We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

On the Energy Efficiency of Virtual Machines' Live Migration in Future Cloud Mobile Broadband Networks

Raad S. Alhumaima, Shireen R. Jawad and
H.S. Al-Raweshidy

Abstract

In this chapter, a live migration of the virtual machine (VM) power consumption (PC) model is introduced. The model proposed an easy and parameterised method to evaluate the power cost of migrating the VMs from one server to another. This work is different from other research works found in the literature. It is not based on software, utilisation ratio or heuristic algorithms. Rather, it is based on converting and generalising the concepts of live migration process and experimental results from other works, which are based on the aforementioned tools. The resulting model eventually converts the power cost of live migration from a function of utilisation ratio to a function of server PC. This means there will be neither a need for additional hardware, a separate software, nor a heuristics-based algorithms to measure the utilisation. The resulting model is simple, on the fly and accurate PC evaluation. Furthermore, the latency cost of live migration process, included the time it take the VM to be completely transferred to the target server, alongside the link distance/delay between the two servers is discussed.

Keywords: live migration, power model, power consumption, virtualisation, C-RAN, power cost

1. Introduction

The relentless growing number of connected mobile devices, along with the abundance of new types of bandwidth-hungry applications, has meant high data rate demands and a huge amount of signalling within the core network (CN). According to Cisco, mobile data traffic is expected to intensify by about 11-fold between 2013 and 2018. Furthermore, mobile device connections will grow to about 10.5 billion by 2018 compared to 7.2 billion in 2013 [1–3]. Additionally, Ericsson reportedly forecasts that in 2021, 150 billion devices will be 5G connected, up from 4.100 billion connections using LTE technology. These rising numbers

are alarming and should urge mobile operators to seek out innovative ideas, designs, protocols and advanced digital signal processing (DSP) techniques in order to effectively cope with this explosively high demand for data while simultaneously providing scalable and faster connectivity.

In view of this demand, cloud radio access network (C-RAN) has been suggested by both mobile operators and equipment vendors to introduce cloud computing in 5G cellular networks by pooling the baseband processing units (BBUs) in a shared and centralised data processing centre, known as a BBU pool. In contrast to the legacy eNodeB, the main baseband physical procedures, cooperation and processing of the upper layers in C-RAN are executed in the BBU pool, whereas the simple radio frequency (RF) functions are tackled by the low power consumption (PC) of remote radio heads (RRHs). C-RAN therefore truncates the capital expenditures (CAPEX) and operational expenditures (OPEX) due to lower site leases, reduced maintenance cost and fewer site visits. Other benefits of C-RAN include (i) using advanced DSP, coordination and cognitive radio techniques to process signals through any neighbouring BBU(s) and efficiently utilising the available spectrum; (ii) managing traffic variations by exploiting fewer computing resources and, therefore, not utilising unwanted processors and (iii) reducing cooling PC as well as the total PC.

Recently, the C-RAN has witnessed a rising technology within the pool, which is live migrating from/to the servers. Basically, virtualising the servers has been an efficient way to increase the energy efficiency of the networks by running several BBUs' software using one server [4, 5].

Live migration is an agile concept in nowadays data centres [1]. In virtualisation environment, the interruption-free live migrating of the virtual machines (VMs) form one host server to another is an important issue to sustain running services to the UEs while gaining many benefits. Simply, live migrating of the VM is the movement of one or many VMs for the original host server to another server; this is done when the VM is still running even after it resides in its target server. It is called 'live' since the VM stayed running during the process of migration [4, 6]. Live migration comprises copying memory data on which the VM resides and CPU contents. Practically, an image file is stored in what is called network-attached storage (NAS) rather than the local disc. NAS is accessible by all VMs and operates as HDD drive [7]. This means that physically transferring the local disc is not required. More information about the process of copying a VM can be found in [1]. However, this technique has privileged the date centres with many advantages, as mentioned below:

- Maintenance: this technique represents a solution in case if the source server is required to be decommissioned due to its type promotion. Alongside, it urgently required operating system or hardware maintenance.

- Reachability: the VM usually resides on a host server which is physically located in a certain area. This VM might be serving UEs which are located far away from it. Alternatively, there might be host servers located closer to the UEs. Therefore, such migration will definitely reduce the link delay, channel losses and help improve system administration.

- Load balancing: there might be servers experiencing heavy load due to their position in a dense area or because of the service type they run. In this case, it is beneficial to distribute

the load amongst other servers in the network via migrating the VMs. This is while proportionally considering their processing usage without degrading the performance of the participants [8].

- Off-loading: when the traffic in the network is low, some servers can be selected to switch off so as the network EE is increased. In this case, live migrating the VMs from the chosen servers to other active ones is the solution.

On the other hand, the performance of migration process depends on many factors, such as the memory allocated to the VM, the size of work load it serves and the transmission rate at which the migration is occurring. Eventually, these factors affect the latency of migration and network traffic flow [9]. However, there are increasing disadvantages symbolised by two major aspects:

- Time: the time it takes the process of migration degrades the network performance. The transferred VMs increase latency factor as it means more imposed link delay [10]; this delay means degraded coverage and lower network capacity. In real-time services, this factor is essential and crucial, in contrast to the off-line services where the latency is relieved.

- Energy: the overhead cost of live migration is considerable. Up to 10 W is withdrawn from the destination, and this value increases when the server is the source. This is on the basis of more computations within the tagged server which will be performed in one unit of time [11].

In this chapter, we will try to model these costs which are synchronised with this technology, aiming to speculate the cost of such process prior to migration. The structure of this chapter is as follows: in Section 2, the process of live migration is modelled. In Section 3, the results are presented. Finally, the summary is given in Section 4.

2. Live migration power model

The increased PC cost due to migrating a single or group of VMs to the destination server is also counted. It has been noted that the power cost of the source server $\left(P_{cost}^{source}\right)$ is changeable according to the utilisation of the CPU. We have translated this practical value to more understandable data to avoid the need to measurements' server. First, the extracted cost is redrawn as a function of sever utilisation (utl) of [11] instead of a function of downtime (latency) of migration. **Figure 1** shows the utilisation against the source server power cost.

