Energy Efficiency Comparison of Hypervisors

Congfeng Jiang[1,2,*], Yumei Wang[1], Dongyang Ou[1],
Youhuizi Li[1], Jilin Zhang[1], Jian Wan[1]

[1]Key Laboratory of Complex Systems Modeling and Simulation

Hangzhou Dianzi University

Hangzhou, Zhejiang, China

{jiangcongfeng, wyumei1, oudongyang1}gmail.com,

{huizi, jilin.zhang,wanjian}@hdu.edu.cn

Bing Luo[2], Weisong Shi[2]

[2]Mobile and Internet Systems Laboratory

Department of Computer Science

Wayne State University

Detroit, MI, USA

{bing,weisong}wayne.edu

Abstract—**Current cloud data centers are fully virtualized for service consolidation and power/energy reduction. Although virtualization could reduce the real-time power consumption and overall energy consumption, the energy characteristics of hypervisors hosting different workloads have not been well profiled or understood thus far. In this study, we investigate the power and energy characteristics of four mainstream hypervisors and a container engine, namely VMware ESXi, Microsoft Hyper-V, KVM, XenServer, and Docker, on six different platforms (three mainstream 2U rack servers, one emerging ARM64 server, one desktop server, and one laptop) with power measurements made over prolonged periods. We use computation-intensive, memory-intensive, and mixed Web server-database workloads to explore the power and energy characteristics of different hypervisors in order to emulate realistic multi-tenant cloud environments. The results of extensive experiments conducted with four workload levels (very light, light, fair, and very heavy) indicate that the hypervisors exhibit different power and energy characteristics. Our findings are as follows. (1) Hypervisors exhibit different power and energy consumptions on the same hardware running the same workload. (2) Although mainstream hypervisors have different energy efficiencies aligned with different workload types and workload levels, no single hypervisor outperforms the other hypervisors on all platforms in terms of power or energy consumption. (3) Although container virtualization is considered as lightweight virtualization in terms of implementation and maintenance, it is essentially not more power-efficient than conventional virtualization technology. (4) Although the ARM64 server has low power consumption, it completes computation tasks with a long execution time and, sometimes, high energy consumption. Further, ARM64 servers have medium energy consumption per database operation for mixed workloads. The results presented in this paper can provide system designers and data center operators with useful insights for power-aware workload placement and virtual machine scheduling.**

Keywords—energy efficiency; power; hypervisor; virtual machine; container virtualization

I. MOTIVATION

The proliferation of cloud computing, big data analytics, e-commerce, and Internet traffic has led to a rapidly growing demand for power by data centers [1, 2]. Various hardware-related approaches have been proposed to increase data center energy efficiency at different levels [3, 4], including circuits and chips [5, 6, 7], memory [8, 9], disk [10, 11], and network traffic routing [12, 13]. Furthermore, the energy efficiency potential of various power- and energy-aware approaches has been explored intensively, e.g., performance tuning [14, 15], application-centric power optimization [16, 17], resource scheduling and allocation [18, 19], and thermal-aware power capping [20, 21]. Resource multiplexing in data centers provides data-center-wide power management opportunities [22, 23]. In recent years, renewable energy and liquid cooled systems have been introduced into modern data centers to further reduce their carbon footprint [24, 25]. In addition, ARM64 servers are regarded as competitive candidates in the server market owing to their lower power consumption compared with servers based on conventional x86 processors.

In cloud data centers, server virtualization and consolidation are widely deployed to reduce power and energy consumption [26]. In general, over-commitment or over-subscription is adopted to further reduce energy costs more actively. For a virtualized platform, the hypervisor or virtual machine monitor (VMM) acts as the equivalent operating system and is responsible for scheduling resources and hosting the guest operating system. In contrast to classical heavyweight virtualization, container virtualization (or containerization) is a lightweight and simple technique that can be scaled up to host a larger number of applications. Thus, it has emerged as a new paradigm for cloud data centers. It is easier to deploy multiple copies of the same applications/services by using container technology rather than virtual machines because containers only require an operating system, supporting programs and libraries, and system resources to run a specific program without any hardware abstraction. In the containerization environment, the container engine acts as the hypervisor in the conventional virtualized environment, while it leverages the underlying operating system kernel for core resource management and allocation. In this paper, we consider the container engine as the hypervisor for simplification, and we use the two terms interchangeably. In addition, we use the terms virtual machine and container interchangeably unless specified otherwise in a particular context.

Owing to hardware abstraction and the semantic gap between the virtual machine (or container) and the underlying hardware, the virtual machine operating system cannot invoke actual power-aware management as in a non-virtualized environment. Moreover, the use of hypervisors may affect different hardware platforms and guest virtual machines at different levels in terms of energy efficiency. Therefore, in such virtualized environments, power management and energy accounting require accurate knowledge of the energy efficiency characteristics of hypervisors [27]. From the data-center-wide point of view, differentiating the energy efficiencies of various hypervisors can facilitate hypervisor selection, system design, and system operation. This motivates us to investigate and compare the energy efficiencies of different hypervisors.

