots$, slot T), referred to as the time period of interest, in deriving the mixed-integer programming model. The time instance at the boundary between slot t-1 and slot t is referred to as time t.

Economy policies of the SMS provider dictate the length of a time slot, and a typical length of a time slot may be in the order of a few hours to a few days.

Similarly to the integer programming model in our previous paper[5,6], our mixed-integer programming model produces a list of WPs showing at which WP each NetApp server is located in each slot in the time period of interest. Figure 5 shows an example of energy-aware NetApp server migration problem. Our mixed-integer programming model assumes that all necessary information (e.g., at which WPs NetApp servers were initially located, where NetApp clients are located in each slot in the time period of interest, which NetApp clients and NetApp servers communicate in each slot in the time period of interest) is available when solving the model and that the SMS provider solves the model off-line once using a high-performance computer before the time period of interest starts. It is notable that, as all necessary information during the time period of interest is available in solving the model, our mixed-integer programming model finds the optimal WPs for NetApp servers and minimizes the total operating cost of SMS; our model obtains the optimal WPs that achieves the theoretical lower bound of the total operating cost of SMS.

4.2 | Mixed-integer Programming Model

This section develops a mixed-integer programming model. Our model determines when and to which WPs NetApp servers migrate in order to minimize the *total operating cost* of SMS (i.e., a sum of the total monetary penalty and total energy cost of SMS), while keeping the monetary penalty of the non-SMS service at an acceptable level during a window of W successive slots. This window is referred to as the smoothing window and avoids the negative impact of the SMS traffic (i.e., Server-Client, Server-Server, and Server-Migration traffics) from concentrating on a single slot[6]. Figure 5 includes an example of the smoothing windows.

Objective function

The mixed-integer programming model minimizes its objective function of the total operating cost of SMS (i.e., a sum of the total monetary penalty of SMS and the total energy cost of SMS) in the time period of interest (slot 1 through slot T). Components of the total operating costs of SMS are defined below.

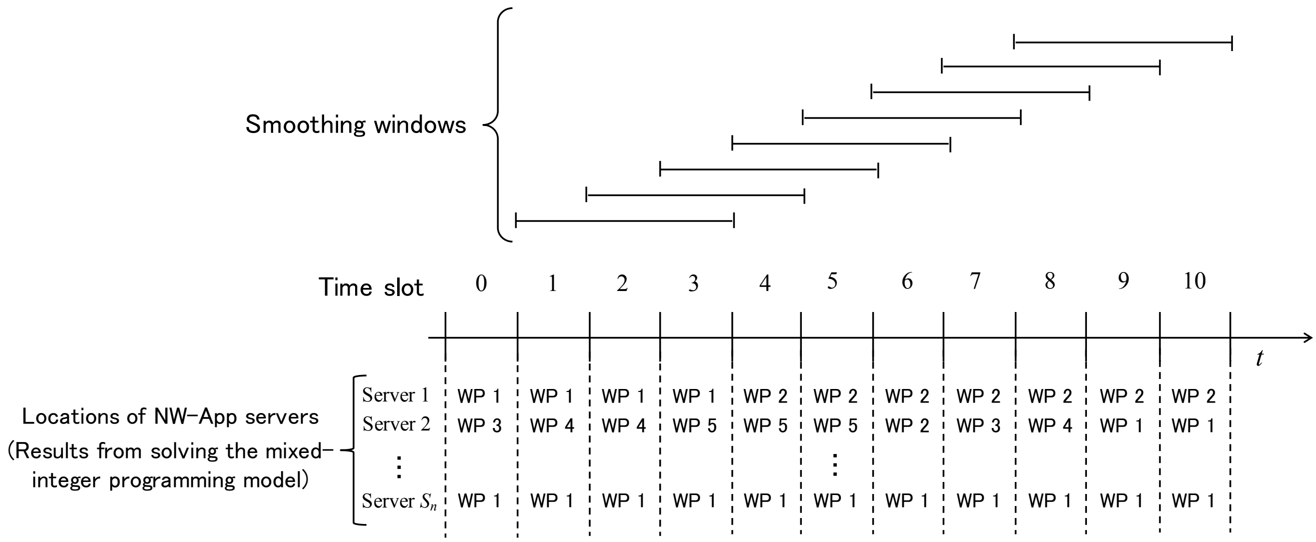


FIGURE 5 An example of energy-aware NetApp server migration in a discrete-time domain

- The total monetary penalty of SMS and the total energy cost of SMS in the time period of interest are the sum of the monetary penalties that the SMS provider pays to its SMS subscribers [5] and the sum of the energy costs that the SMS provider pays to the utility company, respectively, in each slot in the time period of interest (slot 1 through slot T).

The mixed-integer programming model uses an SMS monetary penalty function to calculate the monetary penalty of SMS. For a given NetApp QoS, this function calculates the value of the monetary penalty that the SMS provider pays to its SMS subscribers. Our model assumes the SMS monetary penalty function of the following two types.

- $MP^{SC}_{SMS}(CQ^{SC}_{SMS}(s,l,r,l',t))$ is the SMS monetary penalty function that calculates the monetary penalty of SMS due to NetApp QoS degradation in Server-Client communication, where $CQ^{SC}_{SMS}(s,l,r,l',t)$ is the QoS function that calculates QoS in Server-Client communication between NetApp server s at WP l and NetApp client r at router l' in slot t.
- $MP^{SS}_{SMS}(CQ^{SS}_{SMS}(s,l,s',l',t))$ is the SMS monetary penalty function that calculates the monetary penalty of SMS due to NetApp QoS degradation in Server-Server communication, where $CQ^{SS}_{SMS}(s,l,s',l',t)$ is the QoS function that calculates QoS in Server-Server communication between NetApp server s at WP l and NetApp server s' at WP l' in slot t.

The SMS monetary penalty function and its associated QoS functions may take any forms in the mixed-integer programming model. Each SMS provider has its proprietary financial information and may decide specific function forms based on such proprietary information. Each SMS provider may also adopt its own metrics for QoS (e.g., communication delay, monthly uptime percentage of VMs[43,44]) and measure QoS based on the adopted metrics. Different SMS providers are likely to use different forms of the monetary penalty and QoS functions[43,44].

The mixed-integer programming model uses an SMS energy cost function to calculate the energy cost of SMS. For a given traffic load of routers and a given processing load of WPs, this function calculates the value of the energy cost that the SMS provider pays to the utility company. Our model assumes the SMS energy cost function of the following two types.

- $ECT_{RT}(ECS_{RT}(e,u,t))$ is the SMS energy cost function that calculates the energy cost of a router due to the line-card (of the router) attached to input link e, where $ECS_{RT}(e,u,t)$ is the SMS energy consumption function that calculates the energy consumption of the line-card attached to input link e of a router, when traffic load of the line-card is u in slot t.
- $ECT_{WP}(ECS_{WP}(l,u,t))$ is the SMS energy cost function that calculates the energy cost of a WP, where $ECS_{WP}(l,u,t)$ is the SMS energy consumption function that calculates the energy consumption of WP l, when processing load of WP l is u in slot t.

[5] It is worth noting that our model minimizes the total monetary penalty of SMS by improving the overall average NetApp QoS from the SMS provider's perspective, which may result in degradation in NetApp QoS for some NetApp clients.

Similarly to the SMS monetary penalty function and its associated QoS functions, the SMS energy cost function and its associated energy consumption functions may take any forms in the mixed-integer programming model. Each SMS provider may decide specific function forms based on its own proprietary financial information and its own metrics for QoS.

Constraints

The mixed-integer programming model minimizes the total operating cost of SMS under the constraints that the total monetary penalty of the non-SMS service is below a given threshold, U, in a smoothing window.

The mixed-integer programming model uses a non-SMS service monetary penalty function to calculate the monetary penalty of the non-SMS service. For a given QoS in the non-SMS service, this function calculates the value of the monetary penalty that the non-SMS service provider pays to its subscribers. Our model assumes the following type of the non-SMS service monetary penalty function.

- $MP_{NSMS}(CQ_{NSMS}(\mathbf{T^t_{SC}}, \mathbf{T^t_{SS}}, \mathbf{T^t_{SM}}))$ calculates the monetary penalty of the non-SMS service, where $CQ_{NSMS}(\mathbf{T^t_{SC}}, \mathbf{T^t_{SS}}, \mathbf{T^t_{SM}})$ is the QoS function that calculates QoS of the non-SMS service, when the traffic volumes of three types of SMS traffic (Server-Client, Server-Server, and Server Migration traffics) in slot t are given by traffic matrices $\mathbf{T^t_{SC}}$, $\mathbf{T^t_{SS}}$, and $\mathbf{T^t_{SM}}$, respectively. Elements $d^{l,l',t}_{SC}$ in $\mathbf{T^t_{SC}}$, $d^{l,l',t}_{SS}$ in $\mathbf{T^t_{SS}}$, and $d^{l,l',t}_{SM}$ in $\mathbf{T^t_{SM}}$ represent the transmission rate of Server-Client traffic, Server-Server traffic, and Server Migration traffic from location l (WP or router) to location l' in slot t, respectively.

Similarly to the SMS monetary penalty function and its associated QoS functions, the non-SMS service monetary penalty function and its associated QoS function may take any forms in the mixed-integer programming model. Each SMS provider may decide specific function forms based on its own proprietary financial information and its own metrics for QoS.

Mathematical expressions

The following presents a mathematical formulation of the mixed-integer programming model. Tables 2 and 3 show notations and decision variables used in the mixed-integer programming model, respectively.

- Objective function: Minimize the total operating cost of SMS.

$$\begin{aligned}\text{minimize} \quad \sum_{t\in\mathbb{T}} \Bigg\{ & \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} \sum_{l\in\mathbb{L}} \sum_{r\in\mathbb{R}^t_s} MP^{SC}_{SMS}(CQ^{SC}_{SMS}(s, l, r, u(r), t))\ o^t_s v^t_{s,l} \\ & + \sum_{n\in\mathbb{N}} \sum_{s,s'\in\mathbb{S}_n} \sum_{l,l'\in\mathbb{L}} MP^{SS}_{SMS}(CQ^{SS}_{SMS}(s, l, s', l', t))\ o^t_s v^t_{s,l} o^t_{s'} v^t_{s',l'} \\ & + \sum_{e\in\mathbb{E}} ECT_{RT}(ECS_{RT}(e, u^t_e, t)) + \sum_{l\in\mathbb{L}} ECT_{WP}(ECS_{WP}(l, u^t_l, t)) \Bigg\}\end{aligned} \tag{3}$$

  Eq.(3) is the total operating cost of SMS. It is the objective function of the mixed-integer programming model and is to be minimized. It consists of four terms: (1) the monetary penalty of SMS due to QoS degradation in Server-Client communication, (2) the monetary penalty of SMS due to QoS degradation in Server-Server communication, (3) the energy cost of SMS due to energy consumption of routers, and (4) the energy cost of SMS due to energy consumption of WPs.

- Constraint (1): The total monetary penalty of the non-SMS service in the smoothing window is less than or equal to a given threshold U.

$$\sum_{t\in w} MP_{NSMS}(CQ_{NSMS}(\mathbf{T^t_{SC}}, \mathbf{T^t_{SS}}, \mathbf{T^t_{SM}})) \leq U \quad \forall w \in \mathbb{W} \tag{4}$$

  The left hand side of Eq.(4) is the total monetary penalty of the non-SMS service over a smoothing window. This total monetary penalty is calculated as the sum of the monetary penalties of the non-SMS service incurred in every slot of the smoothing window. Eq.(4) says that this total monetary penalty must be less than or equal to the threshold (U).

- Constraint (2): All NetApp servers reside on WPs.

$$\sum_{l\in\mathbb{L}} v^t_{s,l} \leq 1 \qquad \forall s \in \mathbb{S}_n, \forall n \in \mathbb{N}, \forall t \in \mathbb{T} \tag{5}$$

TABLE 2 A list of notations.

