Estimation of the Cost of VM Migration

Waltenegus Dargie
Chair of Computer Networks
Faculty of Computer Science
Technical University of Dresden
01062 Dresden, Germany
Email: waltenegus.dargie@tu-dresden.de

***Abstract*—One of the mechanisms to achieve energy efficiency in virtualized/cloud environments is consolidation of workloads on an optimal number of servers and switching-off of idle or underutilized servers. Central to this approach is the migration of virtual machines at runtime. In this paper we investigate the cost (migration time) of virtual machines migration. We shall show that migration time exponentially increases as the available network bandwidth decreases; migration time linearly increases as the RAM size of a virtual machine increases. Furthermore, the power consumption of both the destination and the source servers remain by and large the same for a fixed network bandwidth, regardless of the VM size. Interestingly, for the same combination of virtual machines, different orders of migrations resulted in different migration time. We observed that migrating resource intensive virtual machines first yields the shortest migration time. In general, the migration time should be modeled as a random variable since the factors that affect it cannot be known except in a probabilistic sense. Therefore, we propose a probabilistic approach to quantify the cost of virtual machines migration.**

***Index Terms*—Cloud computing, energy-efficient computing, migration cost, migration time, server consolidation, virtual machine migration, workload consolidation**

I. Introduction

One of the advantages of cloud computing is its flexible resource configuration [1]. From the service provider point of view this feature enables to dynamically expand and shrink resource demand, ideally tailoring the cost of leased resources to the anticipated workload [2]. Similarly, the infrastructure provider can avoid the inefficient use of resources, because idle or underutilized resources can be switched off [3].

Workload consolidation (or server consolidation) [4], [5], [6] is one of the mechanisms to efficiently utilize resources in cloud computing. In this scenario, the infrastructure provider regularly monitors the size of incoming workload and the distribution of workload within the cloud environment to determine the number of servers it should make available. Because of the dynamic nature of received workloads, some of the existing servers can be overloaded whereas other can be underutilized [7]. To balance the load distribution within the cloud and to avoid underutilization, workload consolidation can take place. In this process, virtual machines (VM) encapsulating services or applications [8], [9] can be migrated from one server to another without actually stopping the applications [10], [11]. However, VM migration introduces execution latency to the applications and energy overhead to the infrastructure provider [12], [13].

The cost of migration depends on many factors including (1) the memory content and the memory update rate of each virtual machine, (2) the total number of virtual machines to be migrated, (3) the available network bandwidth for migration, and (4) the workload of the source and the destination servers at the time of migration.

Considering the relative short duration required to migrate virtual machines, it can be assumed that the available network bandwidth during migration remains unchanging. Interestingly, for a fixed network bandwidth, experiment results show (refer to Section III) that the power consumption of both the source and the destination servers also remains fairly the same during VM migration. Consequently, in order to estimate the cost of service migration, it is sufficient to set the available network bandwidth to a fixed value and:

- to express the cost of migration in terms of (1) and (2) (above); and,
- to determine the migration time only, since this term affects the migration latency as well as the energy overhead of migration.

In this paper, we first experimentally analyze and then quantitatively express the relationships between the migration time and the size of virtual machines. The study is useful to quantitatively investigate the cost and gain of workload (server) consolidation in virtualized (cloud) environments. Moreover, we shall show that:

- The migration time exponentially decreases as the communication bandwidth increases.
- The average power consumption of the source as well as the destination server linearly increases as the network bandwidth increases.
- The migration time and the energy consumption of both the source and the destination servers linearly increase as the size (memory content) of the virtual machine increases.

The remaining part of this paper is organized as follows: In Section II, we introduce the experiment environment we set up to examine the cost of virtual machine migration. In Section III, we discuss the experiment results and identify the important factors that influence virtual machine migration. In Section IV, we present a quantitative (statistical) approach to estimate the cost of migration. Finally, in Section V, we provide concluding remarks and outline our future work.

978-1-4799-3572-7/14/$31.00 ©2014 IEEE

II. Experiment Setting

A. Live Migration

Live migration enables a virtual machine to be physically moved from one server[1] to another in a transparent fashion without actually stopping none of the applications and services being executed inside the virtual machine [10], [11], [12], [13]. The current virtualization technology (based on hypervisors) requires a network attached storage (NAS) to store the images of all the virtual machines that can be migrated. This way the actual content that should be exchanged between a source and a destination server during the migration of a virtual machine is limited to the in-memory state and the content of the CPU registers.