The original data extracted is curve-fitted to a flexible and simple quadratic Eq. (2), with coefficients, cof1 = 0.0011189, cof2 = −0.25916 and cof3 = 16.315:

$$P_{cost}^{source} = cof1 * utl^2 + cof2 * utl + cof3 \tag{1}$$

However, this description is valid for a server with specific characteristics; there might be slight change in this curve when the type of server is changed. To cover this issue, Eq. (2) is
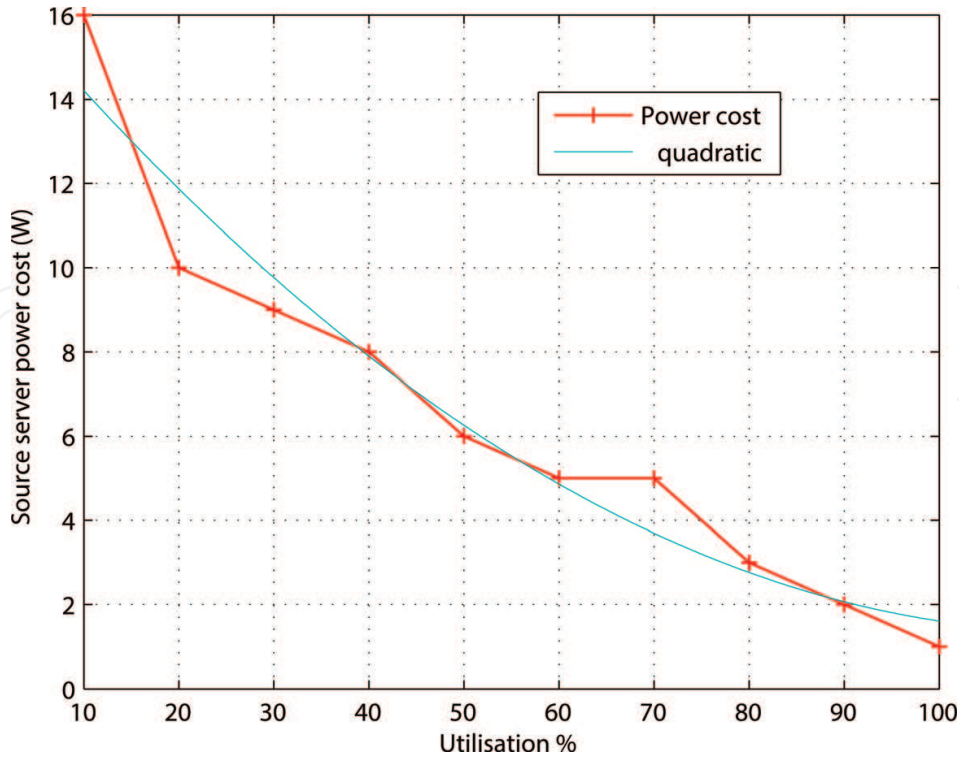
Figure 1. Utilisation against the source server power cost.

generalised by adding a constant called (*scof*), which is a real number. The latter will scale up and down the output of **Figure 1**, so as it matches all possible costs. Therefore, the model in Eq. (2) has been updated, as follows:

$$P_{cost}^{source} = \left(cof1 * utl^2 + cof2 * utl + cof3\right) + scof \tag{2}$$

There has been an objection with measuring the PC using utilisation ratio of a server; the reasons have been mentioned in [5]. Briefly, this method does not offer simplicity for many reasons: it requires real experiment or measurement; it is expensive, because it requires to be physically available at the data centre to observe the utilisation values. On top of that, the service provider (SP) propriety devices are not available to be tested. Therefore, this value is considered, but as a function of maximum and minimum PC of the virtualised server. This means that the real-time measurement power costs which are measured as a function of utilisation ratio are now transferred to a function of PC. To do so, we have used the formula in Eq. (3) to convert the utilisation of a server to a PC. In this case, the power cost of migration can be known directly from the server PC. For example, **Figure 2** shows the utilisation and PC conversions of a server:

$$utl = 100 * \left(P_{server}^{BBU} - min\left(P_{server}^{BBU}\right)\right) / \left(max\left(P_{server}^{BBU}\right) - min\left(P_{server}^{BBU}\right)\right) \tag{3}$$
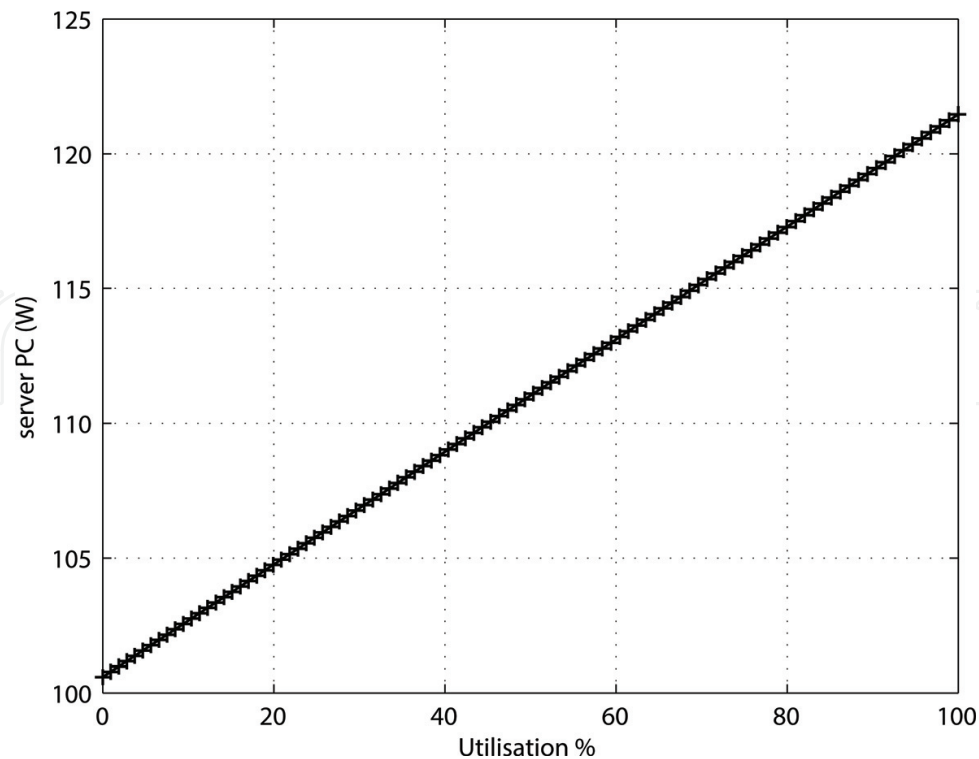
where P_{server}^{BBU} is the BBU server.