Symbol	Meaning
$\mathbb{T}$	The time period of interest (slot 1 through slot T).
$\mathbb{N}$	A set of all NetApps in the network.
$\mathbb{S}_n$	A set of all NetApp servers (NetApp server 1 through S_n) of NetApp n. S_n is the total number of NetApp servers in NetApp n.
o_s^t	A variable that represents the state of a NetApp server s in slot t. $o_s^t = 1$, if NetApp server s is active in slot t, or $o_s^t = 0$, otherwise.
x_s	Amount of the computing resources of a WP that an active NetApp server s requires.
C_l	Capacity of WP l (i.e., the total amount of computing resources available at WP l).
$\mathbb{L}$	A set of all WPs (WP 1 through L) in the network. L is the total number of WPs in the network.
F_s	WP where NetApp server s is initially located.
$\mathbb{R}_s^t$	A set of active NetApp clients that communicate with NetApp server s in slot t.
$u(r)$	Location of NetApp client r.
$\mathbb{E}$	A set of all links in the substrate network.
B_e	Capacity of link e [Gbps].
$t_{SC}^{s,r,t}$	Transmission rate of Server-Client traffic between NetApp server s and client r in slot t [Gbps].
$t_{SS}^{s,s',t}$	Transmission rate of Server-Server traffic between NetApp servers s and s' in slot t [Gbps].
$t_{SM}^{s,t}$	Transmission rate of Server Migration traffic that migrating NetApp server s generates in slot t [Gbps].
$t_{NSMS}^{e,t}$	Transmission rate of the non-SMS service traffic on link e in slot t [Gbps].
$R_{l,l'}^e$	A variable to indicate whether a route includes a given link or not. $R_{l,l'}^e = 1$, if the traffic from location (WP or router) l to location l' flows through link e, or $R_{l,l'}^e = 0$, otherwise.
$\mathbb{W}$	A set of all smoothing windows of length p (slots) in the time period of interest (= $\{\{1, 2, \cdots, p\}, \{2, 3, \cdots, p+1\}, \cdots, \{T-p+1, T-p+2, \cdots, T\}\}$.
U	Threshold on the total monetary penalty of the non-SMS service in the smoothing window.
$\mathbf{T_{SC}^t}$	Traffic matrix of Server-Client traffic. Its element $d_{SC}^{l,l',t}$ represents the transmission rate of the Server-Client traffic from location l to location l' in slot t.
$\mathbf{T_{SS}^t}$	Traffic matrix of Server-Server traffic. Its element $d_{SS}^{l,l',t}$ represents the transmission rate of the Server-Server traffic from location l to location l' in slot t.
$\mathbf{T_{SM}^t}$	Traffic matrix of Server Migration traffic. Its element $d_{SM}^{l,l',t}$ represents the transmission rate of the Server Migration traffic from location l to location l' in slot t.

The left hand side of Eq.(5) is the number of WPs where a NetApp server resides. Eq.(5) says that this number must be at most one.

- Constraint (3): The total amount of the computing resources required by the active NetApp servers on a WP is less than or equal to the capacity of the WP.

$$\sum_{n\in\mathbb{N}}\sum_{s\in\mathbb{S}_n} x_s o_s^t v_{s,l}^t \le C_l \quad \forall l \in \mathbb{L}, \forall t \in \mathbb{T} \tag{6}$$

The left hand side of Eq.(6) is the total amount of the computing resources required by the active NetApp servers on a WP. Eq.(6) says that this number must be less than or equal to the capacity of the WP (C_l).

TABLE 3 Decision variable.

Symbol	Meaning
$v^t_{s,l}$	A variable to indicate whether NetApp server s is at WP l in slot t. $v^t_{s,l} = 1$, if NetApp server s is at WP l in slot t, or $v^t_{s,l} = 0$, otherwise.
u^t_e	Traffic load of link e in slot t
u^t_l	Processing load of WP l in slot t

- Constraint (4): The traffic load of a link is equal to the utilization of the link in a given slot.

$$u^t_e = \left\{ \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} \sum_{l\in\mathbb{L}} \sum_{r\in\mathbb{R}^t_s} t^{s,r,t}_{SC} v^t_{s,l} (R^e_{u(r),l} + R^e_{l,u(r)}) + \sum_{n\in\mathbb{N}} \sum_{s,s'\in\mathbb{S}_n} \sum_{l,l'\in\mathbb{L}} t^{s,s',t}_{SS} o^t_s v^t_{s,l} o^t_{s'} v^t_{s',l'} R^e_{l,l'} \right.$$
$$\left. + \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} \sum_{l,l'\in\mathbb{L}} t^{s,t}_{SM} v^t_{s,l} v^{t+1}_{s,l'} R^e_{l,l'} + t^{e,t}_{NSMS} \right\} / B_e \qquad \forall e \in \mathbb{E}, \forall t \in \mathbb{T} \tag{7}$$

  The left hand side of Eq.(7) is the traffic load of a given link. The right hand side of Eq.(7) calculates the traffic load of a given link as the ratio of the total traffic that flows through the link and the capacity of the link. The numerator of the right hand side of Eq.(7) is a sum of the transmission rates of (1) Server-Client traffic, (2) Server-Server traffic, (3) Server Migration traffic, and (4) non-SMS service traffic that flow through the link, and this sum represents the total traffic that flows through the link. The denominator of the right hand side of Eq.(7) is the capacity of the link.

- Constraint (5): The traffic load of a link connected to a line card of a router is less than or equal to 1 in a given slot.

$$u^t_e \le 1 \qquad \forall e \in \mathbb{E}, \forall t \in \mathbb{T} \tag{8}$$

- Constraint (6): The traffic load of a link is greater than or equal to 0 in a given slot.

$$u^t_e \ge 0 \qquad \forall e \in \mathbb{E}, \forall t \in \mathbb{T} \tag{9}$$

- Constraint (7): The processing load of a WP is equal to the utilization of the WP in a given slot.

$$u^t_l = \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} x_s o^t_s v^t_{s,l} / C_l \qquad \forall l \in \mathbb{L}, \forall t \in \mathbb{T} \tag{10}$$

  The left hand side of Eq.(10) is the processing load of a given WP. The right hand side of Eq.(10) calculates the processing load of a given WP as the ratio of the total computing resources that NetApp servers on the WP require and the capacity of the WP. The numerator of the right hand side of Eq.(10) is a sum of the computing resources required by each active NetApp server running on the WP, and this sum represents the total computing resources that NetApp servers on the WP require. The denominator of the right hand side of Eq.(10) is the capacity of the WP.

- Constraint (8): The processing load of a WP is less than or equal to 1 in a given slot.

$$u^t_l \le 1 \qquad \forall l \in \mathbb{L}, \forall t \in \mathbb{T} \tag{11}$$

- Constraint (9): The processing load of a WP is greater than or equal to 0 in a given slot.

$$u^t_l \ge 0 \qquad \forall l \in \mathbb{L}, \forall t \in \mathbb{T} \tag{12}$$

5 | NUMERICAL RESULTS

In this section, we numerically solve the mixed-integer programming model developed in this paper and compare the total operating cost of SMS obtained from the model developed in this paper with that obtained in our previous paper[5,6]. We first explain in subsection 5.1 the network configurations and parameter values assumed in the numerical examples, and we then show

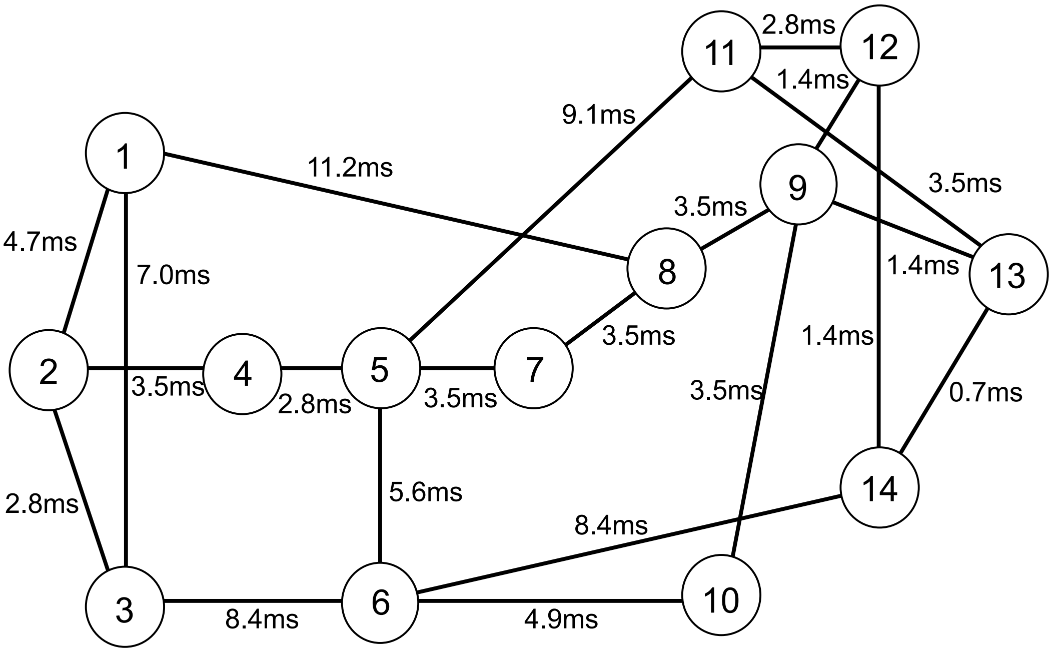


FIGURE 6 NSFNET

numerical examples in subsection 5.2. For ease of explanation, we refer to the model developed in this paper “the current model" and the model developed in our previous paper “the previous model" in the following. We numerically solved the current and previous models using IBM CPLEX solver[45] on a computer with two CPUs (2.1 GHz each) sharing a 256 GB main memory.

It is notable that the total operating cost of SMS obtained from the current model is always lower than or equal to that from the previous model. This is because the current model minimizes the total operating cost (a sum of the total monetary penalty and the total energy cost) of SMS, and the previous model, on the contrary, minimizes only the total monetary penalty of SMS without considering how the locations of NetApp servers impact the total energy cost of SMS. It is also notable that the current model achieves the theoretical lower bound of the total operating cost of SMS, as explained in Section 4.1.

5.1 | Network Configurations and Parameter Values of the Numerical Examples

Network configurations

In the numerical examples, we assume 14-node NSFNET (Figure 6)[46] and 25-node JPN25 (a nation-wide network in Japan) (Figure 7)[47]. Network configurations and parameter values of the NSFNET and JPN25 used in our numerical results are chosen from the actual network configurations and parameter values of these networks[46,47]. They are as follows. There are a total of $N = 14$ nodes on the NSFNET and $N = 25$ nodes on the JPN25, where N is the number of nodes. Capacity of each link on the NSFNET and the JPN25 is 10 Gbps. Link propagation delays range from 1.4 ms to 11.2 ms on the NSFNET and range from 0.14 ms to 4.66 ms on the JPN25.[6] On both networks, we assume that the propagation delay of a link dominates the delay that a traffic experiences on the link, namely, the packet transmission time on the link and the packet queuing delay at a node are both negligible.[7]

We assume a single NetApp on both the NSFNET and JPN25. NetApp consists of 1 NetApp server and N NetApp clients, where N is the number of nodes. N NetApp clients are distributed over N nodes with one NetApp client assigned to each node. The probability that a NetApp client is either active or inactive in a slot is equal and 0.5. Each node is a WP capable of running 1 NetApp server. The NetApp server is initially at a WP of node 1 (WP 1).

[6] The detailed values of the link propagation delays on the JPN25 are available at the public website[47].

[7] These assumptions are based on the following. When using a typical Internet link speed of 10 Gbits/sec[48], transmission time of a typical 1500 byte packet[49] is 0.0012 ms. This is multiple orders of magnitude smaller than the propagation delays of the NSFNET and JPN25. In addition, existing studies[50] show that a buffer for 10-20 packets in a core router achieves high TCP throughputs. A small buffer in a core router results in a negligible queuing delay.