The most frequently employed migration approach is the so-called "pre-copy" approach [10], [14], [15] which consists of the following three phases:

1) *Pre-Copy Phase*: At this stage, the VM continues to run while its memory content is iteratively copied page by page from the source server to the destination server. As each round takes some amount of time, some of the memory pages on the source server may be modified and may no longer be in sync with the copy version on the destination server. These pages have to be re-transmitted to ensure memory consistency.
2) *Pre-Copy Termination Phase*: Without any stop condition, the pre-copy phase may go on indefinitely. While a stop condition depends on the specific hypervisor implementation, it typically takes one of the following thresholds into account: (1) the number of iterations exceeds a pre-defined threshold ($n > n_{th}$), (2) the total amount of memory that has already been transmitted exceeds a pre-defined threshold ($m > m_{th}$), or (3) the number of pages modified in the previous round drops below a pre-defined threshold ($p < p_{th}$).
3) *Stop-and-Copy Phase*: At this stage the hypervisor suspends the VM to prevent further page writing and copies the remaining modified pages as well as the state of the CPU registers to the destination host. After the migration process is completed, the hypervisor on the destination host resumes the VM.

B. Setting

For our experiment setting we employed two homogeneous servers and a network attached storage, all of them connected with each other via a Gigabit Ethernet switch. The two servers run Fedora 15 (Linux kernel v. 2.6.38, x86_64) as an operating system and KVM[2] as a hypervisor. Each server employs two Intel i5-680 dual core 3.6 MHz processors, 4 GB DDR3-1333 SDRAM memory and a Gigabit Ethernet Network Interface Card (NIC). The NAS system consists of Intel Xeon E5620 Quad-Core 2.4 GHz processor, 10 GB DDR3-1333 SDRAM memory and Gigabit Ethernet NIC.

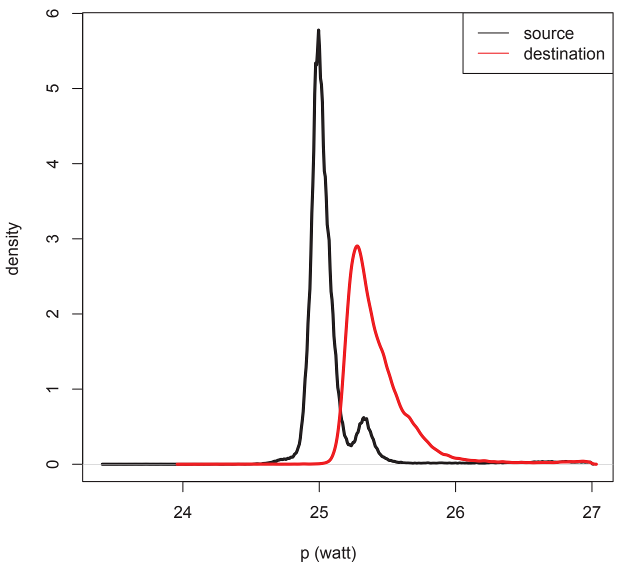

Fig. 1. The density functions of the power consumption of the source and the destination servers in their idle state (no virtual machine was running on the servers).

For each experiment, we run a virtual machine on one of the servers and migrated it to the other server while the virtual machine executes a custom-made μ-benchmark. We varied the RAM size of the benchmark from 800 to 2000 MB in step of 100 MB. The virtual machine has a single virtual processor and runs Fedora 15 as its operating system. A client machine connected to the network initiates the live migration. We varied the available network bandwidth during migration from 10 to 100 MBps (Megabyte per second) in step of 10 MBps. For each configuration, we repeated the experiment 25 times. We introduced a 30 s idle state between iterations to avoid experiment bias.

We employed two Yokogawa WT210 digital power analyzers to measure the overall AC power consumption of both servers. The devices can measure DC as well as AC power consumption at a rate of 10 Hz and a DC current between 15 μA and 26 A with an accuracy of 0.1 %.

The aim of the experiment is (1) to examine the contribution of network bandwidth and the RAM size of virtual machines to the cost of migration – migration time, power consumption, and energy overhead and (2) to quantitatively express the cost of migration in terms of these quantities. We assume that the purpose of migration is to switch off idle or underutilized servers. Therefore, we do not give emphasis on the *type* of workload the virtual machines execute, since, by assumption, these virtual machines do not appreciably utilize the server's resources prior to migration.

III. Experiment Results

Fig. 1 displays the (unscaled) density function of the power consumption of the source and destination servers at an idle state. The figure shows that even at an idle state, the power consumption of both servers is neither deterministic nor constant. Hence, it should be considered as a random variable. Interestingly, for both servers, the idle power consumption

[1] In the context of this paper, the term *server* refers to a physical machine.