In this paper, we do not intend to compare the power and performance of different servers as in previous studies [28, 29, 30] (see related work in Section IV). Instead, we compare the energy efficiencies of hypervisors on the same server. In particular, we attempt to answer the following questions:

(Q1) Do different hypervisors have approximately the same energy efficiency for a virtual machine running the same application on the same hardware platform? If not, how much is the difference? Answering this question is essentially equivalent to investigating how we can choose hypervisors to host virtual machines.

(**Q2**) Does one hypervisor have different energy efficiencies on different hardware platforms running the same virtual machine and same application? Answering this question is equivalent to investigating whether more than one type of hypervisor needs to be deployed to harvest the energy efficiency variability of existing servers or whether a single hypervisor can be deployed on all the servers to simplify the management cost for a given workload.

(Q3) Does one hypervisor have different energy efficiencies on the same hardware platform running different applications within its virtual machines? Answering this question is equivalent to investigating whether hypervisor affinity exists across different applications or workloads.

(**Q4**) Should we always use newer hardware to achieve better energy efficiency in a virtualized environment? Do these platforms significantly differ in terms of energy efficiency in a virtualized environment?

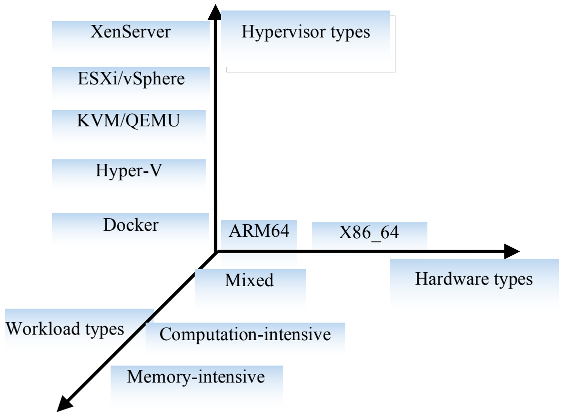


Fig. 1. Comparison of energy efficiency of hypervisors in three orthogonal dimensions

It is worth noting that although most research attention is currently focused on publicly known large-scale cloud data centers that consume less than 5% of the data center electricity in the U.S., many small, medium, corporate, and multi-tenant data centers continue to operate with low energy efficiency [31]. Therefore, in this study, we investigate the energy efficiency of not only rack servers used in typical large-scale cloud data centers but also desktop servers, ARM64 servers, and laptops used in other data centers. We selected four mainstream hypervisors, namely Microsoft Hyper-V, VMware ESXi, KVM/QEMU, and XenServer, as well as the Docker container engine, because these hypervisors are widely deployed in existing virtualized data centers. Further, we conducted extensive experiments in three orthogonal dimensions: hardware, hypervisor, and workload (see Fig. 1). Specifically, we compared the energy efficiency of five hypervisors (including Docker) on six hardware platforms running three types of workload at four workload levels. We measured the real-time power and calculated the energy consumption of each experiment and micro-operation.

The remainder of this paper is organized as follows. Section II describes the details of the experiment methodology and system setup. Section III presents the experiment results and discusses the observations made and insights gained with regard to the energy efficiency of the hypervisors. Section IV summarizes the related work. Finally, Section V concludes the paper and briefly explores directions for future work.

II. EXPERIMENT METHODOLOGY AND SETUP

A. Experiment Methodology

We conducted experiments in three orthogonal dimensions: hardware, hypervisor, and workload. For example, for a specific hardware, we ran different hypervisors on it, where each hypervisor hosted a number of virtual machines running computation-intensive, memory intensive and mixed workloads. In contrast to previous studies, we conducted a fine-grained energy efficiency comparison of hypervisors under a series of workload intensities to obtain the energy proportionality profile of the hardware and hypervisors. Previous studies on energy efficiency comparison have focused on a fixed workload level for each workload type; such an approach is not comprehensive because different hardware platforms have different energy proportionalities. Instead, we classified the workload intensity into four workload levels, namely very light, light, fair, and very heavy, in order to emulate a realistic multi-tenant virtualized cloud environment. For computation-intensive workloads, the fair workload level can stress the system at more than 95% utilization steadily, while the very heavy workload level saturates the system. Our methodology provides an opportunity to investigate the energy proportionality of not only the hardware itself but also the hypervisor, and our experiments show that the power and energy consumption change significantly on the same hardware with varying workload levels when running different hypervisors.