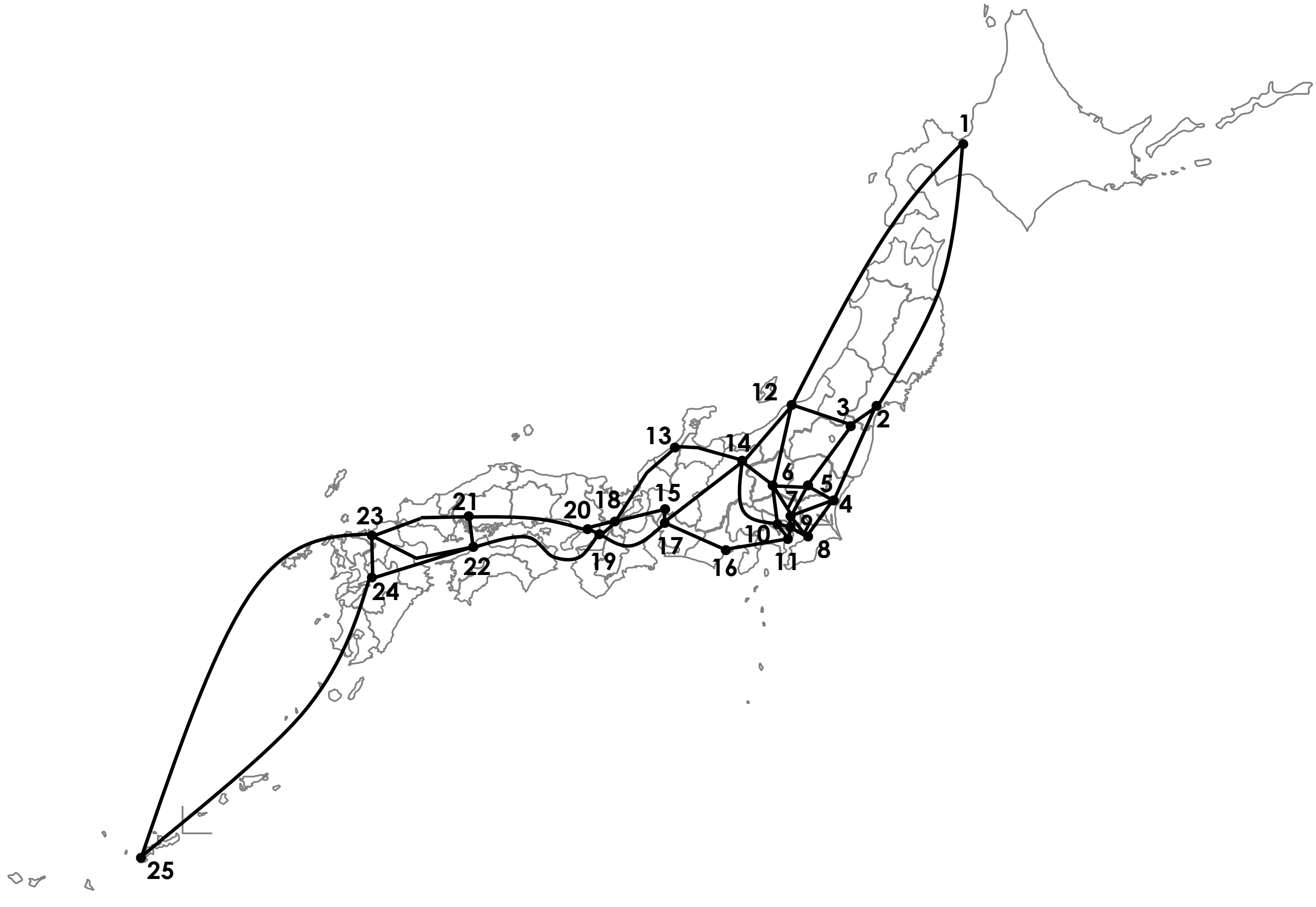


FIGURE 7 JPN25

Each link on the NSFNET and the JPN25 supports two types of traffic; the non-SMS service traffic and the traffic generated by NetApps (Server-Client, Server-Server, and Server Migration traffics) in SMS. With respect to the non-SMS service traffic, we assume that its transmission rate ($t^{e,t}_{NSMS}$) is either 1 Gbps (with probability 0.5) or 0 Gbps (with probability 0.5). We further assume that its transmission rate ($t^{e,t}_{NSMS}$) stays the same within a slot but it may change slot to slot. With respect to the traffic generated by NetApps in SMS, we assume that the transmission rate ($t^{s,r,t}_{SC}$) of Server-Client traffic is 1 [Mbps], the transmission rate ($t^{s,r,t}_{SS}$) of Server-Server traffic is 0, and the transmission rate ($t^{s,t}_{SM}$) of Server Migration traffic is 1 [Gbps]. The transmission rate of Server-Server traffic is 0, because we assume a single NetApp server in the numerical examples, and thus, no Server-Server traffic occurs in the numerical examples. We assume the shortest hop path routing algorithm for Server-Client communication and NetApp server migration.

SMS provider - SMS subscriber business model

In the numerical examples, we assume that the SMS provider and its SMS subscribers follow a simplified business model described below. In the numerical examples, we also assume the parameter values described below for the business model.

- The SMS provider and its SMS subscribers agree on the end-to-end delay of smaller than or equal to q_{max} in each slot between a NetApp server and active NetApp clients. Namely, the SMS provider agrees to provide the following QoS.

$$CQ^{SC}_{SMS}(s, l, r, l', t) = q \leq q_{max} \tag{13}$$

  where q [ms] is the end-to-end delay between WP l (where NetApp server s runs) and node l' (where NetApp client r locates) in time slot t. In the numerical examples, we assume $q_{max} = 10$ [ms] with the NSFNET and $q_{max} = 5$ [ms] with the JPN25.

- The time period of interest (T) is constant, and $T = 10$ slots and $T = 5$ slots for the NSFNET and for the JPN25, respectively.
- Server-Client communication and Server-Server communication both can potentially suffer from QoS degradation. In the numerical examples, however, there is no communication between NetApp servers (Server-Server communication), as there is only 1 NetApp server assumed in the network. Consequently, QoS degradation can occur only in Server-Client communication, and the SMS provider is only responsible for the monetary penalty of SMS, when QoS degrades in Server-Client communication.
- The monetary penalty of SMS incurred in Server-Client communication is assumed to be proportional to the excessive delay above the threshold (q_{max}) in the end-to-end delay between the NetApp server and its N NetApp clients. Namely, the SMS monetary penalty function for Server-Client communication ($MP^{SC}_{SMS}(q)$) is given by the following.

$$MP^{SC}_{SMS}(q) = \max(0, \alpha(q - q_{max})) \tag{14}$$

where q [ms] is the end-to-end delay between WP l (where NetApp server s runs) and node l' (where NetApp client r locates) and is given by the QoS function ($CQ^{SC}_{SMS}(s, l, r, l', t)$) in Eq.(13). The total monetary penalty of SMS is proportional to the excessive amount of delay, $q - q_{max}$, and α is the proportionality constant. In the numerical examples, we vary the value of α to examine how much the total monetary penalty of SMS contributes to the total operating cost of SMS. The monetary penalty is 0, when the end-to-end delay q is small than or equal to the threshold q_{max}.

SMS provider - utility company business model

In order to calculate the energy cost of SMS equipment (routers and WPs) in each slot in the time period of interest, we assume the following energy consumption model, energy cost model, and parameter values related to the business models.

Figure 8 depicts the linear energy consumption model[51] assumed in this paper. Figure 8 (a) shows the relationship between the traffic load (u^t_e) of a line-card in a router and the energy consumption ($ECS_{RT}(e, u^t_e, t)$) of the line-card. This relationship is expressed by Eq.(15) below.

$$ECS_{RT}(e, u^t_e, t) = \begin{cases} 0 & (u^t_e = 0) \\ P^e_C + P^e_V u^t_e & (u^t_e > 0) \end{cases} \tag{15}$$

It is notable that, when a router is connected to M number of links, it has M line-cards, and its energy consumption is a sum of the energy consumption of each line-card. Figure 8 (b) shows the relationship between the processing load (u^t_l) of a WP and the energy consumption ($ECS_{WP}(l, u^t_l, t)$) of the WP. This relationship is expressed by Eq.(16) below.

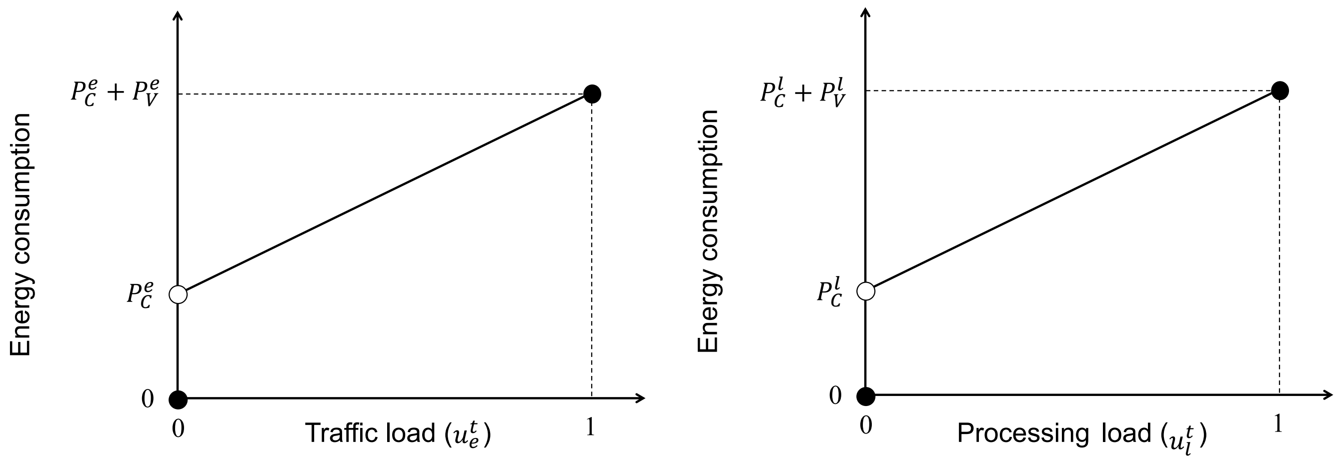
$$ECS_{WP}(l, u^t_l, t) = \begin{cases} 0 & (u^t_l = 0) \\ P^l_C + P^l_V u^t_l & (u^t_l > 0) \end{cases} \tag{16}$$

where P^e_C (P^l_C) is the base energy consumption of a line-card in a router (a WP), and P^e_V (P^l_V) is the slope of the energy consumption required to process an additional amount of traffic load (processing load).[8] P^e_C (P^l_C) is the energy consumption of a line-card (a WP) when it is active and when the traffic load (processing load) is $0 + \epsilon$, i.e., the energy consumption required to keep the line-card in a router (WP) active. Figures 8 (a), 8 (b), Eq.(15) and Eq.(16) show that the energy consumption of a line-card in a router ($ECS_{RT}(e, u^t_e, t)$) and of a WP ($ECS_{WP}(l, u^t_l, t)$) both linearly increase with a slope of P^e_V and P^l_V starting at $u^t_e = u^t_l = 0 + \epsilon$ and reaches its maximum at $u^t_e = u^t_l = 1$. It is notable that the energy consumption $ECS_{RT}(e, u^t_e, t)$ and $ECS_{WP}(l, u^t_l, t)$ are equal to 0, consuming no or little energy, when $u^t_e = u^t_l = 0$, i.e., when there is no traffic on a line-card in a router or when there is no processing load on a WP, namely, when they are in a sleep-mode.

In order to investigate how the base energy consumption (P^e_C and P^l_C) and the slope (P^e_V and P^l_V) of the energy consumption function impact the total operating cost of SMS, we consider two types of routers and WPs in the numerical examples, (1) energy-efficient and (2) non energy-efficient routers and WPs. These two types differ in their values of the base energy consumption P^e_C (P^l_C), i.e., the energy consumption to keep a line-card in a router (a WP) active.

- When a router (a WP) is energy-efficient, the base energy consumption of a line-card (a WP), P^e_C (P^l_C), is 0, and the energy consumption $ECS_{RT}(e, u^t_e, t)$ ($ECS_{WP}(l, u^t_l, t)$) linearly increase starting from 0 at $u^t_e = u^t_l = 0 + \epsilon$ and reaches its

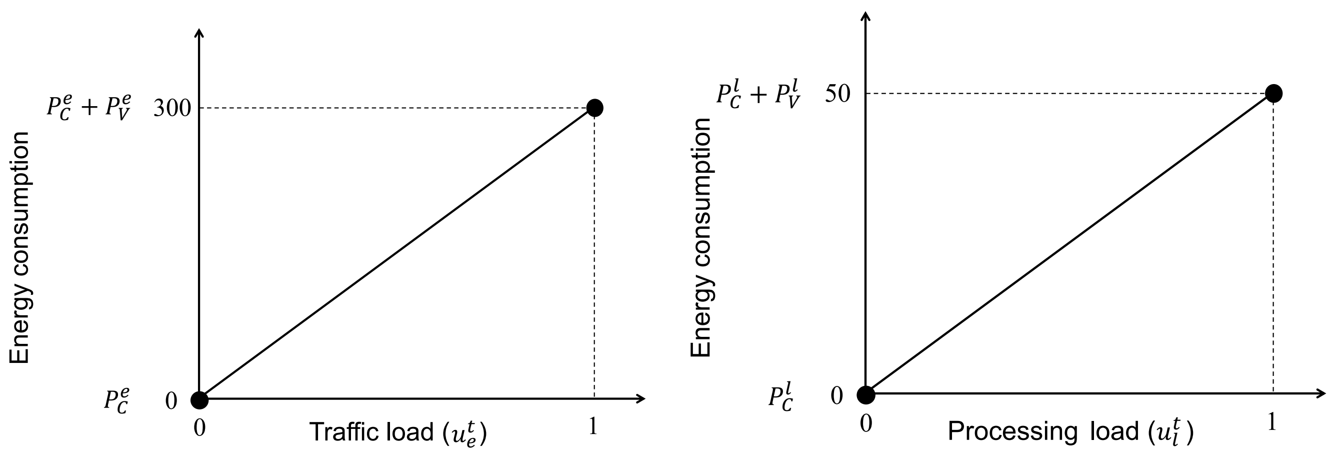
[8] It is notable that, in order to differentiate parameters used in the numerical result section, we use an upper subscript e to indicate a line-card e attached to link e in a router, and an upper subscript l to indicate a WP l.