[2] http://www.linux.com/directory/Software/applications/kernel-based-virtual-machine

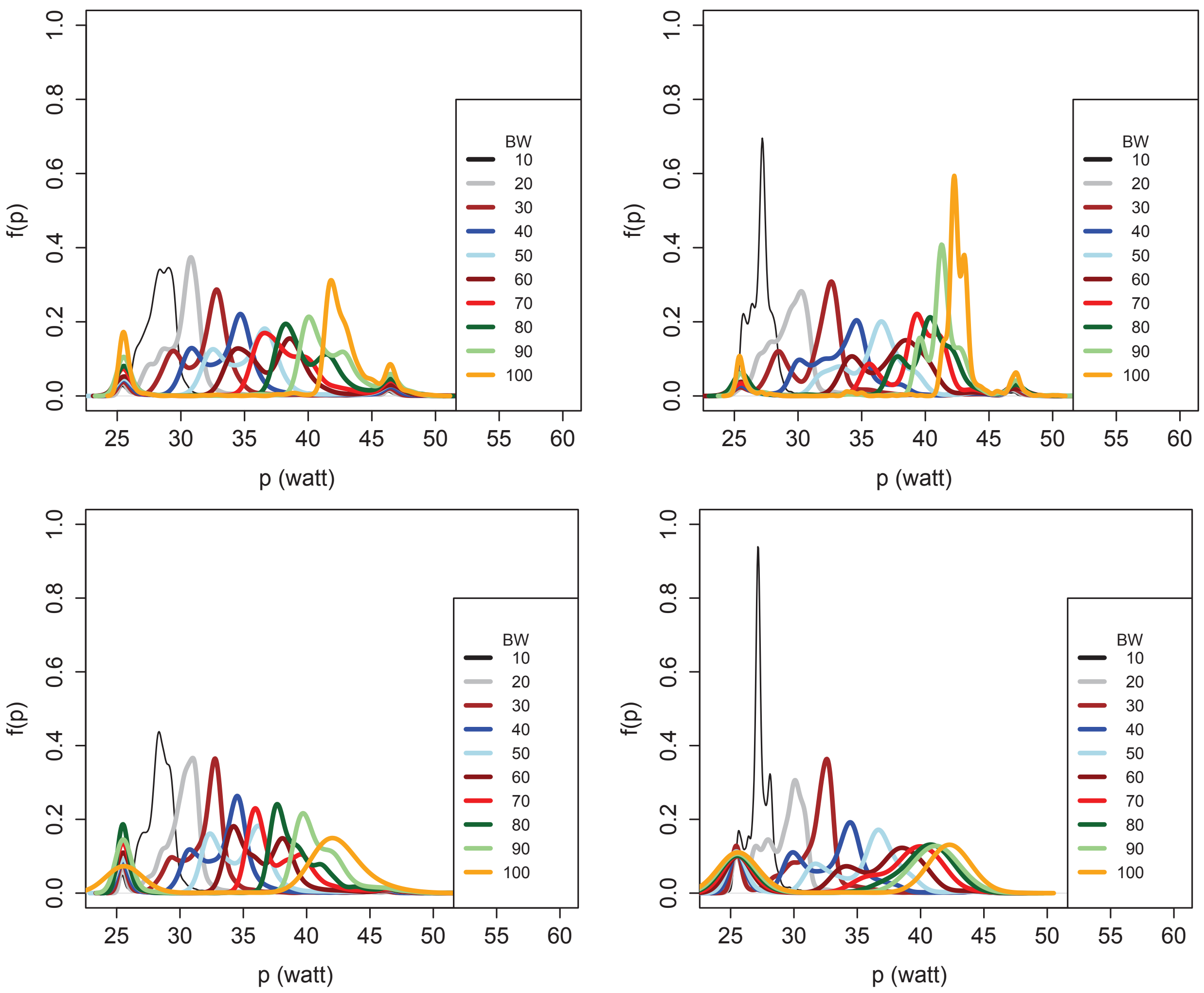

Fig. 2. The power consumption density function, $f(p)$, of the source (left) and destination (right) servers during the migration of virtual machines as a function of network bandwidth (in megabyte per second). The RAM size of the virtual machines was 800 MB (top) and 2 G MB (bottom).

can be modeled as a normally distributed random variable. Secondly, the figure reveals that even if the two servers have identical technical specifications, their idle power consumption is different both in magnitude and in a statistical sense. The average idle power consumption of the source server is always smaller than the destination server. Moreover, the variance of the destination server is greater than the variance of the source server. This simple example illustrates that no two servers can be considered homogeneous in a strict sense.

A. *Network bandwidth*

Fig. 2 shows the density functions of the power consumption of the source and the destination servers during the live migration of idle virtual machines. Repeated migrations were carried out by varying the available network bandwidth from 10 megabyte per second to 100 megabyte per second and the RAM content of the virtual machines from 800 MB to 2 GB. During a single migration, both the size of the VM and the network bandwidth were not changing appreciably. The average CPU utilization of the μ-benchmark running on the virtual machines was below 7%. Therefore, it is reasonable to assume that the migration latency as well as a significant portion of the extra power and energy consumptions of the two servers were due to the migration of the RAM content of the virtual machines.

As can be seen, the power consumption of both the source and the destination servers increased as the available network bandwidth between them increased, because the faster the communication between the two servers the busier the hypervisors become coordinating the VM migration. Interestingly, if we ignore the idle component of the power consumption, the density functions can be approximated by a normal distribution, in which case it is sufficient to express the power consumption by its mean and variance (standard deviation). Furthermore, the mean power consumption increased almost exponentially with an increase in the available network bandwidth.