Figure 2. Utilisation ratio and its PC equivalent.

The 100% utilisation means that the server is fully utilised and experiences maximum PC $\left(\max\left(P_{server}^{BBU}\right)\right)$, while 0% utilisation means that the server is load-free $\left(\min\left(P_{server}^{BBU}\right)\right)$ or in idle mode of operation. The resulting PC and its equivalent power cost are shown in **Figure 3**.

In the receiving side, the power cost in the destination server $\left(P_{cost}^{dest}\right)$ during the migration is experimentally measured in [11], which was reported as 10 W, regardless of the percentage of utilisation in the server. This assumption is also generalised by adding another constant ($rcof$), which is a real number, so as all possibilities of servers' specifications are considered, as follows:

$$P_{cost}^{dest} = 10 + rcof \tag{4}$$

The total number of source servers undergoing live migration is (S_{smig}), the total number of target servers is (S_{rmig}), the number of migrated VMs is (N_{smig}) and the total number of received VMs is (N_{rmig}). The overall PC of live migration-based virtualised C-RAN data centre (P_{vCRAN}^{mig}) is formulated by adding these costs to the total PC of virtualised C-RAN, which is found in [5]. This yields

$$P_{vCRAN}^{mig} = P_{vCRAN} + \sum_{S_{smig}}^{S_{smig}N_{smig}}\sum_{n_{smig}} \left(P_{cost}^{source}\right)_{S_{smig},\, n_{smig}} + \sum_{S_{rmig}}^{S_{rmig}N_{rmig}}\sum_{n_{rmig}} \left(P_{cost}^{dest}\right)_{S_{rmig},\, n_{rmig}} \tag{5}$$

In terms of time, the live-migrated downtime is relatively large; this process can happen within orders of milliseconds [12] or even orders of seconds. In any case, the higher data rate in which
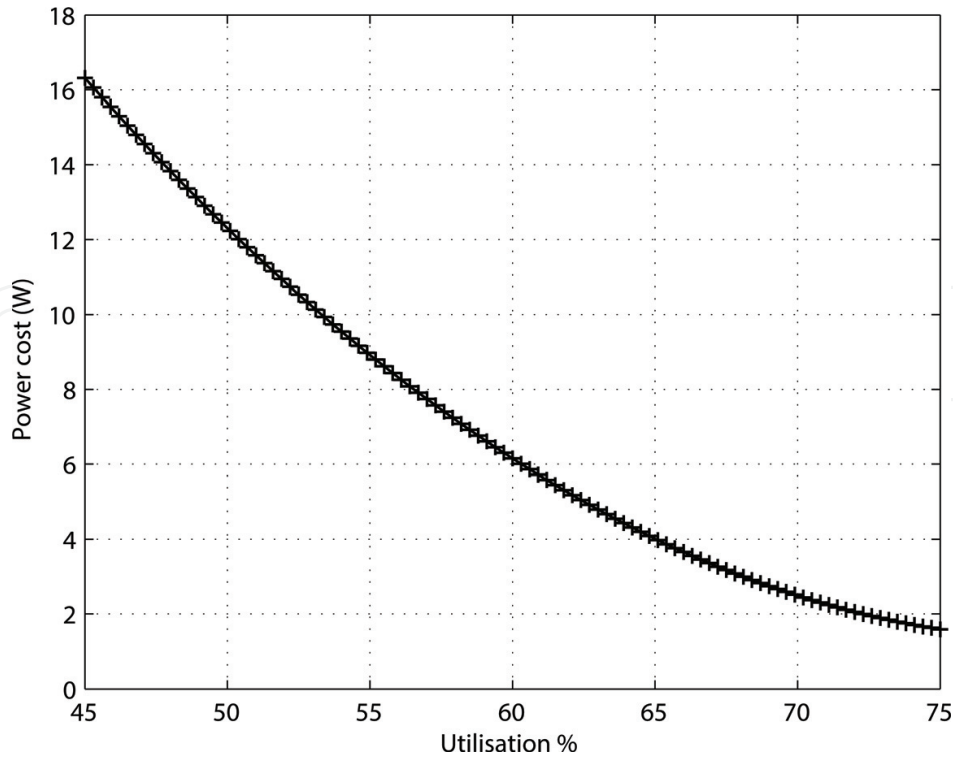
Figure 3. Power cost of live migration and its PC equivalent.

the VM is migrated, the less latency [9, 11]. The latter represents the time it takes the VM to be transferred to other servers seamlessly. As the UEs are still connected while migration, this value does not mean that it is an effective latency to be added to the virtualisation process latency. However, when migrating the VM, the time cost can be established from the channel in which the VM is transferred, which might be wireless or wired [11]. This cost eventually can be added to the modelling of [5].

The time cost due to migrating the VM is (τ_{mig}), which is equivalent to the ($\frac{d_{mig}}{v_{mig}}$), where d_{mig} and v_{mig} denote the distance and speed at which the VM is transferred from the source to target server. v_{mig} is based on the channel type; it can be as fast as the speed light if the channel is wireless, experiencing cable losses if the channel is ethernet via coaxial cable, or can be experiencing a refractive index in case of using an optical fibre channel. However, the effect of τ_{mig} can be major if the VM is moved between distant centres. Therefore, a virtualised data centre experiences a total delay (τ_T) due to virtualisation delay (τ_v) and migration delay τ_{mig}, where