(a) Energy consumption model for a line-card of a router

(b) Energy consumption model for a WP

FIGURE 8 Linear energy consumption model



(a)

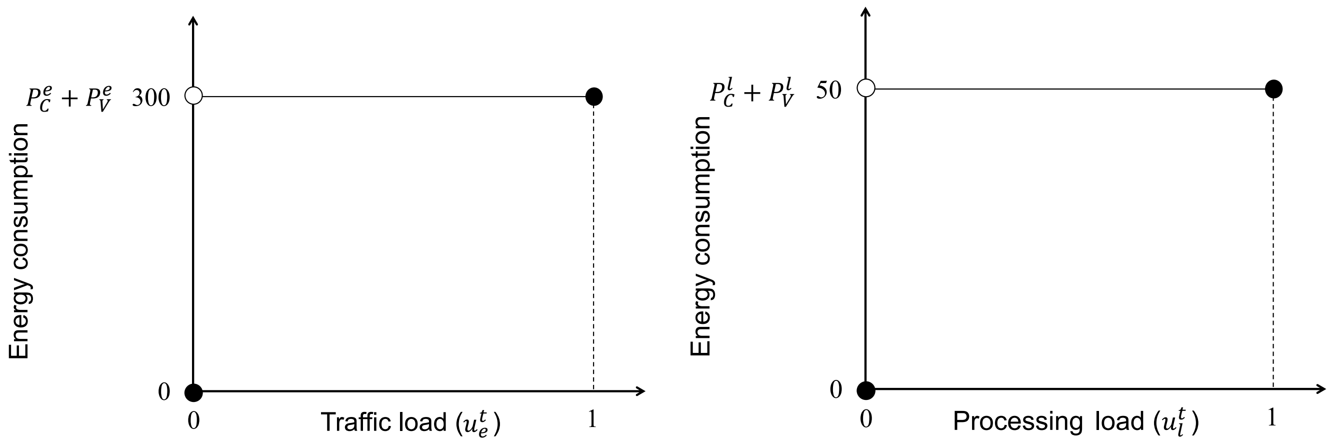
(b)

FIGURE 9 Energy consumption model for (a) an energy-efficient router ($P_C^e = 0$, $P_V^e = 300$), and (b) an energy-efficient WP ($P_C^l = 0$, $P_V^l = 50$)

maximum $P_C^e + P_V^e = 300$ ($P_C^l + P_V^l = 50$) at $u_e^t = u_l^t = 1$ (Figure 9). This is called energy-efficient, as the base energy consumption is 0.

- When a router (a WP) is non energy-efficient, the base energy consumption of a line-card (a WP) is $P_C^e + P_V^e = 300$ ($P_C^l + P_V^l = 50$), and the energy consumption $ECS_{RT}(e, u_e^t, t)$ ($ECS_{WP}(l, u_l^t, t)$) stays at this value for all values of u_e^t (u_l^t) from $u_e^t = u_l^t = 0 + \epsilon$ to $u_e^t = u_l^t = 1$ (Figure 10). This is called non energy-efficient, as the base energy consumption is the same as the maximum energy consumption.

Figures 9 and 10 determine the energy consumption of a line-card of a router and a WP from the traffic load on a line-card of a router and the processing load of a WP. The following energy cost models, Eq.(17) for a line-card of a router and Eq.(18) for a WP, calculate the energy cost (in \$) from the energy consumption (in Watts). Eq.(17) assumes that the energy cost (in \$) of a line-card in a router is equal to the energy consumption (in Watts). Similarly, Eq.(18) assumes that the energy cost (in \$) of a WP is equal to the energy consumption (in Watts) of a WP.



(a) **(b)**

FIGURE 10 Energy consumption model for (a) a non energy-efficient router ($P_C^e = 300$, $P_V^e = 0$), and (b) a non energy-efficient WP ($P_C^l = 50$, $P_V^l = 0$)

$$ECT_{RT}(z) = z \tag{17}$$

$$ECT_{WP}(z) = z \tag{18}$$

where z on the left hand side of the above equations is the energy consumption of a line-card in a router or a WP and is calculated from the energy consumption functions $ECS_{RT}(e, u_e^t, t)$ and $ECS_{WP}(l, u_l^t, t)$. z on the right hand side of the above equations is the corresponding energy cost (in \$).

SMS provider's total net income

With the business models described above, our mixed-integer programming model minimizes SMS provider's total operating cost in the numerical examples as described below:

$$\begin{aligned}
\text{minimize} \ \sum_{t\in\mathbb{T}} \Bigg\{ & \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} \sum_{l\in\mathbb{L}} \sum_{r\in\mathbb{R}_s^t} MP_{SMS}^{SC}(CQ_{SMS}^{SC}(s, l, r, u(r), t))\ o_s^t v_{s,l}^t \\
& + \sum_{n\in\mathbb{N}} \sum_{s,s'\in\mathbb{S}_n} \sum_{l,l'\in\mathbb{L}} MP_{SMS}^{SS}(CQ_{SMS}^{SS}(s, l, s', l', t))\ o_s^t v_{s,l}^t o_{s'}^t v_{s',l'}^t \\
& + \sum_{e\in\mathbb{E}} ECT_{RT}(ECS_{RT}(e, u_e^t, t)) + \sum_{l\in\mathbb{L}} ECT_{WP}(ECS_{WP}(l, u_l^t, t)) \Bigg\}
\end{aligned} \tag{19}$$

$$= \text{minimize} \ \sum_{t\in\mathbb{T}} \left\{ \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} \sum_{l\in\mathbb{L}} \sum_{r\in\mathbb{R}_s^t} \max(0, q - q_{max}) o_s^t v_{s,l}^t + \sum_{e\in\mathbb{E}} (P_C^e \lceil u_e^t \rceil + P_V^e u_e^t) + \sum_{l\in\mathbb{L}} (P_C^l \lceil u_l^t \rceil + P_V^l u_l^t) \right\} \tag{20}$$

where $\sum_{t\in\mathbb{T}} \sum_{n\in\mathbb{N}} \sum_{s\in\mathbb{S}_n} \sum_{l\in\mathbb{L}} \sum_{r\in\mathbb{R}_s^t} \max(0, q - q_{max}) o_s^t v_{s,l}^t$ is the sum of monetary penalties that the SMS provider pays to each SMS subscriber (the total monetary penalty of SMS), and $\sum_{t\in\mathbb{T}} \{\sum_{e\in\mathbb{E}} (P_C^e \lceil u_e^t \rceil + P_V^e u_e^t) + \sum_{l\in\mathbb{L}} (P_C^l \lceil u_l^t \rceil + P_V^l u_l^t)\}$ is the sum of energy cost that the SMS provider pays to the utility company to run the SMS equipment (the total energy cost of SMS). The total net income of SMS is maximized, when the total operating cost of SMS is minimized.

Non-SMS service provider - non-SMS service subscriber business model

In the numerical examples, we assume that the non-SMS service provider (i.e., the SMS provider) and its subscribers (i.e., non-SMS service subscribers) follow a simplified business model described below. In the numerical examples, we also assume the parameter values described below for the business model.

- Three types of SMS traffic (i.e., Server-Client, Server-Server, and Server Migration traffics) can potentially cause the monetary penalty of the non-SMS service. In the numerical examples, however, Server-Server and Server-Client traffics do not cause the monetary penalty of the non-SMS service. This is because there is only 1 NetApp server in the network and, thus, there is no Server-Server traffic generated in the network and because the volume of Server-Client traffic is typically negligible (e.g., 1 [Mbps]) compared to that of Server Migration traffic (e.g., 1 [Gbps]). Server Migration traffic is the only type of SMS traffic that can potentially cause the monetary penalty of the non-SMS service.
- The volume of Server Migration traffic increases as the number of hops that the NetApp server takes to migrate to its new location. Therefore, we assume that both QoS degradation of the non-SMS service and the monetary penalty of the non-SMS service caused by Server Migration traffic are proportional to the number of hops that the NetApp server takes to migrate to its new location. Namely, the monetary penalty function of the non-SMS service caused by Server Migration traffic ($MP_{NSMS}(CQ_{NSMS}(\mathbf{T^t_{SC}}, \mathbf{T^t_{SS}}, \mathbf{T^t_{SM}}))$ is given by the following.

$$MP_{NSMS}(CQ_{NSMS}(\mathbf{T^t_{SC}}, \mathbf{T^t_{SS}}, \mathbf{T^t_{SM}})) = h \times d \tag{21}$$

  where h is the number of hops on the shortest hop path from the current location of the NetApp server to the new location of the NetApp server, and d is the contribution that one hop brings to the monetary penalty of the non-SMS service. In the numerical examples, d is assumed to be 1. The unit of $MP_{NSMS}(CQ_{NSMS}(\mathbf{T^t_{SC}}, \mathbf{T^t_{SS}}, \mathbf{T^t_{SM}})$ is defined similarly to that of $MP^{SC}_{SMS}(s, l, r, l', t)$.

5.2 | Numerical Examples

5.2.1 | Numerical examples: without non-SMS service traffic

We first focus on the cases where there is no non-SMS service traffic and there is only SMS traffic in the substrate network.

Numerical examples of the dominant total energy cost

In the following numerical examples, we consider cases where the total energy cost contributes to the total operating cost of SMS significantly more than the total monetary penalty. We set α to 1 in the monetary penalty function $MP^{SC}_{SMS}(q)$, namely, the monetary penalty of SMS incurred in Server-Client communication is assumed to be equal to the excessive delay above the threshold (q_{max}) in the end-to-end delay between the NetApp server and its N NetApp clients.

Figures 11 and 12 depict the total operating cost, the total monetary penalty and the total energy cost of SMS with NSFNET, when we use two energy consumption models described in Figures 9 and 10, respectively.

In Figure 11, all routers and WPs are energy-efficient (described in Figure 9). Figure 11 depicts the total operating cost of SMS (Figure 11 (a)) obtained from the current model (shown as a red line) and the previous model without considering the total energy cost of SMS (shown as a blue line). This figure also shows the components of the total operating cost of SMS, i.e., the total monetary penalty of SMS (Figure 11 (b)) and the total energy cost of SMS (Figure 11 (c)). On Figure 11, the x-axis is the upper bound (U) on the total monetary penalty of the non-SMS service in the smoothing window.

When the upper bound (U) is 0, Figure 11 (a), 11 (b) and 11 (c), all show that both the current model and the previous model result in the same total operating cost, the total monetary penalty and the total energy cost of SMS. This is because no NetApp server is allowed to migrate in the current model and in the previous model, when U is 0. (The same observation holds for Figure 12 when U is 0.)