B. *VM RAM Size*

Fig. 3 displays the density functions of the power consumption of the source and the destination servers for a fixed network bandwidth and variable RAM size. In contrast

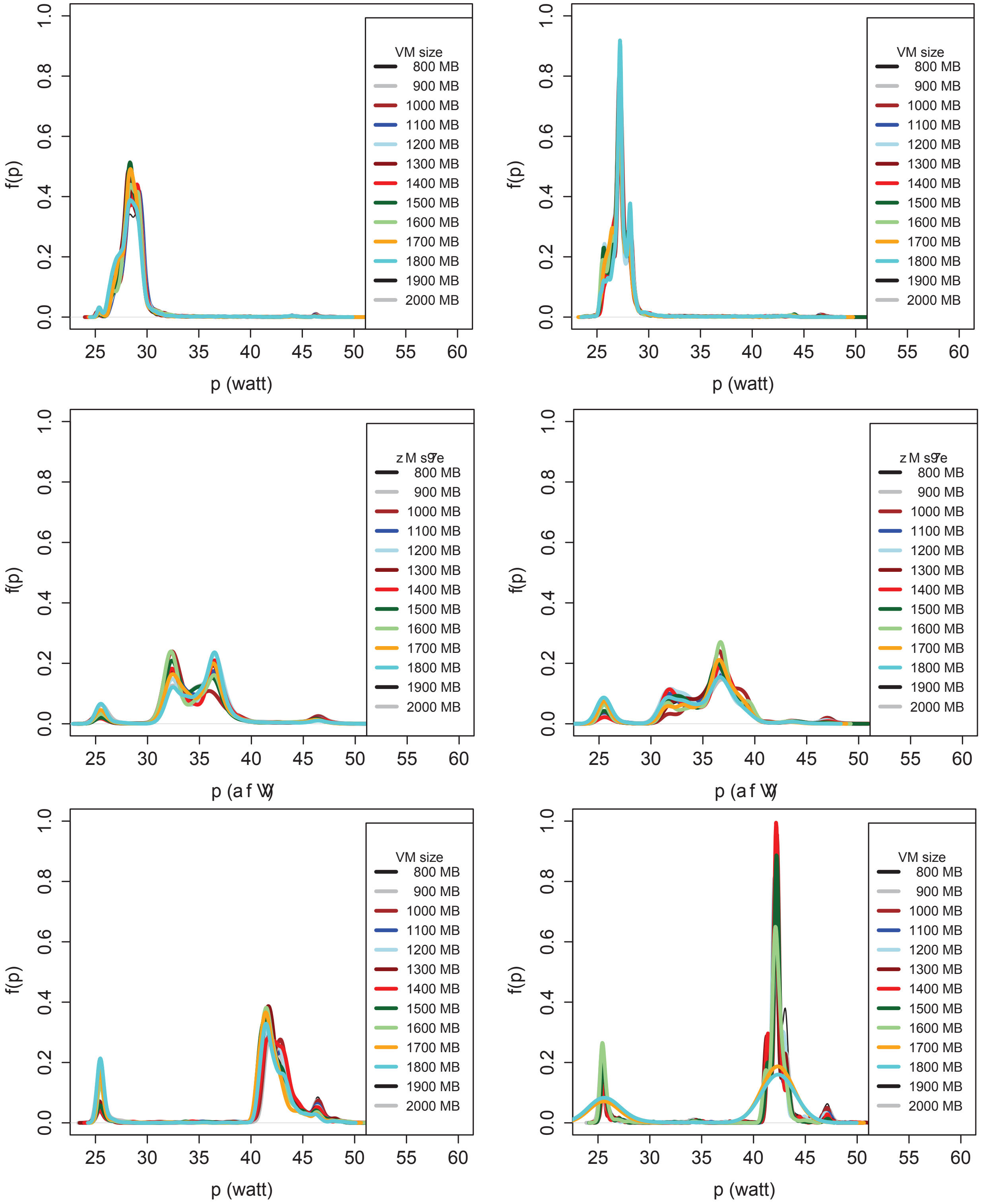

Fig. 3. The power consumption density function $f(p)$ of the source (left) and destination (right) servers during the migration of virtual machines. The available network bandwidth (from top to bottom) was 10, 50, and 100 megabyte per second.

to Fig. 2, the statistics of the power consumption of both servers remain almost the same as we varied the RAM size of the virtual machines. This does not mean, however, that the energy overhead is the same for all the cases. Understandably, the migration time increases as the VM content increases, increasing the energy consumption (which is a function of time) of migration. However, the figures reveal that for a fixed network bandwidth, the power consumption of neither the source nor the destination server plays a role in determining the migration latency of live migration.

C. Migration time (Latency)

The time required for an underutilized or idle server to migrate all the virtual machines it hosts before it can be switched off is an important quantity, because it can affect the infrastructure provider as well as the service provider. As far as the infrastructure (cloud) provider is concerned, the question "when and for how long should a server be switched off?" can be answered by taking the following parameters in to consideration: the time required to empty and switch-off the server as well as to switch-on the server when the workload of the cloud increases, demanding additional resources. For a fast-changing workload, the switch-off duration can be affected by the migration time.