$$\tau_T = \tau_v + \tau_{mig} \tag{6}$$

Approximately, the experimental results in [11] is similar to [9] in terms of power cost and migration time. The latter has measured the cost with respect to the transmission bit rate used to send the VM. While in the former, the power cost is associated with server utilisation. These two measurements are correlated as the highly utilised server means that it is sending using higher bit rate and vice versa. **Figure 4** shows the VM migration bit rate relation with the power cost and migration time.
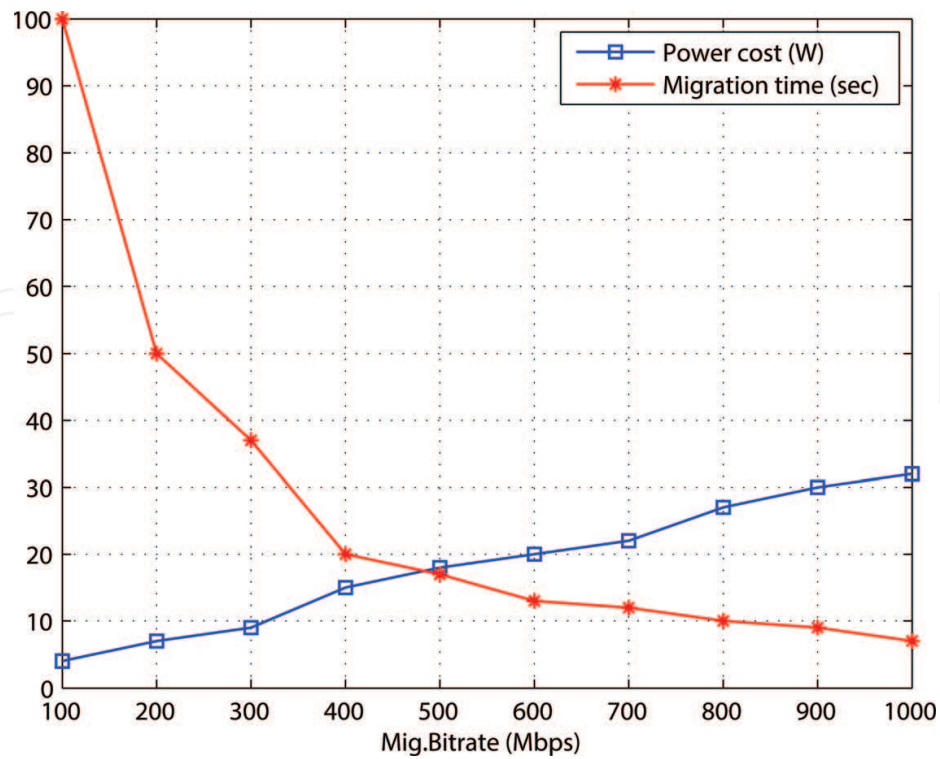
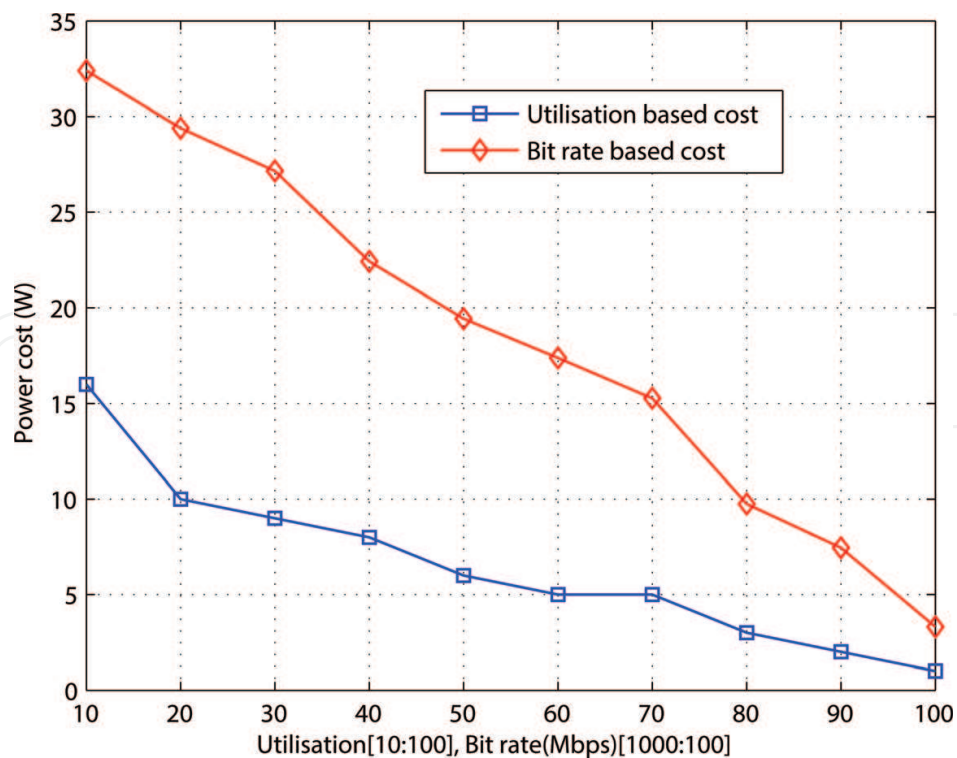**Figure 4.** Migration bit rate as a function of power cost and migration time of [9].

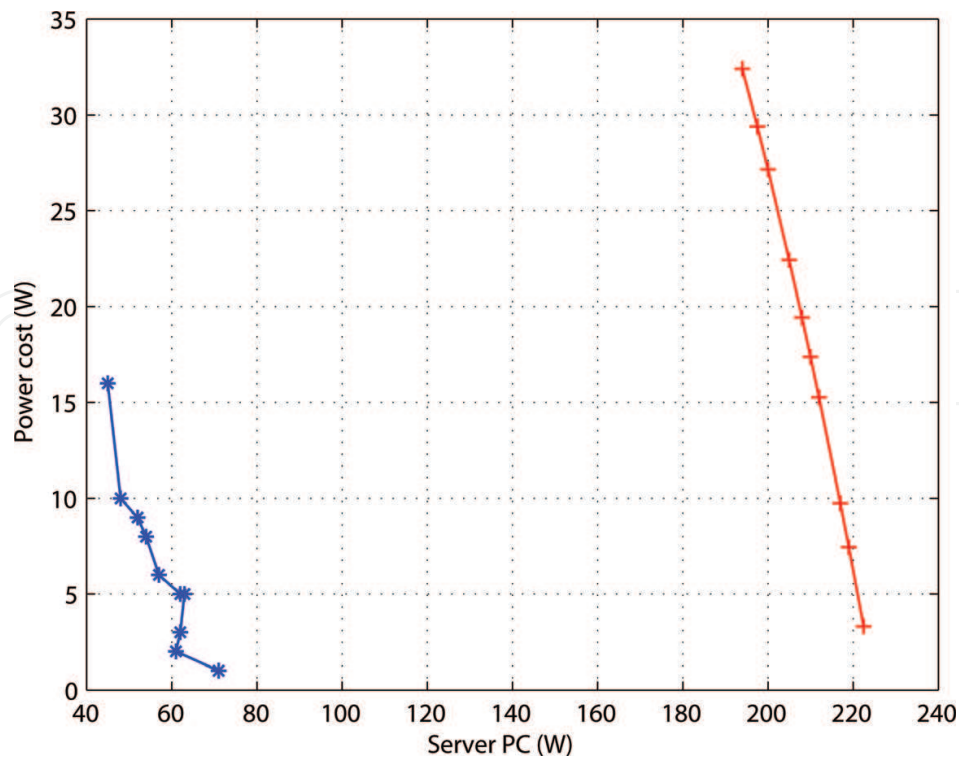**Figure 5.** Difference between the outcomes of two experiments in terms of utilisation and bit rate.