When the upper bound (U) is non-zero, the NetApp server may migrate, when necessary. Figure 11 (a) shows that the total operating cost of SMS of the current model and that of the previous model are approximately 676 and 866, respectively, when the upper bound (U) is non-zero ($U = 2, 4, 6, 8$ and 10). The current model shows approximately 22% lower total operating cost of SMS than the previous model. This is because, compared to the previous model, the current model significantly decreases the total energy cost of SMS, a dominant element of the total operating cost of SMS, as shown in Figure 11 (c). This significant decrease in the total energy cost of SMS is explained as follows. When SMS equipment is energy-efficient, the total energy cost of SMS depends on the total volume of Server Migration traffic because energy cost of routers is proportional to routers' traffic

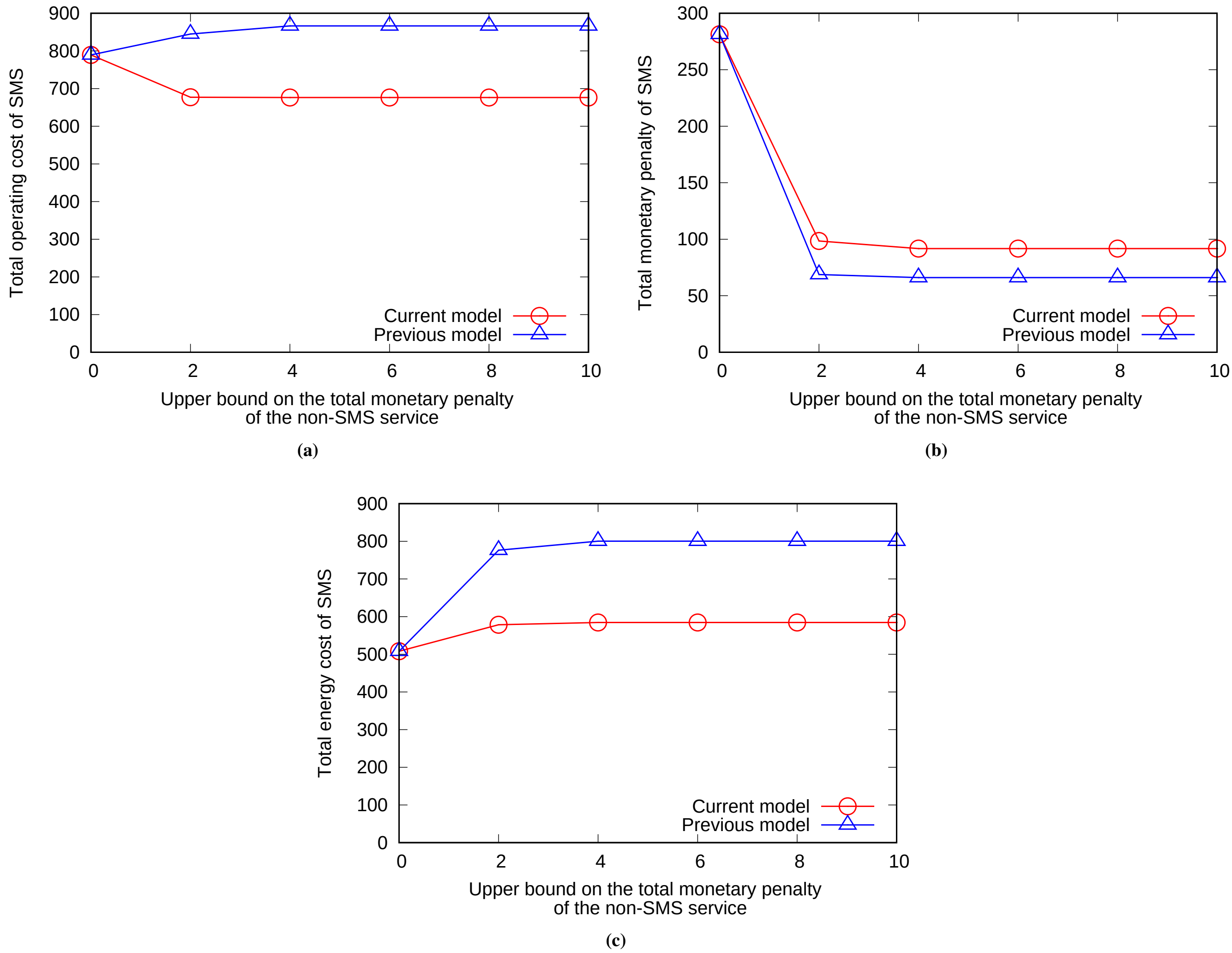


FIGURE 11 NSFNET, (a) total operating cost, (b) total monetary penalty, (c) total energy cost of SMS (no non-SMS service traffic, the total energy cost of SMS dominates the total operating cost of SMS ($\alpha = 1$), energy-efficient SMS equipment)

load, which is dominated by Server Migration traffic. In the current model, the NetApp server tends to migrate to a WP that is close to NetApp clients (in the hop count) within the initial few slots in the time period of interest in order to reduce the total monetary penalty of SMS and stays at the WP for the remaining duration, refraining from further migrating to another WP in order to avoid increasing the total energy cost of SMS due to Server Migration traffic. On the other hand, in the previous model, the NetApp server migrates to a WP that only minimizes the total monetary penalty of SMS without minimizing the total energy cost of SMS, namely, the NetApp server migrates to a WP that is close to the location of active NetApp clients, reducing the total monetary penalty of SMS (as seen in Figure 11 (b)), but possibly increasing the total energy cost of SMS (as seen in Figure 11 (c)).

In Figure 12, all routers and WPs are non energy-efficient (described in Figure 10). Figure 12 (a) shows that the total operating cost of SMS of the current model is approximately 42,865 and approximately 10% lower than the total operating cost of SMS (approximately 47,846) of the previous model. This is because, compared to the previous model, the current model decreases the total energy cost of SMS, a dominant element of the total operating cost of SMS, as shown in Figure 12 (c). This is explained as follows. When SMS equipment is non energy-efficient, the total energy cost of SMS depends on the number of line-cards of routers in active-state, because each line-card of a router in active-state consumes the base energy, which is equal to the maximum energy. In the current model, the NetApp server tends to migrate to a WP so that the Server Migration traffic caused by this NetApp server migration activates the least number of inactive line-cards on the migration route in order to avoid increasing the

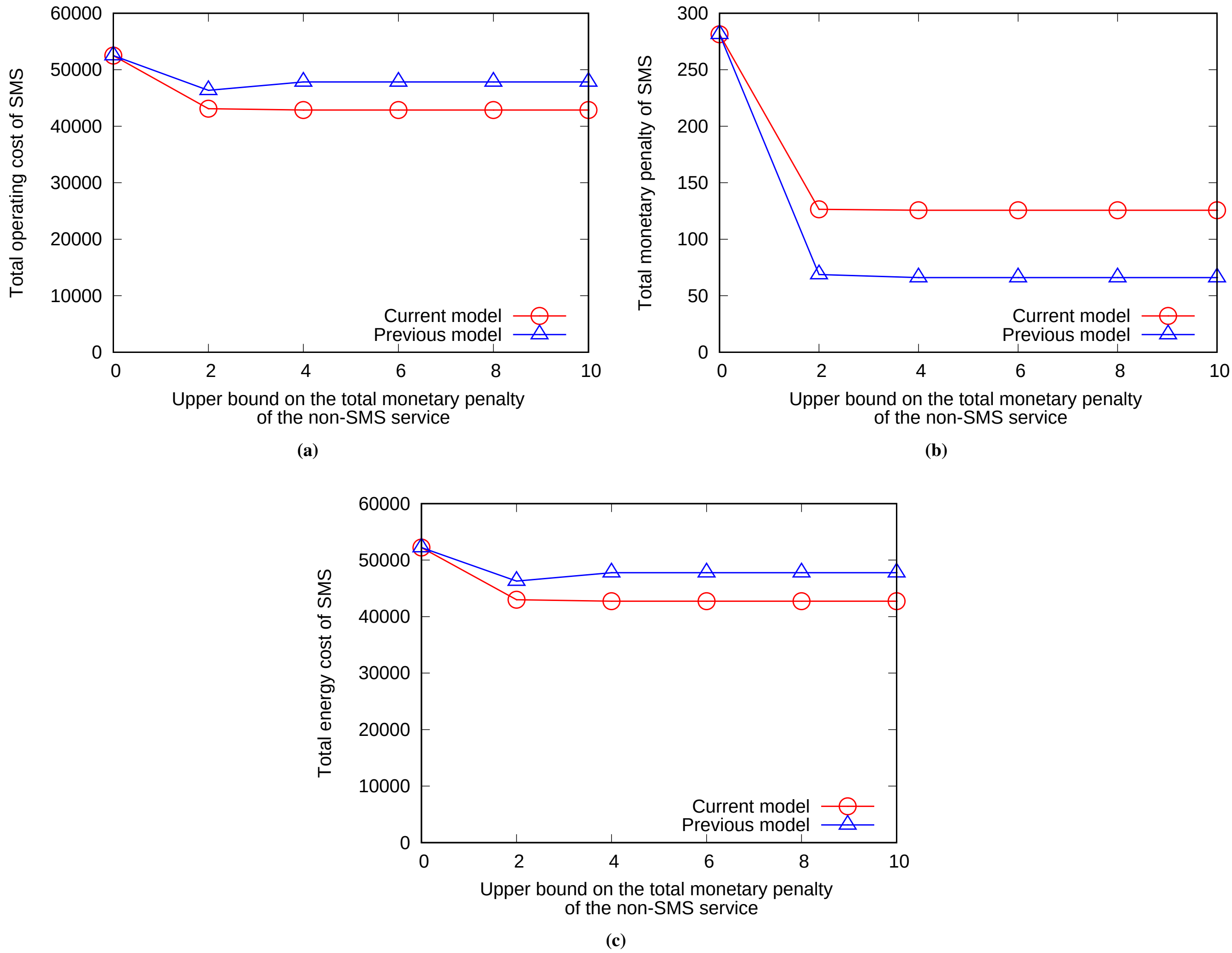


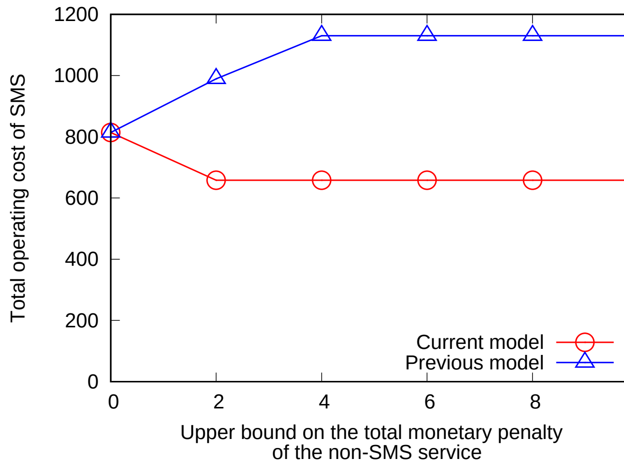
FIGURE 12 NSFNET, (a) total operating cost, (b) total monetary penalty, (c) total energy cost of SMS (no non-SMS service traffic, the total energy cost of SMS dominates the total operating cost of SMS ($\alpha = 1$), non energy-efficient SMS equipment)

total energy cost of SMS. On the other hand, in the previous model, similarly to Figure 11, the NetApp server migrates to a WP that only minimizes the total monetary penalty of SMS without minimizing the total energy cost of SMS.

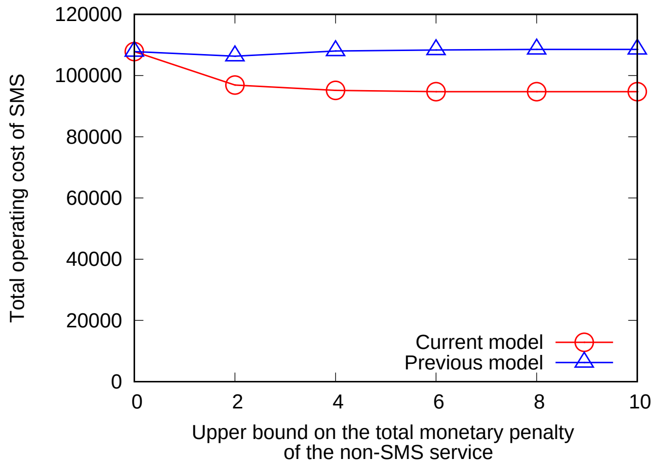
Figure 13 depicts the total operating cost of SMS with JPN25 when we use two energy consumption models, energy-efficient SMS equipment (Figure 13 (a)) and non energy-efficient SMS equipment (Figure 13 (b)). Parameter values used in Figures 13 (a) and 13 (b) are same as those used in Figures 11 and 12, respectively. Figures 13 (a) and 13 (b) show the similar results as those with the NSFNET (Figures 11 (a) and 12 (a)). However, compared to the results with NSFNET, the difference in the total operating cost of SMS between the current model and the previous model is larger with JPN25 (e.g., the current model shows up to 42% lower total operating cost of SMS than the previous model in Figure 13 (a)). This is because JPN 25 is larger in the number of nodes than NSFNET, and consequently, it magnifies the benefits of the current model over those of the previous model.