As far as the service provider is concerned, migration can degrade the performance of individual services, because the virtual machine monitors on both servers require extra resources to coordinate the migration. If the migration duration is long, then the performance degradation may not be acceptable to the application users.

Fig. 4 displays the relationship between migration time, network bandwidth, and the RAM size of a virtual machine. The migration time increases linearly with the RAM size of a virtual machine. The migration time decreases exponentially as the available network bandwidth increases linearly. It must be remembered that the power consumption of both servers increases exponentially as the available network bandwidth increases linearly. Hence, there is a trade-off between decreasing migration time and increasing power consumption.

IV. Quantitative Cost Estimation

The number of active virtual machines a server in a cloud environment hosts at any given time cannot be known in advance, since service providers may come and go freely. In addition, the cloud provider may freely move virtual machines from one server to another for reasons of load-balancing, fault tolerance, efficient resource management, etc. Therefore, it may not be possible to know the exact number of virtual machines (**n**) at the time workload consolidation takes place. Similarly, the size of each of these virtual machines (**s**) is a dynamic quantity, because their workload fluctuates overtime.

Consequently, the time required to migrate virtual machines (**t**) and the extra energy consumed by both the source ($\mathbf{e}_s$) and the destination servers ($\mathbf{e}_d$) during virtual machine migration should be modeled as random variables.

Compared to the idle energy consumption, the energy overhead introduced during virtual machine migration is indeed small and can be disregarded [12], [13]. But the same cannot be said of the migration latency, because the migration latency can affect both the cloud infrastructure as well as the QoS the service provider wishes to attain.

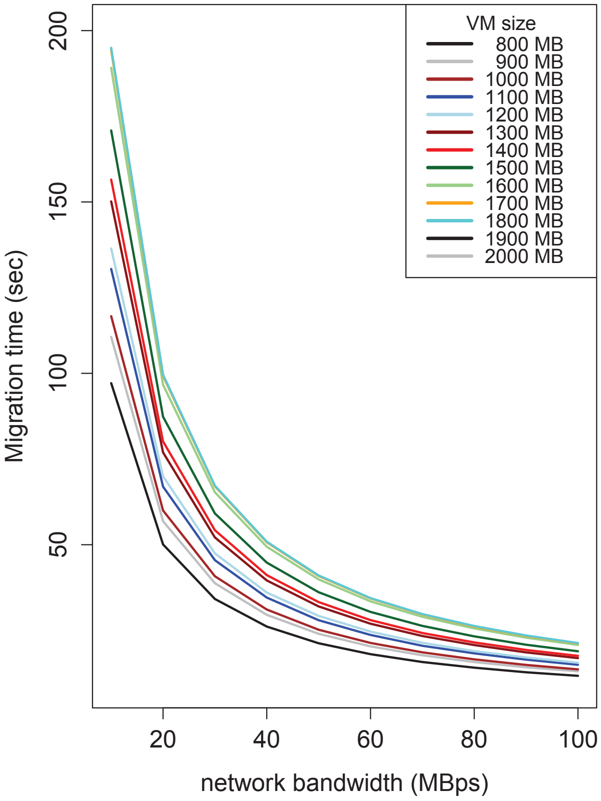

Fig. 4. Migration latency as a function of the network bandwidth and the RAM size of a virtual machine.

The time required to switch off an underutilized or an idle server[3] depends on the time required to migrate all the virtual machines it hosts. Because this time depends on quantities which are random by nature, it should also be considered as a random variable. Consequently, its statistical properties (density function, distribution function, expected value, variance, and autocorrelation) depend on the statistical properties of the quantities that influence it.

To make our analysis tractable, we assume that the available network bandwidth during migration does not change appreciably. As a result, the statistics of the total migration time depend on two random variables, namely, the number of virtual machines and the RAM size of each of these virtual machines.

In the following subsections we use boldface small letters to represent random variables. The distribution of a random variable is represented by F and its density by f. In most practical cases, the two most significant properties of a random variable are the expected value and the variance. We assume that the reader is familiar with the definition of these two properties.

[3] Because the same argument can apply for an overloaded server, we only pursue the case of underutilized server in the subsequent sections.

A. Single Virtual Machine

If there is only a single virtual machine at the time of migration, the cost of migrating this virtual machine can be expressed in terms of the statistics of the RAM size of the virtual machine:

$$\mathbf{t} = f(\mathbf{s}) \tag{1}$$

If we assume a linear relationship between **t** and **s**[4]:

$$\mathbf{t} = a\mathbf{s} + b \tag{2}$$

Then, we can express the statistics of **t** in terms of the statistics of **s**. For example, the probability distribution function (CDF) of **t** is expressed as: $F(t) = P\{\mathbf{t} \leq t\}$ where **t** is the migration latency defined as a random variable and t is a specific size. But using Equation 2, we can expressed the CDF of **t** in terms of **s** because:

$$F(t) = P\{\mathbf{t} \leq t\} = P\{a\mathbf{s} + b \leq t\} = P\left\{\mathbf{s} \leq \frac{t-b}{a}\right\} \tag{3}$$

which is the CDF of **s** expressed in terms of t: $F_s\left(\frac{t-b}{a}\right)$.