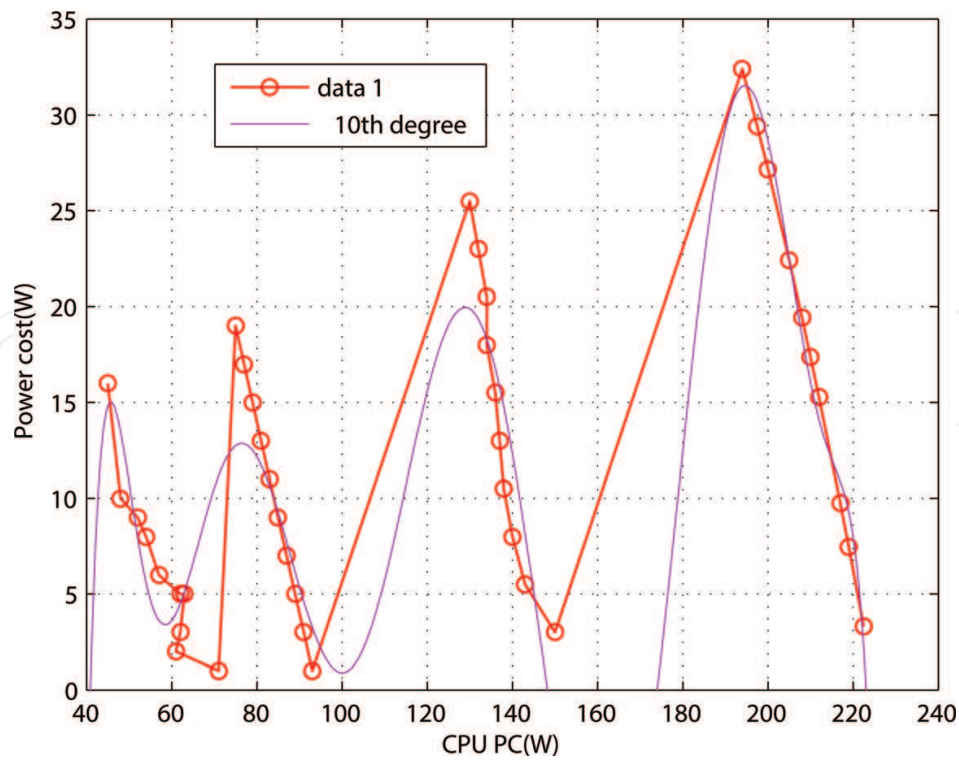**Figure 6.** Difference between the outcomes of two experiments in terms of PC.

Figure 7. Difference between the outcomes of two experiments.

Figure 5 shows the differences between the outcomes of [9, 11] in terms of bit rate and utilisation, respectively. On the other hand, **Figure 6** shows the difference between the two works in terms of PC. However, this difference is originated from the characteristics of the two operated servers. Nevertheless, the behaviour of the outcomes is identical. To extend these models to a general formulation, another two data sets have been added for servers that are commercially available and consume from (130–150 W) to (75–93 W) [13]. Together, there will be four groups of data sets. These latter produce **Figure 7**, which is then poly-fitted using linear regression to produce the following model.

$$
\begin{aligned}
P_{cost}^{source} = {} & p1 * P_{server}^{BBU\ 10} + p2 * P_{server}^{BBU\ 9} + p3 * P_{server}^{BBU\ 8} + p4 * P_{server}^{BBU\ 7} \\
& + p5 * P_{server}^{BBU\ 6} + p6 * P_{server}^{BBU\ 5} + p7 * P_{server}^{BBU\ 4} + p8 * P_{server}^{BBU\ 4} \\
& + p9 * P_{server}^{BBU\ 2} + p10 * P_{server}^{BBU} + p11
\end{aligned}
\tag{7}
$$

with coefficients: $p1$ = -8.6139e-17, $p2$ = 1.1482e-13, $p3$ = -6.6938e-11, $p4$ = 2.2424e-08, $p5$ = -4.7684e-06, $p6$ = 0.00067105, $p7$ = -0.063154, $p8$ = 3.9177, $p9$ = -153.09, $p10$ = 3399.7 and $p11$ = -32561.

Subsequently, the value of P_{cost}^{source} is updated.

3. Results

Figure 8 shows the power cost when a virtualised server holding 60 VMs is migrating 10 VMs; it also shows the power cost due to receiving the same number of VMs. Furthermore, it shows the total cost of both cases. This source cost was derived from the virtualised server consumption as mentioned above. At $N = 60$, the virtualised server PC is known. From this consumption, the power cost is obtained (about 4.9195 W) and multiplied by the number of migrated VMs. On the other hand, 10 W times the number of received VMs is the power cost in the receiving side.

Figure 9 shows the increasing cost because of the increasing number of migrated VMs, up to 50 VMs, while the virtualised server already hosted 60 VMs.

As we can see from **Figures 8** and **9**, migrating the VMs is not power-effective process. To migrate and receive 10 VMs, the power cost is about 150 W, which is more than the original server PC. However, this cost is influenced by the period at which these VMs are transferred. Intuitively, the longer period these VMs are migrated, the less power cost and more efficient system. Therefore, algorithms/methods are needed to optimise at what time the VM is required to be moved. This can be based on several parameters, such as the number of UEs connected, load balancing requirement, position of the servers, etc.