Numerical examples of the comparable total monetary penalty and total energy cost

In the following numerical examples, we consider cases where both the total monetary penalty and the total energy cost comparably contribute to the total operating cost of SMS. We vary the value of α from 10 to 1,000 in the monetary penalty function $MP_{SMS}^{SC}(q)$.



(a)



(b)

FIGURE 13 JPN25, total operating cost of SMS, (a) energy-efficient SMS equipment, (b) non energy-efficient SMS equipment (no non-SMS service traffic, the total energy cost of SMS dominates the total operating cost of SMS ($\alpha = 1$))

Figure 14 depicts the total operating cost, the total monetary penalty and the total energy cost of SMS with NSFNET when we use the energy-efficient model described in Figure 9. In Figure 14, all routers and WPs are energy-efficient. Figure 14 (a) shows that the total operating cost of SMS of the current model (approximately 1,374) is almost identical to that of the previous model (approximately 1,460). This is explained as follows. Because both the total monetary penalty and the total energy cost comparably contribute to the total operating cost of SMS, the NetApp server in the current model migrates among WPs such that both the total monetary penalty and the total energy cost decrease. As a result, the NetApp server in the current model often migrates to a WP to which the NetApp server in the previous model migrates, i.e., migration of the NetApp servers in the current model becomes more similar to that in the previous model compared to the case with the dominant total energy cost (Figure 11).

Although the total operating cost of SMS is almost identical between the current model and the previous model, Figure 14 (c) shows that the current model achieves approximately 13% to 15% lower total energy cost than the previous model. This is because the NetApp server in the current model tends to stay at a WP in order to avoid generating the Server Migration traffic and causing the energy cost of routers to increase rather to migrate among WPs in order to reduce monetary penalty. This occurs when the reduction in the energy cost by avoiding the Server Migration traffic is higher than the reduction in the monetary penalty by migrating to a new WP. Similar results are obtained for non energy-efficient model. (Figures are not included due to space limitations.)

Figure 15 depicts the total operating cost, the total monetary penalty and the total energy cost of SMS with JPN25 when we use the energy-efficient model described in Figure 9. The total operating cost of SMS of the current model is approximately 22% lower than that of the previous model (when the upper bound (U) is non-zero and NetApp servers are allowed to migrate). This is because the NetApp server migrates less often in the current model than in the previous model, resulting in the smaller Server Migration traffic and lower energy cost due to Server Migration traffic.

When all routers and WPs are non energy-efficient in JPN25, the total operating cost of SMS of the current model is almost identical to that of the previous model (Figures are not included due to space limitations). This is because, similarly to Figure 14, in order to reduce both the total monetary penalty and the total energy cost, the NetApp server in the current model often migrates to a WP to which the NetApp server in the previous model migrates, i.e., migration of the NetApp servers in the current model becomes similar to that in the previous model.

Numerical examples of the dominant total monetary penalty

In the following, we briefly summarize numerical examples where the total monetary penalty contributes to the total operating cost of SMS significantly more than the total energy cost. In the numerical examples, we set α to 10,000 in the monetary penalty function $MP^{SC}_{SMS}(q)$. Figures are not included due to space limitations.

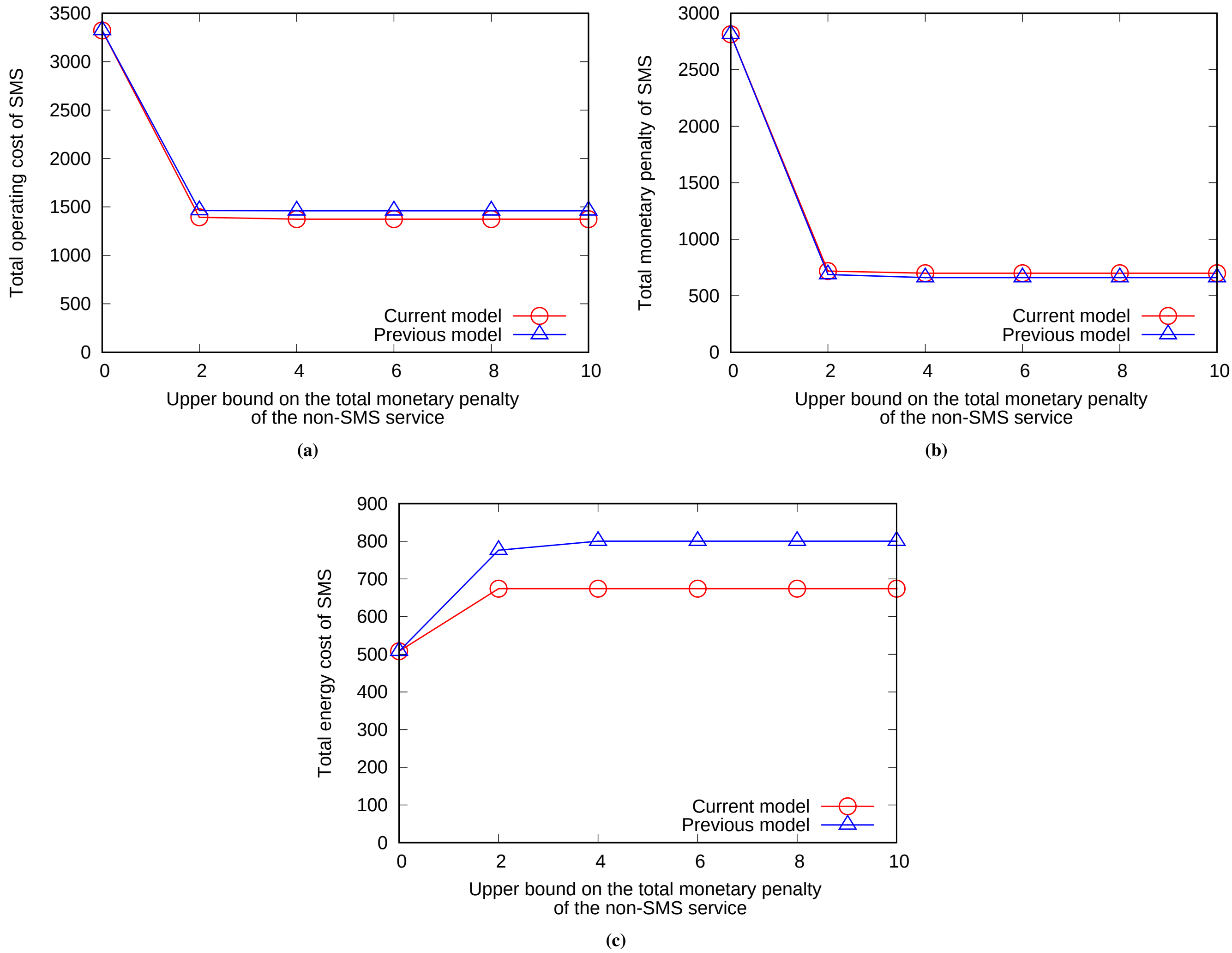


FIGURE 14 NSFNET, (a) total operating cost, (b) total monetary penalty, (c) total energy cost of SMS (no non-SMS service traffic, both the total monetary penalty and the total energy cost of SMS comparably contribute to the total operating cost of SMS ($\alpha = 10$), energy-efficient SMS equipment)

The total operating cost, the total monetary penalty and the total energy cost of SMS are almost identical between the current model and the previous model for both NSFNET and JPN25 in each of the two energy consumption models described in Figures 9 and 10. This is because the total monetary penalty of SMS dominates the total operating cost of SMS, and consequently, the NetApp server in the current model migrates primarily to minimize the total monetary penalty of SMS, i.e., the NetApp server in the current model behaves almost identically to that in the previous model without considering the total energy cost.

5.2.2 | Numerical examples: with the non-SMS service traffic

We next focus on the cases where there exits the non-SMS service traffic in addition to the SMS traffic in the substrate network.

Numerical examples of the dominant total energy cost

In the following numerical examples, we consider cases where the total energy cost contributes to the total operating cost of SMS significantly more than the total monetary penalty. We set α to 1 in the monetary penalty function $MP^{SC}_{SMS}(q)$.

Figures 16 and 17 depict the total operating cost, the total monetary penalty and the total energy cost of SMS with NSFNET when we use two energy consumption models described in Figures 9 and 10, respectively.

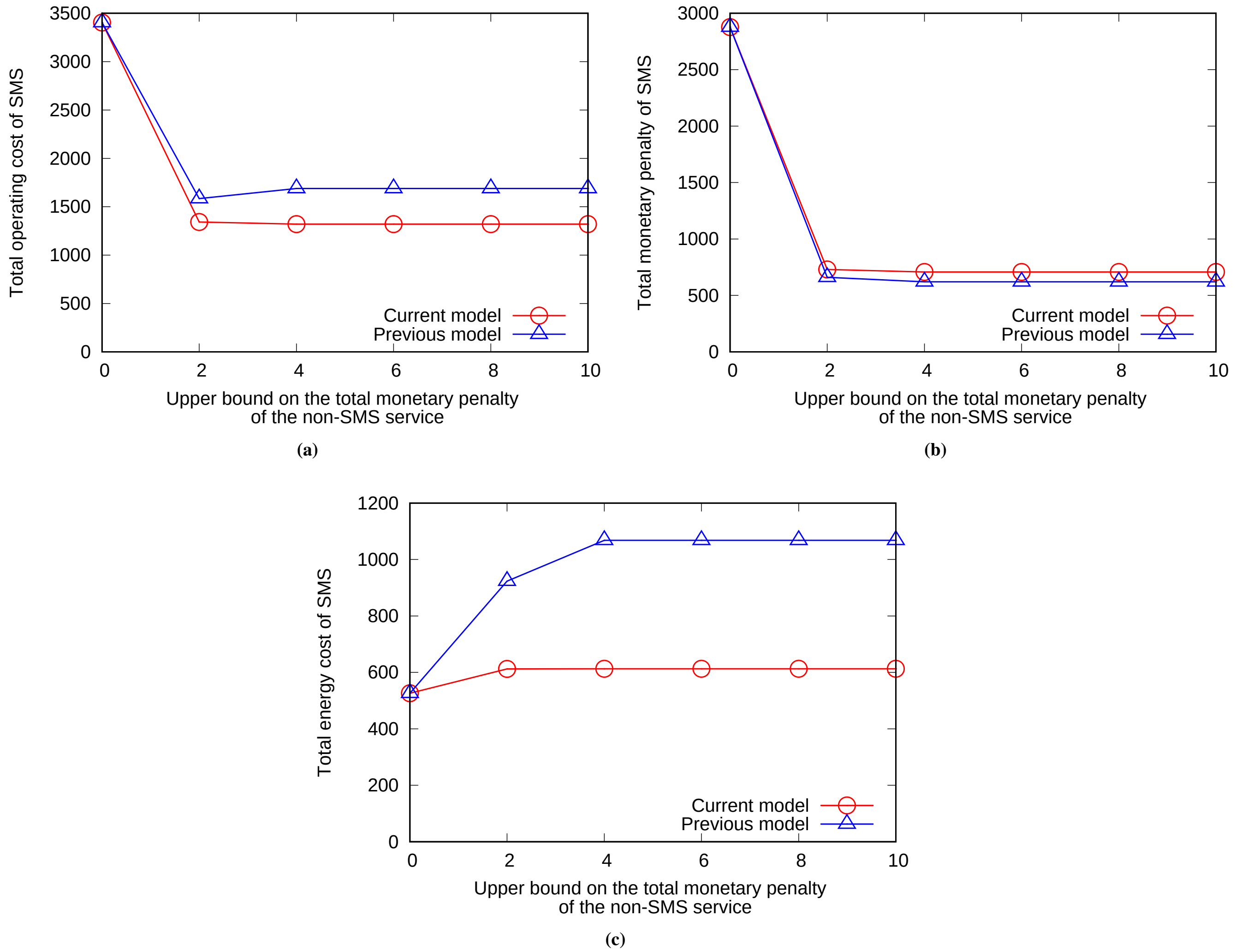


FIGURE 15 JPN25, (a) total operating cost, (b) total monetary penalty, (c) total energy cost of SMS (no non-SMS service traffic, both the total monetary penalty and the total energy cost of SMS comparably contribute to the total operating cost of SMS ($\alpha = 10$), energy-efficient SMS equipment)

When all routers and WPs are energy-efficient (Figure 16), the total operating cost of SMS of the current model (approximately 6,844) is almost identical to that of the previous model (approximately 7,034). This is because, unlike in Figure 11 (c), the difference between the total energy cost of the current model and that of the previous model is not significant, as seen in Figure 16 (c), and also because the total energy cost of SMS, not the monetary penalty of SMS, dominates the total operating cost of SMS, as seen in Figure 16 (b) and Figure 16 (c).

