Ulrich Lampe, Markus Kieselmann, André Miede, Sebastian Zöller, Ralf Steinmetz: **On the Accuracy of Time Measurements in Virtual Machines.** In: Rong Chang, Ephraim Feig, Alan Sussman, Liana L. Fong: Proceedings of the 6th International Conference on Cloud Computing (CLOUD 2013), p. 103-104, Institute of Electrical and Electronics Engineers (IEEE), June 2013. ISBN 978-0-7685-5028-2.

On the Accuracy of Time Measurements in Virtual Machines

Ulrich Lampe*, Markus Kieselmann*, André Miede†, Sebastian Zöller* and Ralf Steinmetz*
**Multimedia Communications Lab (KOM), Technische Universität Darmstadt, Darmstadt, Germany*
Contact Author: ulrich.lampe@KOM.tu-darmstadt.de
†*Fakultät für Ingenieurwissenschaften, HTW des Saarlandes, Saarbrücken, Germany*

***Abstract*—While cloud computing permits access to a large pool of experimental infrastructure, the most common form – virtual machines – has been shown to exhibit substantial deficits with respect to the accuracy of time measurements. In our ongoing work, we provide a detailed analysis of these deficits based on various machine configurations. Preliminary results indicate that not the use of virtualization as such, but the potentially uncontrollable utilization of the physical host is a decisive factor for the accuracy of time measurements.**

***Keywords*-virtual machine, time measurement, accuracy**

I. Introduction

A key element of cloud computing is elasticity, i. e., the ability to access Information Technology (IT) resources in a flexible and affordable fashion [1]. This characteristic is of great benefit for researchers, who frequently require large capacities on short term in order to conduct scientific experiments. Due to their functional flexibility, Virtual Machine (VM) instances are preferably used in this context.

Unfortunately, our past work has provided indications that VMs suffer from deficits with respect to the accuracy of time measurements [2]. In the paper at hand, we report initial findings from our ongoing research work, which aims to more thoroughly examine the time measurement accuracy in physical and cloud-based virtual environments under consideration of various influence factors, including virtualization as such and different virtualization solutions.

II. Experimental Setup

In this work, we pursue the same approach as in our past research [2]: We repeatedly measure the computation time of a deterministic function in order to quantify potential inaccuracies in time measurement. Specifically, our Java-based measurement tool features a simple counter function. It automatically adapts to a given integer argument a, such that a computation time of approximately $a \times 10$ ms can be expected, and returns the practically observed computation time as result. The tool can be configured to conduct a series of measurements, referred to as batches. Each batch comprises a number of calls of the aforementioned counter function based on a given set of arguments. More formally, b batches are subsequently executed, with each batch comprising c method calls using the individual arguments $a \in A = \{2^0, 2^1, \ldots, 2^m\}$, respectively. The parameters b, c, and m can be freely chosen by the user.

The aim of our work is to quantify the impact of different potential influence factors on the accuracy of time measurements, which constitutes the *dependent variable*. Thus, we employ a multitude of different *machine configurations* in our experiments, where the influence factors are modeled through four *independent variables*.

The first variable, *machine type*, may correspond to *VM*, as common form of cloud infrastructure, or Physical Machine (*PM*), as traditional experimental infrastructure. The second variable, *deployment model*, comprises VMs from a *public cloud* (Amazon EC2), a *private cloud* that is operated at our institute (KOM), and VMs on a *local host* computer. As third independent variable, we regard the *virtualization software*, considering the commercial VMWare *ESXi* and the open-source software *Xen*. As fourth independent variable, we take into account *host utilization*, i. e., computational load that the PM or host system for the VMs is subjected to. We consider the cases of *low load*, i. e., the PM exclusively hosts one instance of the measurement tool or VM, *high load*, where the system runs multiple measurement tool or VM instances in parallel, and *random load*, i. e., the host utilization is out of our control sphere and potentially fluctuates.

In principal, we follow a full-factorial approach in our experiments. That is, we examine each possible combination of values for the four independent variables that were introduced in the previous section. Because some combinations are mutually exclusive (e. g., Amazon does not provide a choice between different virtualization systems), our experiment comprises a total of 8 different machine configurations. For every configuration, we conducted $b = 20$ experimental batches with $c = 100$ method calls each. The set of applicable arguments was specified as $A = \{2^0 = 1, ..., 2^9 = 512\}$, i. e., $m = 9$. Thus, we obtained a total *sample* of $20,000$ runtime observations per configuration. Given that Windows and Linux exhibited similar time measurement accuracy in our past work [2], we chose Ubuntu Server 12.04.1 LTS as default operating system for the experiments.

III. Preliminary Results

In accordance with our previous work [2], we use the normalized standard deviation, i. e., the *Coefficient of Variation* (CV), as measure of accuracy. It is given by the ratio between the standard deviation (commonly denoted as σ) and the mean value of the observations (μ) in a sample. The

Table I: Observed time measurement accuracies, i. e., *coefficients of variation*, by machine configuration. Values in parentheses denote the rank among all configurations for the given argument, ordered from most accurate (1) to least accurate (8). Abbreviations: M/T (Machine Type), D/M (Deployment Model), V/S (Virtualization Software), H/U (Host Utilization).