Figure 9 exhibits the cost and total PC when the number of both migrated and received VMs is equal. However, sending and receiving a VM by more than one server at the same time represent another facility that can be added and offered by the model. This is when the number of migrated and received VMs is different. **Figure 10** then shows a three-dimensional plot, and
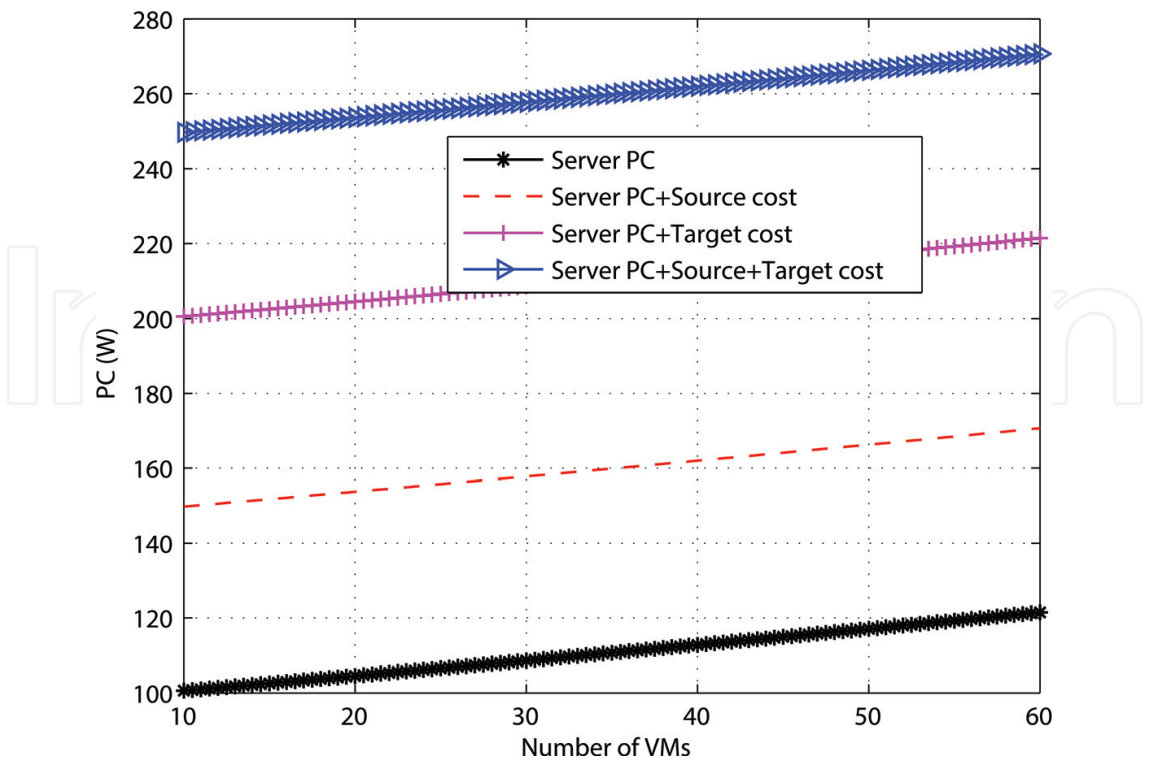
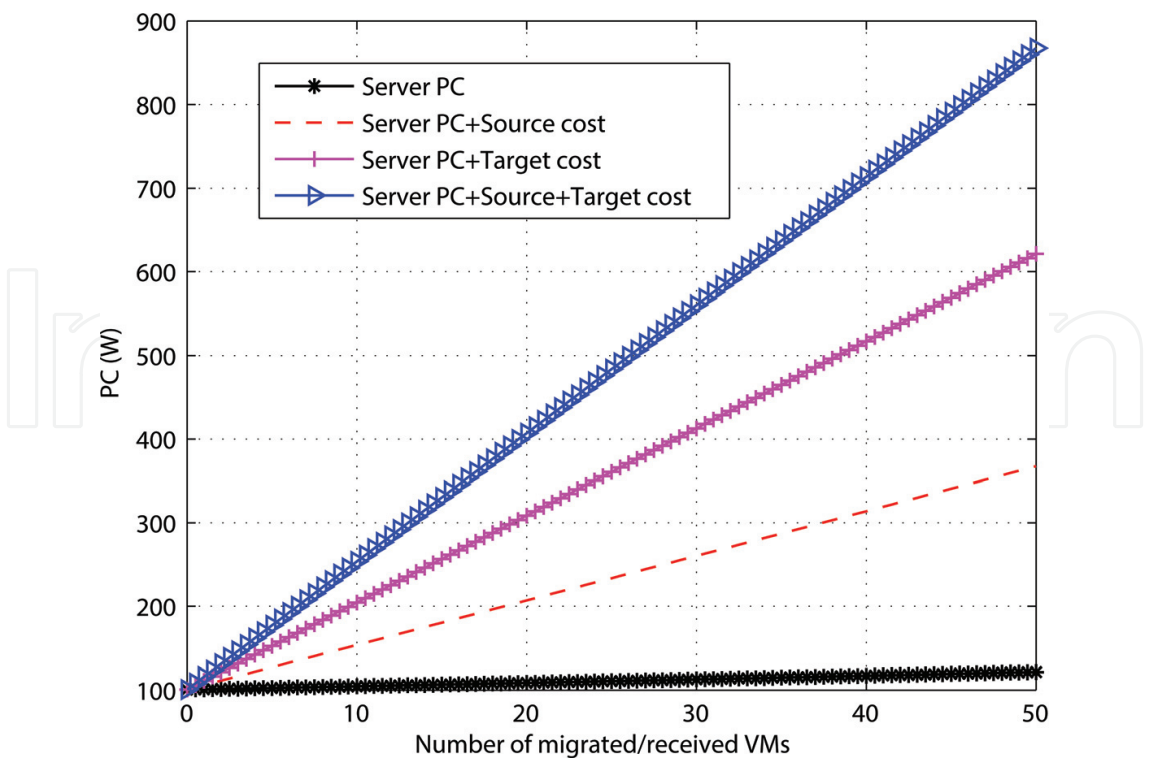**Figure 8.** PC of source server hosting 60 VMs; it shows the power cost of migrating 10 VMs.