Likewise, the probability density function of **t** can be expressed in terms of the probability density function of **s**:

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}F_s\left(\frac{t-b}{a}\right) = \frac{1}{a}f_s\left(\frac{t-b}{a}\right) \tag{4}$$

If we are interested in the expected migration latency, then, it suffices to know the average size of the virtual machine, because:

$$E[\mathbf{t}] = E[a\mathbf{s} + b] = aE[\mathbf{s}] + b \tag{5}$$

Similarly, the variance of **t** can be expressed in terms of the variance of **s**, since:

$$\sigma_t^2 = E\left[(\mathbf{t} - \eta_t)^2\right] = E\left[\mathbf{t}^2\right] - (\eta_t)^2 \tag{6}$$

where η_t is the expected migration latency. Expressing η_t in terms of the expected size of the virtual machine, η_s, reduces Fig. 6 to:

$$\sigma_t^2 = a^2\sigma_s^2 \tag{7}$$

B. Two Virtual Machines

If we have two virtual machines to migrate, the migration cost **t** is the summation of the migration time of the individual virtual machines. This is because existing virtual machine monitors can migrate one virtual machine at a time:

$$\mathbf{t} = \mathbf{t}_1 + \mathbf{t}_2 \tag{8}$$

To get the statistical properties of **t**, we need the joint probability density function $f(t_1, t_2)$. If the two virtual machines are statistically independent, then, the probability density function of **t** can be expressed as:

$$f(t) = \int_0^{\infty} f(t_1)\, f(t - t_1)\, dt_1 \tag{9}$$

where the second term of the integrand is the density of $\mathbf{t}_2$ expressed in terms of $t - t_1$.

If we are interested in the expected value of **t** only, then, it suffices to know the expected values of $\mathbf{t}_1$ and $\mathbf{t}_2$:

$$E[\mathbf{t}] = E[\mathbf{t}_1 + \mathbf{t}_2] = E[\mathbf{t}_1] + E[\mathbf{t}_2]$$

The expected values of $\mathbf{t}_1$ and $\mathbf{t}_2$ in turn can be expressed using Equation 5. Similarly, the variance of **t** can be expressed in terms of the variance of $\mathbf{t}_1$ and $\mathbf{t}_2$:

$$\sigma_t^2 = \sigma_{t_1}^2 + \sigma_{t_2}^2 \tag{10}$$

C. Multiple Virtual Machines

In general, the number of virtual machines that should be migrated cannot be known in advance (before the decision to migrate them was made). Hence, the total migration time can be expresses as follows:

$$\mathbf{t} = \sum_{i=1}^{\mathbf{n}} \mathbf{t}_i \tag{11}$$

where $\mathbf{t}_i$ is the time needed to migrate the i-th virtual machine and **n** and $\mathbf{t}_i$ are statistically independent for all i.

We need the n-dimensional joint density function, $f(t_1, t_2, ..., t_n)$, to evaluate Equation 11, which is not easy to obtain. If, however, we are interested in the expected value and the variance of **t**, then we can evaluate Equation 11 first as a conditional mean by fixing $\mathbf{n} = n$ and then by computing the expected value of the conditional mean with respect to **n**[5]. Hence,

$$E[\mathbf{t}|\mathbf{n} = n] = E\left[\sum_{i=1}^{n} \mathbf{t}_i\right] \tag{12}$$

Notice that the upper bound of the summation expression is no longer a random variable because we are calculating the conditional expected value, $E[\mathbf{t}|\mathbf{n} = n]$. Once n is fixed, the conditional migration time depends only on the statistics of $\mathbf{t}_i$.

Suppose $\mathbf{t}_i$ are independent, identically distributed (i.i.d.) random variables with mean η_i and variance σ_i^2. In this case the conditional expected value of **t** can be expressed as follows:

$$E[\mathbf{t}|\mathbf{n} = n] = E\left[\sum_{i=1}^{n} \mathbf{t}_i\right] = \sum_{i=1}^{n} E[\mathbf{t}_i] = n\eta_i \tag{13}$$

Hence, the expected value of **t** is given as:

[4] Akoush et al. [16] and Strunk [17] confirm that **t** is linearly related with **s**. Our analysis is not limited to linear models, however.