#	Machine Configuration				Function Argument ($\approx a \times 10$ ms Computation Time)															
	M/T	D/M	V/S	H/U	$a = 1$		$a = 2$		$a = 4$		$a = 8$		$a = 16$		$a = 32$		$a = 64$		$a = 128$	
1	PM	n/a	n/a	Low	0.0328	(3)	0.0403	(3)	0.0005	(1)	0.0003	(1)	0.0002	(1)	0.0003	(2)	0.0001	(1)	0.0001	(2)
2	PM	n/a	n/a	High	0.5527	(5)	0.3876	(5)	0.3152	(5)	0.2814	(7)	0.2501	(7)	0.2290	(7)	0.2078	(8)	0.1795	(8)
3	VM	Local	ESXi	Low	0.0290	(2)	0.0203	(2)	0.0226	(4)	0.0085	(4)	0.0057	(4)	0.0045	(4)	0.0195	(4)	0.0086	(4)
4	VM	Local	ESXi	High	1.6590	(8)	1.2490	(8)	0.8356	(8)	0.5877	(8)	0.4783	(8)	0.3420	(8)	0.1934	(7)	0.0822	(7)
5	VM	Local	Xen	Low	0.2463	(4)	0.0733	(4)	0.0029	(2)	0.0005	(2)	0.0002	(2)	0.0001	(1)	0.0001	(2)	0.0001	(1)
6	VM	Local	Xen	High	1.0830	(6)	0.7112	(7)	0.4449	(7)	0.2489	(6)	0.1135	(6)	0.0597	(6)	0.0296	(6)	0.0157	(5)
7	VM	Private	ESXi	Low	0.0118	(1)	0.0074	(1)	0.0029	(3)	0.0022	(3)	0.0022	(3)	0.0017	(3)	0.0017	(3)	0.0010	(3)
8	VM	Public	Xen	Rnd.	1.1398	(7)	0.6293	(6)	0.3576	(6)	0.1783	(5)	0.0974	(5)	0.0499	(5)	0.0271	(5)	0.0159	(6)

CV numerically represents the dependent variable in our experiments. *Higher* values indicate *lower* accuracy and vice versa. Hence, in the case of ideal accuracy, the observed CV would correspond to zero. An overview of all considered machine configurations and corresponding results is provided in Table I. Results for the arguments $a \geq 256$ have been omitted due to space limitations.

Given the findings of our previous work [2], which indicated general deficits of VMs with respect to time measurements, our new experiments provide some surprises. Specifically, the VM from the *private* cloud at our institute provide the best accuracy for small arguments, i. e., $a \leq 2$, among all tested configurations (cf. #7 in Table I). For increasing arguments, the VM loses some ground to the PM (cf. #1 in Table I). Yet, the results confirm the deficits of *public* clouds with respect to time measurements. Notably, the VM instance from Amazon EC2 exhibit very high CVs, i. e., low accuracy, for most arguments, specifically for $a \leq 4$ (cf. #8 in Table I).

Concerning the two virtualization solutions, ESXi and Xen, we obtained mixed results. On the basis of the locally hosted VMs and low utilization, the observed CVs indicate some advantages for ESXi with respect to small arguments (i. e., $a \leq 2$), while the relative performance of Xen improves with growing arguments (cf. #3 and #5 in Table I). In addition, Xen achieves more favorable accuracy once high host utilization comes into play; in this case, ESXi generally appears to perform very poorly (cf. #4 in Table I).

From the above discussion, one may conclude that the host utilization plays a key role in the accuracy of measurements, and this is strikingly confirmed in our experiments. Regardless of the machine type and virtualization software, imposing additional load on the physical host results in sharp increases in the observed CVs (cf. #4 and #6 in Table I). The same applies for the physical machine (cf. #2 in Table I).

IV. Conclusions and Outlook

The initial experimental results in this paper to some extent relativize the findings of our previous work: Most notably, we have found that contemporary virtualization technology as such does *not* necessarily imply deficits with respect to the accuracy of time measurements. In fact, some of the lowest CVs, i. e., best accuracies, among all machine configurations in our experiments were observed on VM instance from a private cloud. Likewise, those VMs that were hosted on a single physical host performed very similarly to a "raw" PM.

Our experiments have shown that a different influence factor, namely host utilization, is the key determinant for time measurement inaccuracies. Unfortunately, this is the very factor that commonly lies out of the control sphere of the end user when leasing resources from a public cloud. In fact, from the viewpoint of a cloud provider, the consolidation of multiple VMs on a single physical host is highly desirable in order to reduce operational cost. The same also applies to a private cloud in principal, even though the level of control may be higher for the end user in such deployment model. To state it more explicitly, virtualization does not appear to hurt the accuracy time measurement, but high host utilization – which is a key benefit of virtualization – does.

Hence, our results confirm the most important recommendation of our previous work: If accurate time measurements, specifically in the sub-second range, are required in scientific experiments, dedicated PMs should be preferred over VMs. Yet, if host utilization as the key influence factor can be effectively controlled by the end user, VMs may also provide acceptable accuracy.

For the future, we plan to extend our work through a longitudinal (long-term) study design, as well as the consideration of additional commercial cloud providers, programming languages, and real-time operating systems.

Acknowledgments

This work has been sponsored in part by the E-Finance Lab e. V., Frankfurt a. M., Germany (www.efinancelab.de).

References

[1] D. Owens, "Securing Elasticity in the Cloud," *Communications of the ACM*, vol. 53, no. 6, pp. 46–51, 2010.

[2] U. Lampe, A. Miede, N. Richerzhagen, D. Schuller, and R. Steinmetz, "The Virtual Margin of Error," in *Proc. of CLOSER 2012*, 2012.