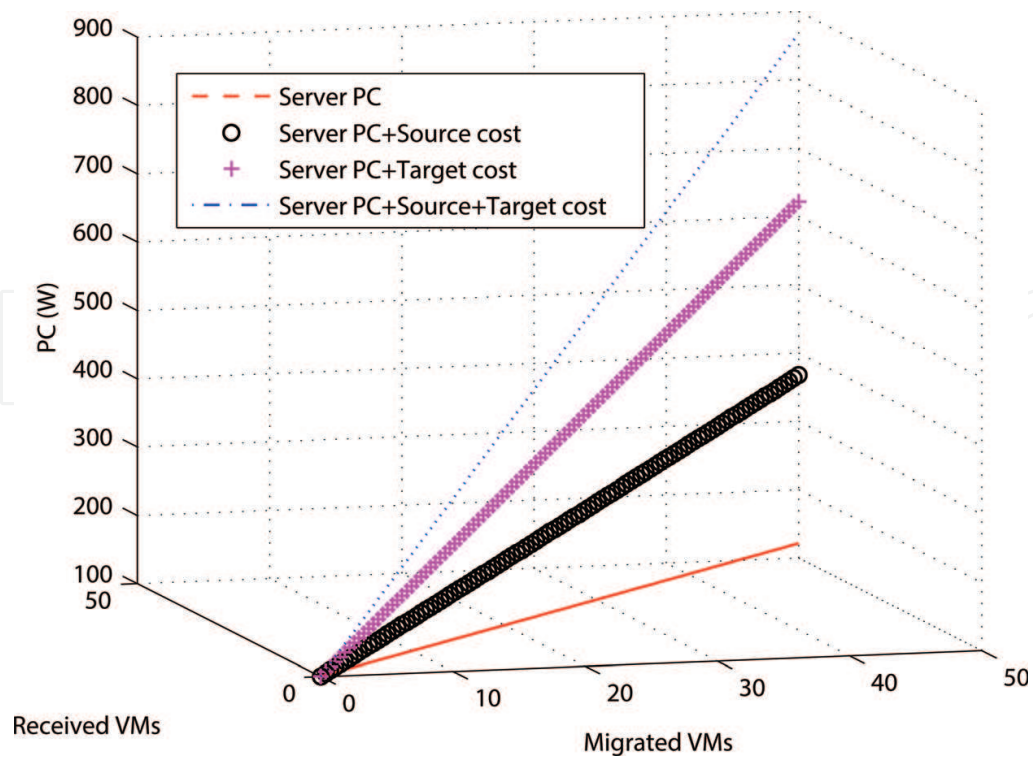**Figure 9.** PC of source server hosting 60 VMs; it shows the power cost of migrating 10 VMs.

Figure 10. PC of source server hosting 60 VMs; it shows the power cost of migrating 10 VMs.
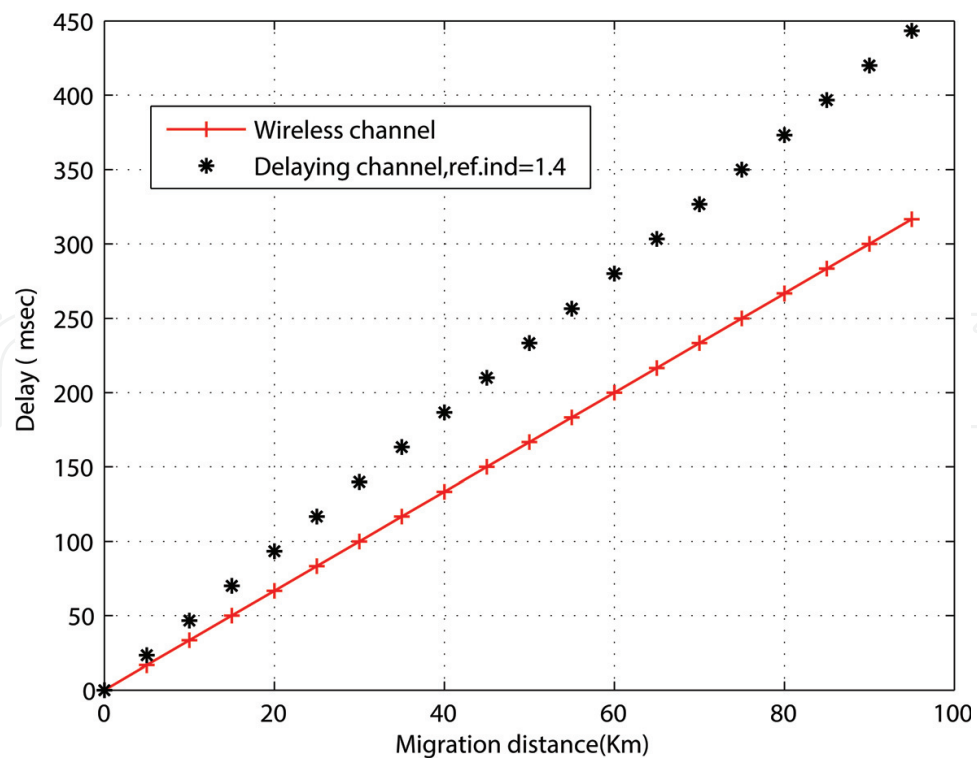