[5] If we have two random variables **x** and **y** and the conditional density function $f(x|y)$, then the conditional expected value of **x** is expressed as $E[\mathbf{x}|y] = \int_{-\infty}^{\infty} xf(x|y)\,dx$. Then the expected value of **x** can be expressed as: $E[\mathbf{x}] = E[E[\mathbf{x}|\mathbf{y}]] = \int_{-\infty}^{\infty} E[\mathbf{x}|y]\, f(y)\, dy$. Note also that $f(x, y) = f(x|y)\, f(y) = f(y|x)\, f(x)$.

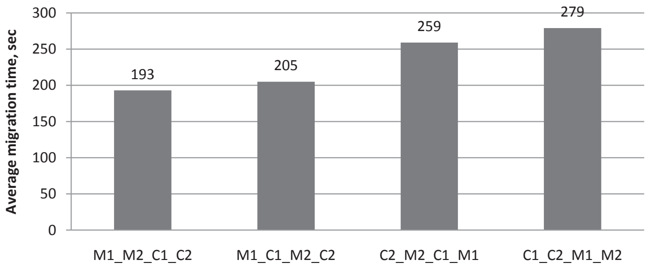

Fig. 5. Effect of the order of migration on migration time when resource-intensive virtual machines are migrated one after the other.

$$E\left[\mathbf{t}\right] = E\left[E\left[\mathbf{t}|\mathbf{n}\right]\right] = E\left[\mathbf{n}\right]\eta_i = \eta_n\eta_i \tag{14}$$

where η_n is the expected number of virtual machines to be migrated.

To compute the variance of $\mathbf{t}$, we shall make use of the relation: $\sigma_t^2 = E\left[t^2\right] - \left(E\left[\mathbf{t}\right]\right)^2$. Since we already know $\left(E\left[\mathbf{t}\right]\right)^2 = \left(\eta_n\eta_i\right)^2$, it suffices to compute $E\left[t^2\right]$. Following the same reasoning pattern in Equation 12:

$$E\left[\mathbf{t}^2|\mathbf{n}=n\right] = E\left[\left(\sum_{i=1}^{n}\mathbf{t}_i\right)^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\mathbf{t}_i\mathbf{t}_j\right] \tag{15}$$

It must be remembered that we have assumed that the migration times $\mathbf{t}_i$ amd $\mathbf{t}_j$ $\forall i \neq j$ are statistically independent. Hence,

$$E\left[\mathbf{t}_i\mathbf{t}_j\right] = \begin{cases}\sigma_i^2 + \eta_i & i = j\\ \sigma_i^2 & i \neq j\end{cases} \tag{16}$$

The double sum in Equation 15 contains n terms for $i = j$ and $n^2 - n$ terms for $i \neq j$; hence it equals $\left(\sigma_i^2 + \eta_i^2\right)n + \eta_i^2\left(n^2 - n\right) = \eta_i^2 n^2 + \sigma_i^2 n$.

Subsequently,

$$E\left[\mathbf{t}^2\right] = E\left[E\left[\mathbf{t}|\mathbf{n}\right]\right] = E\left[\eta_i^2\mathbf{n}^2 + \sigma_i^2\mathbf{n}\right] \tag{17}$$

D. *Effect of Workload Type on Migration Time*

The expressions we developed so far to quantify the cost of VM migration do not take into consideration the types of applications (workloads) the virtual machines host at the time of migration and the sequence of migration. This is because we assumed that the server is underutilized and the virtual machines do not consume resources appreciably. If, on the other hand, the virtual machines are actively utilizing resources (for example, CPU cycles, interconnects, and RAM bandwidth), then estimating the cost of migration becomes complex.

Interestingly, for an active server, the migration time is also influenced by the order of migration (i.e., which of the virtual machines are migrated first or last) [18]. Fig. 5 compares the average time required for migrating four virtual machines with different sequences. The virtual machines hosted four different SPEC CPU 2006 benchmarks: 464.h264ref (integer operation), 444.namd (floating point operation), 429.mcf (integer operation), and 401.bzip (integer operation). The first two benchmarks were predominantly CPU-intensive whereas the last two were both CPU- and memory-intensive benchmarks (with high memory read-write rate).

Fig. 5 displays the four possible migration orders. Since the source server requires sufficient resources to initiate and coordinate the migration process, removing the resource-intensive benchmarks first speeds up the migration.

V. Conclusion

One of the mechanisms to ensure the efficient utilization of computing resources in a cloud environment is workload consolidation. Typically, workload consolidation takes place to transfer the workload of underutilized servers and then to switch them off completely. During this process, virtual machines encapsulating active services and applications are migrated from the underutilized servers (source servers) to servers which can be loaded optimally (destination servers).

Live migration may cause performance degradation of many applications, both within and outside the virtual machine being migrated. This is because the virtual machine monitors in the source and the destination servers should allocate resources to orchestrate the migration.

In this paper we investigated the factors that affect the cost of live migration of virtual machines. Experiment results revealed that the power consumption of both the source and the destination servers proportionally increased as the available network bandwidth decreased, but the power consumption of both servers was not affected by the RAM size of the virtual machine being migrated. Likewise, the migration time exponentially decreases as the network bandwidth increased. The migration time increased linearly as the RAM size of the virtual machine increased.