The difference between the total energy cost of the current model and that of the previous model is not significant in Figure 16 (c). This is because the total energy cost of routers due to the non-SMS service traffic is significant and it outweighs the energy cost reduction that the current model brings from decreasing the energy cost due to Server Migration traffic, making the total energy cost of the current model similar to that of the previous model. It is notable that the total energy cost of routers due to the non-SMS traffic is significant, as the non-SMS service traffic is much higher than the SMS traffic as seen below; (a) The non-SMS service traffic is 0.5 Gbps on the average, and it is placed on each and every link of all 42 links of the NSFNET and in each and every slot of the time period of interest, regardless of whether the NetApp server migrates or not, (b) the Server-Client traffic (SMS traffic) is 1 Mbps, and it is placed only on the links between the NetApp server and active NetApp clients, and (c) the Server Migration traffic (SMS traffic) is 1 Gbps, and it is placed only on the links between the WP from which the NetApp

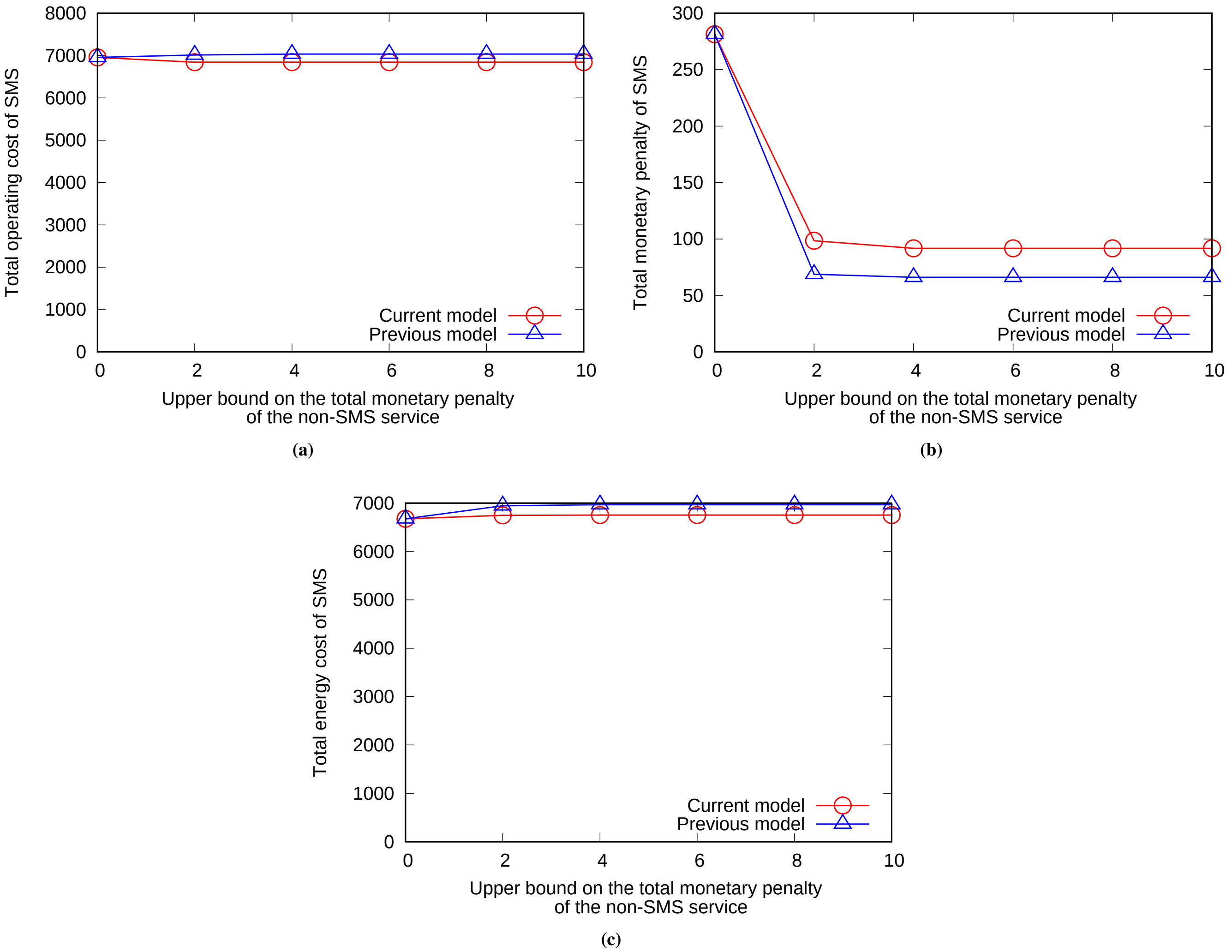
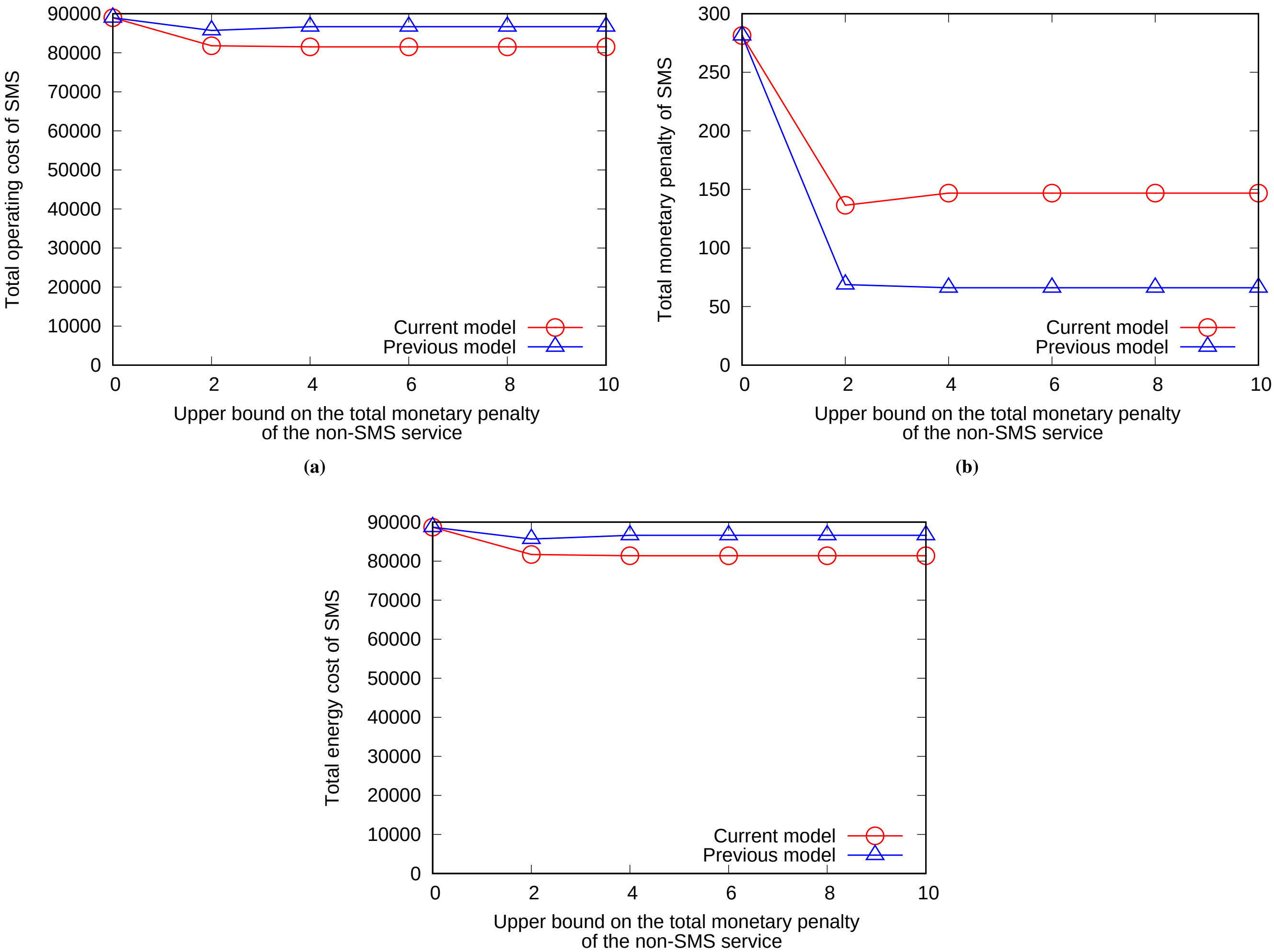


FIGURE 16 NSFNET, (a) total operating cost, (b) total monetary penalty, (c) total energy cost of SMS (non-SMS service traffic, the total energy cost of SMS dominates the total operating cost of SMS ($\alpha = 1$), energy-efficient SMS equipment)

server migrates and the new WP to which the NetApp server migrates and only in the slot when the NetApp server migrates.[9] (a) is significant compared to (b) and (c).

When all routers and WPs are non energy-efficient (Figure 17), the total operating cost of SMS of the current model is approximately 81,526 and approximately 6% lower than that of the previous model (approximately 86,666), as seen in Figure 17 (a). This is because the total energy cost of SMS of the current model is approximately 6% lower than that of the previous model, as seen in Figure 17 (c). This is explained as follows. In Figure 17, similarly to Figure 16, we observe that the total energy cost of SMS is dominated by the total energy cost of routers. We also observe, however, that unlike in Figure 16 (c), the total energy cost of routers in Figure 17 (c) is not always dominated by the total energy cost of routers due to the non-SMS service traffic: it is the total energy cost of routers due to the SMS traffic (i.e., Server-Client traffic and Server Migration traffic) that significantly contributes to the total energy cost of SMS. This is because, when a non energy-efficient router is active, it is subject to at least its base energy consumption, which is equal to the maximum energy consumption, even when it supports a small amount of traffic (such as Server-Client traffic whose bit-rate is only 1 Mbps). Therefore, concentrating the traffic on a smaller number of routers saves more energy than spreading the traffic over a larger number of routers, each of which consuming the maximum energy consumption. The NetApp server in the previous model does not consider the number of routers in active-state and migrates in

[9] It is notable that there is only 1 NetApp server in the numerical examples, and thus, there is no Server-Server traffic (SMS traffic) in the numerical examples.



(c)

FIGURE 17 NSFNET, (a) total operating cost, (b) total monetary penalty, (c) total energy cost of SMS (non-SMS service traffic, the total energy cost of SMS dominates the total operating cost of SMS ($\alpha = 1$), non energy-efficient SMS equipment)

order to minimize only the total monetary penalty of SMS, while the NetApp server in the current model migrates to reduce the number of routers in active-state. Namely, in the current model, the NetApp server migrates to a WP such that the smallest number of routers are involved in supporting Server Migration traffic and Server-Client traffic. This accounts for difference in the total energy cost of SMS between the current model and the previous model in Figure 17 (c).

Numerical examples of the comparable total monetary penalty and total energy cost

In the following, we briefly summarize numerical examples where both the total monetary penalty and the total energy cost comparably contribute to the total operating cost of SMS. We vary the value of α from 10 to 1,000 in the monetary penalty function $MP^{SC}_{SMS}(q)$.

The total operating cost, the total monetary penalty and the total energy cost of SMS are almost identical between the current model and the previous model for both NSFNET and JPN25 in each of the two energy consumption models described in Figures 9 and 10. As explained in “Numerical examples of the comparable total monetary penalty and total energy cost" of subsection 5.2.1, this is because the total monetary penalty of SMS contributes to the total operating cost of SMS as much as the total energy cost does, and consequently, the NetApp server in the current model often migrates to a WP that is close to active NetApp clients in order to decrease the total monetary penalty of SMS, i.e., migration of the NetApp servers in the current model becomes similar to that in the previous model.