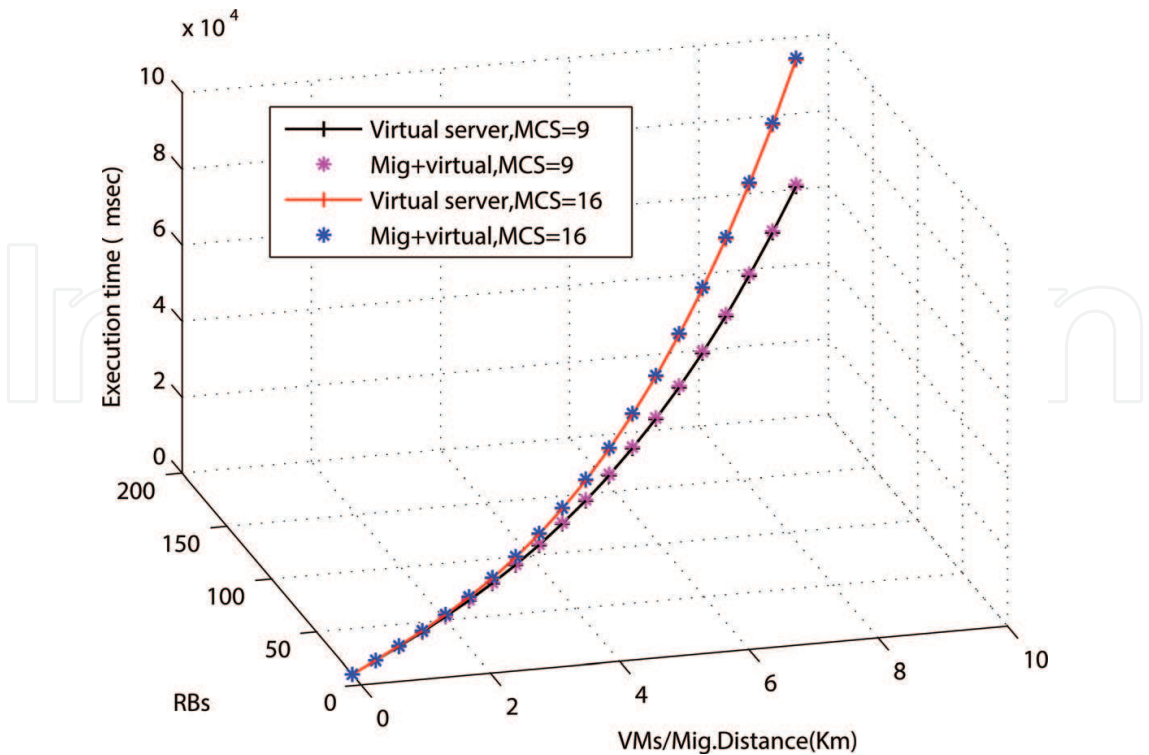**Figure 11.** Delay comparison of wireless and delaying channel.

Figure 12. Total delay of virtualised server while migrating a VM to different distances.

x and y axes represent the number of migrated and received VMs, while z axis shows the PC. This figure shows the cost of migrating or receiving 50 VMs; also it shows the cost of both cases. The virtualised server holds 60 VMs.

Figure 11 shows the channel delay experienced due to migrating a VM to a distant data centre. This figure compares wireless with lossy channel with refractive index = 1.4.

However, in both cases, wireless or delaying channel, this amount of delay is neglected compared to the virtualisation process delay. **Figure 12** shows the total delay of virtualisation and migration. This covers the wireless, up to an optical fibre channel with refractive index which is equivalent to 1.4. This is when the virtualised server is hosting 10 VMs, each processing 200 RBs and migrating 1 VM to variable distances (up to 100 km).

4. Summary

A model has been presented to demonstrate the PC cost in a virtualised based data centres, specifically for live migration case. By using the proposed model, the power and time cost can be assessed. Since there is no simple expression to describe such costs, this model has been proposed. It has converted the experimental results which are based on the server utilisation and migration bit rate to a simple-to-adapt mathematical formulation. The model enables the network providers to decide about the EE of such technology in the data centres and virtualised servers. The cost of repeated migration can cause a considerable amount of power lost. Nevertheless, by using a quick

and simple calculation of the model, a decision about when and where the migration occurs can be easily made. On the other hand, the expected time cost due to migration has been shown. However, the time cost is negligible compared to the virtualisation. This time was the cost of the distance between the two cooperated servers, while the time of the migration process itself can reach more than 7 s. This was not counted within this modelling since the VM is not terminated from serving the attached UEs and it is still live.

Author details

Raad S. Alhumaima*, Shireen R. Jawad and H.S. Al-Raweshidy

*Address all correspondence to: raad_1990_1990@yahoo.com

WNCC, School of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge, Middlesex, UK

References

[1] Strunk A, Dargie W. Does live migration of virtual machines cost energy?. In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA); March 2013. pp. 514-521

[2] Alhumaima RS, Al-Raweshidy HS. Evaluating the energy efficiency of software defined-based cloud radio access networks. IET Communications. 2016;**10**(8):987-994

[3] Alhumaima RS, Khan M, Al-Raweshidy HS. Component and parameterised power model for cloud radio access network. IET Communications. 2016;**10**(7):745-752

[4] Zhang J, Ren F, Lin C. Delay guaranteed live migration of virtual machines. In: INFOCOM, 2014 Proceedings IEEE: IEEE; 2014. pp. 574-582

[5] Alhumaima RS, Al-Raweshidy HS. Modelling the power consumption and trade-offs of virtualised cloud radio access networks. IET Communications. 2017;**11**(7):1158-1164

[6] Zheng J, Ng TSE, Sripanidkulchai K, Liu Z. Pacer: A progress management system for live virtual machine migration in cloud computing. IEEE Transactions on Network and Service Management. December 2013;**10**:369-382

[7] Ruan Y, Cao Z, Cui Z. Pre-filter-copy: Efficient and self-adaptive live migration of virtual machines. IEEE Systems Journal. Dec 2016;**10**:1459-1469

[8] Agarwal A, Shangruff R. Live migration of virtual machines in cloud. International Journal of Scientific and Research Publications. 2012;**2**(6):1-5

[9] Liu H, Jin H, Xu C-Z, Liao X. Performance and energy modeling for live migration of virtual machines. Cluster computing. 2013;**16**(2):249-264

[10] Strunk A. Costs of virtual machine live migration: A survey. In: 2012 IEEE Eighth World Congress on Services. June 2012. pp. 323-329

[11] Huang Q, Gao F, Wang R, Qi Z. Power consumption of virtual machine live migration in clouds. In: Communications and Mobile Computing (CMC), 2011 Third International Conference on; April 2011. pp. 122-125

[12] Clark C, Fraser K, Hand S, Hansen JG, Jul E, Limpach C, Pratt I, Warfield A. Live migration of virtual machines. In: Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation - Volume 2, NSDI'05, (Berkeley, CA, USA), pp. 273-286, USENIX Association; 2005

[13] Buildcomputers, Power Consumption of PC Components in Watts (http://www.buildcomputers.net/power-consumption-of-pc-components.html), 2017 (accessed March 19, 2017)