If we consider the available network bandwidth during migration as a fixed quantity, the cost of migration, particularly, the total migration time ($\mathbf{t}$), depends on the number of virtual machines to be migrated ($\mathbf{n}$) and the size of each virtual machine ($\mathbf{s}$), both of which cannot be known in advance except in a probabilistic sense. Therefore, we modeled these quantities as random variables and derived expressions to estimate the statistics of $\mathbf{t}$ in terms of $\mathbf{n}$ and $\mathbf{s}$.

Our approach is suitable for underutilized servers in which hosted virtual machines do not utilize computing resources appreciably. If, however, the virtual machines consume resources significantly, accounting the cost of migrations becomes complex. In fact, even the order of migration itself can have a considerable effect. We have experimentally shown that first migrating virtual machines hosting resource-intensive applications has a shorter migration time than first migrating less resource-intensive virtual machines.

Acknowledgment

The experiments presented in this paper have been conducted by my colleague Dr. Anja Strunk during her tenure at the Chair of Computer Networks, TU Dresden.

This work has been partially funded by the German Research Foundation (DFG) under project agreement: SFB 912/1 2011.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Commun. ACM*, vol. 53, pp. 50–58, Apr. 2010.

[2] R. Buyya, "Market-oriented cloud computing: Vision, hype, and reality of delivering computing as the 5th utility," in *Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, CCGRID '09, (Washington, DC, USA), pp. 1–, IEEE Computer Society, 2009.

[3] M. Lin, A. Wierman, L. L. H. Andrew, and E. Thereska, "Dynamic right-sizing for power-proportional data centers," *2011 Proceedings IEEE INFOCOM*, pp. 1098–1106, Apr. 2011.

[4] D. Abts, M. R. Marty, P. M. Wells, P. Klausler, and H. Liu, "Energy proportional datacenter networks," *SIGARCH Comput. Archit. News*, vol. 38, no. 3, pp. 338—-347, 2010.

[5] Q. Zhu, J. Zhu, and G. Agrawal, "Power-aware consolidation of scientific workflows in virtualized environments," in *High Performance Computing, Networking, Storage and Analysis (SC), 2010 International Conference for*, pp. 1–12, Nov 2010.

[6] C. Curino, E. P. Jones, S. Madden, and H. Balakrishnan, "Workload-aware database monitoring and consolidation," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, (New York, NY, USA), pp. 313–324, ACM, 2011.

[7] W. Dargie, A. Strunk, and A. Schill, "Energy-aware service execution," *2011 IEEE 36th Conference on Local Computer Networks*, pp. 1064–1071, Oct. 2011.

[8] I. Rodero, J. Jaramillo, A. Quiroz, M. Parashar, and F. Guim, "Towards energy-aware autonomic provisioning for virtualized environments," in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing - HPDC '10*, (New York, New York, USA), p. 320, ACM Press, 2010.

[9] B. Sotomayor, R. S. Montero, I. Llorente, and I. Foster, "Virtual infrastructure management in private and hybrid clouds," *Internet Computing, IEEE*, vol. 13, no. 5, pp. 14–22, 2009.

[10] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "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.

[11] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Black-box and gray-box strategies for virtual machine migration," in *Proceedings of the 4th USENIX Conference on Networked Systems Design &#38; Implementation*, NSDI'07, (Berkeley, CA, USA), pp. 17–17, USENIX Association, 2007.

[12] A. Strunk and W. Dargie, "Does Live Migration of Virtual Machines cost Energy?," in *2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, (Barcelona, Spain), pp. 514–521, 2013.

[13] K. Rybina, W. Dargie, A. Strunk, and A. Schill, "Investigation into the energy cost of live migration of virtual machines," in *SustainIT*, pp. 1–8, 2013.

[14] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast transparent migration for virtual machines," in *Proceedings of the Annual Conference on USENIX Annual Technical Conference*, ATEC '05, (Berkeley, CA, USA), pp. 25–25, USENIX Association, 2005.

[15] H. Jin, L. Deng, S. Wu, X. Shi, and X. Pan, "Live virtual machine migration with adaptive, memory compression," in *Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on*, pp. 1–10, Aug 2009.

[16] S. Akoush, R. Sohan, A. Rice, A. Moore, and A. Hopper, "Predicting the performance of virtual machine migration," in *Modeling, Analysis Simulation of Computer and Telecommunication Systems (MASCOTS), 2010 IEEE International Symposium on*, pp. 37–46, Aug 2010.

[17] A. Strunk, "A lightweight model for estimating energy cost of live migration of virtual machines," in *IEEE CLOUD*, pp. 510–517, 2013.

[18] K. Rybina, A. Patni, and A. Schill, "Analysing the Migration Time of Live Migration of Multiple Virtual Machines," in *4th International Conference on Cloud Computing and Services Science (CLOSER 2014)*, 2014.