Numerical examples of the dominant total monetary penalty

In the following, we briefly summarize numerical examples where the total monetary penalty contributes to the total operating cost of SMS significantly more than the total energy cost. We set α to 10,000 in the monetary penalty function $MP_{SMS}^{SC}(q)$. Figures are not included due to space limitations.

The total operating cost, the total monetary penalty and the total energy cost of SMS are almost identical between the current model and the previous model for both NSFNET and JPN25 in each of the two energy consumption models described in Figures 9 and 10. As explained in "Numerical examples of the dominant total monetary penalty" of subsection 5.2.1, this is because the total monetary penalty of SMS dominates the total operating cost of SMS, and consequently, the NetApp server in the current model migrates primarily to minimize the total monetary penalty of SMS, i.e., the NetApp server in the current model behaves almost identically to that in the previous model without considering the total energy cost.

6 | CONCLUSIONS

In this paper, we have proposed server migration service (SMS). In SMS, NetApp servers migrate among WPs in the network (1) to optimally locate themselves in relation to NetApp clients and minimize the monetary penalty of SMS and (2) to optimally distribute traffic load over routers and processing load over (physical) computers and minimize the total energy cost of SMS.

The mixed-integer programming model for SMS developed in this paper determines when and to which locations NetApp servers migrate and achieves the best possible total operating cost of SMS, while preventing NetApp servers from excessively migrating and adversely impacting QoS of the non-SMS service that share the substrate network resources with SMS.

The simulation results have shown that the mixed-integer programming model developed in this paper achieves up to 42% lower total operating cost of SMS compared to the model that only minimizes the monetary penalty of SMS without considering the energy cost of SMS.

Future work includes developing a model to determine locations and capacities of WPs in the network.

ACKNOWLEDGMENT

This research and development work was supported by the MIC/SCOPE #162108003.

References

1. Amazon EC2, http://aws.amazon.com/ec2.
2. Google Computer Engine, https://cloud.google.com/compute.
3. Microsoft Azure, https://azure.microsoft.com/en-us/.
4. V. Clincy and B. Wilgor, "Subjective evaluation of latency and packet loss in cloud-based game," In Proc. 10th International Conference on Information Technology: New Generations, pp. 473–476, 2013.
5. Y. Fukushima, T. Murase, T. Yokohira, and T. Suda, "Optimization of server locations in server migration service," In Proc. International Conference on Networking and Services (ICNS), pp. 200–206, 2013.
6. Y. Fukushima, T. Murase, G. Motoyoshi, T. Yokohira, and T. Suda, "Determining server locations in server migration service to minimize monetary penalty of dynamic server migration," J Netw Syst Manage, Vol. 26, pp. 993–1033, 2018.
7. S. Weisong, C. Jie, Z. Quan, L. Youhuizi, and X. Lanyu, "Edge computing: vision and challenges," IEEE Internet of Things Journal, Vol. 3, Iss. 5, pp. 637–646, 2016.

8. F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the Internet of Things,” In Proc. the MCC Workshop on Mobile Cloud Computing, pp. 13–16, 2012.

9. K. Chen, K. Xu, K. Xi, and J. H. Chao, “Intelligent virtual machine placement for cost efficiency in geo-distributed cloud systems,” In Proc. IEEE International Conference on Communications (ICC), pp. 3498–3503, 2013.

10. M. Anan and N. Nasser, “SLA-based optimization of energy efficiency for green cloud computing,” In Proc. IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2015.

11. M. Anan, N. Nasser, A. Ahmed, and A. Alfuqaha, “Optimization of power and migration cost in virtualized data centers,” . In Proc. IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–5, 2016.

12. W. Quanwang, I. Fuyuki, Z. Qingsheng, and X. Yunni, “Energy and migration cost-aware dynamic virtual machine consolidation in heterogeneous cloud datacenters,” IEEE Trans on Serv Comp, 2016.

13. A. Beloglazov and R. Buyya, “Energy efficient allocation of virtual machines in cloud data centers,” In Proc. IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGC), pp. 577–578, 2010.

14. H. Goudarzi and M. Pedram, “Energy efficient virtual machine replication and placement,” In Proc. IEEE International Conference on Cloud Computing (CLOUD), pp. 750–757, 2012.

15. S. Ranjan, J. Rolia, H. Fu, and E. Knightly, “QoS-driven server migration for Internet data centers,” In Proc. tenth International Workshop on Quality of Service (IWQoS), pp. 3–12, 2002.

16. B. G. Assefa, O. Ozkasap, I. Kizil, M. Aloqaily, and O. Bouachir, “Energy efficiency in SDDC: considering server and network utilities,” In Proc. IEEE Symposium on Computer and Communications (ISCC), 6 pages, 2020.

17. B. Kantarci, L. Foschini, A. Corradi, and T. Mouftah, “Inter-and-intra data center VM-placement for energy-efficient large-scale cloud systems,” In Proc. International Workshop on Management and Security technologies for Cloud Computing (ManSec-CC), pp. 708–713, 2012.

18. K. Atefeh, K. G. Saurabh, and B. Rajkumar, “Energy and carbon-efficient placement of virtual machines in distributed cloud data centers,” In Proc. European Conference on Parallel Processing, pp. 317–328, 2013.

19. M. Alicherry and V. T. Lakshman, “Network aware resource allocation in distributed clouds,” In Proc. IEEE Conference on Computer Communications (INFOCOM), pp. 963–971, 2012.

20. H. Ichihara, Y. Koizumi, H. Ohsaki, K. Hato, J. Murayama, and M. Imase, “On the integrated control of virtual machine live migration and traffic engineering for cloud computing,” In Proc. IEEE Global Communications Conference (GLOBECOM), pp. 1629–1634, 2012.

21. A. Yousefpour, A. Patil, G. Ishigaki, I. Kim, X. Wang, H. C. Cankaya, Q. Zhang, W. Xie, and J. P. Jue, “FOGPLAN: a lightweight QoS-aware dynamic fog service provisioning framework,” IEEE Internet of Things Journal, Vol. 6, No. 3, pp. 5080–5096, 2019.

22. M. P. J. Mahenge, C. Li, and C. A. Sanga, “Collaborative mobile edge and cloud computing: tasks unloading for improving users’ quality of experience in resource-intensive mobile applications,” In Proc. International Conference on Computer and Communications Systems (ICCS), pp. 322–326, 2019.

23. M. G. Arani, A. Souri, F. Safara, and M. Norouzi, “An efficient task scheduling approach using moth-flame optimization algorithm for cyber-physical system applications in fog computing,” Trans Emerging Tel Tech., 2020.

24. J. Ahn, J. Lee, D. Niyato, and H. S. Park, “Novel QoS-guaranteed orchestration scheme for energy-efficient mobile augmented reality applications in multi-access edge computing,” IEEE Transactions on Vehicular Technology, Vol. 69, No. 11, 2020.

25. P. Zhao, P. Wang, X. Yang, and J. Lin, “Towards cost-efficient edge intelligent computing with elastic deployment of container-based microservices,” IEEE Access, Vol. 8, pp. 102947–102957, 2020.

26. Y. Ren, F. Zeng, W. Li, and L. Meng, "A low-cost edge server placement strategy in wireless metropolitan area networks," In Proc. International Conference on Computer Communication and Networks (ICCCN), 6 pages, 2018.

27. M. Daraghmeh, I. Al-Ridhawi, M. Aloqaily, Y. Jararweh, and A. Agarwal, "A power management approach to reduce energy consumption for edge computing servers," In Proc. International Conference on Fog and Mobile Edge Computing (FMEC), pp. 259–264, 2019.

28. K. Sonbol, O. Ozkasap, I. Al-Oqily, and M. Aloqaily, "EdgeKV: decentralized, scalable, and consistent storage for the edge," Journal of Parallel and Distributed Computing, pp. 28–40, 2020.

29. P. Yadav and S. Kar, "Evaluating the impact of region based content popularity of videos on the cost of CDN deployment," In Proc. National Conference on Communications (NCC), 6 pages, 2020.

30. Q. Shuai, K. Wang, F. Miao, and L. Jin, "A cost-based distributed algorithm for load balancing in content delivery network," In Proc. International Conference on Intelligent Human-Machine Systems and Cybermetrics (IHMSC), pp. 11–15, 2017.

31. G. C. Harrison, M. D. Chess, and A. Kershenbaum, "Mobile agents: Are they a good idea? Technical report,", IBM Research Division, 1995.

32. B. D. Lange and M. Oshima, "Seven good reasons for mobile agents," ACM Comm., Vol. 42, Iss. 3, pp. 88–89, 1999.

33. Y. Chen, D. Guo, M. Z. A. Bhuiyan, M. Xu, G. Wang, and P. LV, "Towards profit optimization during online participant selection in compressive mobile crowsensing," ACM Transactions on Sensor Networks, Vol. 15, No. 4, pp. 38.1–38.29, 2019.

34. Y. Wu, H. Huang, N. Wu, Y. Wang, M. Z. A. Bhuiyan, and T. Wang, "An incentive-based protection and recovery strategy for secure big data in social networks," Information Sciences, Vol. 508, pp. 79–91, 2020.

35. T. Hara, M. Tsukamoto, and S. Nishio, "A scheduling method of database migration for WAN environments," In Proc. Brazilian Symposium on Database (SBBD), pp. 125–136, 1999.

36. Y. Fukushima, T. Yokohira, T. Murase, and T. Suda, "Power-aware server location decision in server migration service," In Proc. International Conference on Information and Communication Technology Convergence (ICTC), pp. 150–155, 2016.

37. G. Quadrio, A. Bujari, E. C. Palazzi, D. Ronzani, D. Maggiorini, and A. L. Ripamonti, "Network analysis of the steam in-home streaming game system," In Proc. the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom), pp. 475–476, 2016.

38. S. C. Lee, "The revolution of StarCraft network traffic," In Proc. the 11th Annual Workshop on Network and System Support for Games (NetGames), Article No. 18, 2012.

39. S. Nathan, P. Kulkarni, and U. Bellur, "Resource availability based performance benchmarking of virtual machine migrations," In Proc. the 4th ACM/SPEC International Conference on Performance Engineering (ICPE), pp. 387–398, 2013.

40. S. Aleksic, "Analysis of Power Consumption in Future High-Capacity Network Nodes," J Opt Commu Net, Vol. 1, Iss. 3, pp. 245–258, 2009.

41. A. Francini and D. Stiliadis, "Performance bounds of rate-adaptation schemes for energy-efficient routers," In Proc. International Conference on High Performance Switching and Routing (HPSR), pp. 175–182, 2010.

42. G. Semeraro, G. Magklis, R. Balasubramonian, H. D. Albonesi, S, Dwarkadas, and L. M. Scott, "Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling," In Proc. International Symbosium on High-Performance Computer Architecture (HPCA), pp. 29–40, 2002.

43. Amazon EC2 service level agreement, http://aws.amazon.com/ec2-sla/.

44. Google Compute Engine service level agreement, http://cloud.google.com/compute/sla.

45. IBM ILOG CPLEX Optimizer, http://www-01.ibm.com/software/ integration/optimization/cplex-optimizer/.

46. R. Ramaswami and N. K. Sivarajan, “Design of logical topologies for wavelength-routed optical networks,” IEEE J Sel Are Comm, Vol. 14, Iss. 5, pp. 840–851, 1996.

47. Japan Photonic Network Model, http://www.ieice.org/cs/pn/jpn/jpnm.html.

48. L. Xue, C. Cui, and S. Kumar, “Experimental evaluation of the effect of queue management schemes on the performance of high speed TCPs in 10Gbps network environment,” In Proc. International Conference on Computing, Networking, and Communications (ICNC), pp. 315–319, 2012.

49. A. Vishwanath, J. Zhu, K. Hinton, R. Ayre, and R. Tucker, “Estimating the energy consumption for packet processing, storage and switching in optical-IP routers,” In Proc. Optical Fiber Communication Conference (OFC), paper OM3A.6, 2013.

50. M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden, “Part III: Routers with very small buffers,” SIGCOMM Comp Comm rev, Vol. 35, pp. 83–90, 2005.

51. E. Yetginer, T. Uekae, T. I. Kislasi, and N. G. Rouskas, “Power efficient traffic grooming in optical WDM networks,” In Proc. IEEE Global Communications Conference (GLOBECOM), pp. 1838–1843, 2009.

How to cite this article: Y. Fukushima, T. Suda, T. Murase, Y. Tarutani, and T. Yokohira (2022), Minimizing the monetary penalty and energy cost of server migration service, *Trans Emerging Tel Tech*, *2022*.