B. Experiment Setup

To compare the energy efficiencies of the hypervisors comprehensively, we selected six typical hardware platforms as our testbed. The testbed configuration is summarized in Table 1, and the platforms are sorted by CPU release date (Intel Q8300 is the oldest). At present, some processors and motherboards developed for laptops are used in data centers owing to their good power performance. We selected the Lenovo W541 laptop as a representative of such low-power platforms. We ran the following hypervisors on the different platforms: Windows Server 2012 R2 for Hyper-V (data center version with GUI), CentOS Linux 7 (v1503 server with GUI) for KVM/QEMU 1.5.3, ESXi 6.0/vSphere 6.0, and XenServer 6.5.0. For a reasonable comparison, the virtual machines ran the same operating system with the same software configuration, including the kernel version and software stack. For example, all the virtual machines ran the same x64 version of CentOS 7 with Linux kernel 3.10. All the power data were measured using a Watts up? .NET power meter.

Table 1. Platform configurations

Platform	Type	CPU	Cores/threads	Memory	Storage
HP s5280t	Desktop server	Intel Q8300	4/4	8 GB	1 TB HDD
HP DL380 G6	2U server	Intel Xeon E5530 x2/wVT-x &VT-d	8/16	80GB	2 TB SSD RAID0
Intel S2600GZ	2U server	Intel Xeon E5-2680x2/w VT-x &VT-d	16/32	64GB	2 TB HDD RAID1
Lenovo W541	Laptop	Intel i7- 4710MQ with VT-x	4/8	32 GB	2 TB SSD
APM X-C1	ARM64 server	APM X-Gene 2	8/8	16 GB	2 TB SSD
Lenovo RD450	2U server	Intel Xeon E5-2620v3/w VT-x &VT-d	12/24	192GB	3.3 TB HDD RAID 0

C. *Experiment Workload and Workload Level Classification*

On top of hypervisors, we ran a number of virtual machines with three workloads: one computation-intensive workload, one memory-intensive workload and one mixed Web server and database workload.

Computation-intensive workload: We used a prime number computation program written in C, namely PrimeSearch, as the computation-intensive workload. In one execution of PrimeSearch, it calculates and searches for prime numbers in 10 intervals: (1,1000000), (1,2000000), (1,3000000), (1,4000000), (1,5000000), (1,6000000), (1,7000000), (1,8000000), (1,9000000), and (1,10000000). These 10 intervals are 10 sub-searching tasks. The completion time of searching in each interval is calculated and the sum of the completion times of the 10 sub-tasks is considered as the task completion time of one PrimeSearch execution.

Memory-intensive workload: We used STREAM [44] as a synthetic benchmark to stress the memory system of the tested servers, and we measured the memory bandwidth (in MB/s) and the corresponding computation rate for simple vector kernels.

Mixed workload: We used a two-tier Web server and database synthetic environment as the mixed workload, namely LAMP (Linux, Apache, MySQL, and PHP). In the mixed LAMP workload, we used PHP scripts in Web pages to insert data records into MySQL database via Apache Web server. Each execution tries to insert 1000000 rows of data. The PrimeSearch and LAMP workloads are static workloads. During each experiment, all the virtual machines ran the same code; thus, each machine contributed equally to the workload on the system being tested.

For convenience, we refer to the above-mentioned workloads as computation-intensive, memory-intensive, and mixed workloads, respectively, in the remainder of the paper.

Workload level: In our experiments, we considered the number of virtual machines running concurrently within the same physical server as the workload indicator. However, because different server platforms are configured with different processor sockets, processor cores, and execution threads, we ran the virtual machines in proportion to the physical processor cores of the tested server. Although it is possible to compare the energy efficiencies of different servers with the same number of virtual machines of different hypervisors, it is reasonable to compare them with the number of virtual machines proportional to the hardware configuration (especially the number of processor cores) in order to investigate the energy efficiency scalability on different platforms. Here, we define four workload levels, namely *1/4*, *1/2*, *1/1*, and *2/1,* which correspond to very light, light, fair, and very heavy workloads, respectively. We considered the workload level as the workload intensity indicator. These workload levels imply that the number of virtual machines running concurrently within a physical server is one-fourth of (1/4), one-half of (1/2), equal to (1/1), and two times (2/1) the number of physical processor cores, respectively. For example, when we ran very heavy computation-intensive workload on the Lenovo RD450 server, we ran ***24*** virtual machines. By contrast, we ran only ***8*** virtual machines on the HP s5280t server for very heavy workload because the HP s5280t server has only 4 physical processor cores (see Table 1 for processor configurations).

III. EXPERIMENTAL OBSERVATIONS AND INSIGHTS

A. *Computation-intensive Workloads*

First, we run the PrimeSearch workload on each platform. The power and energy results are shown in Fig. 2, while the completion times are shown in Fig. 3. Further, the power and energy variations (ratio of the highest to the lowest) of different hypervisors for different workload running on different platforms are summarized in Tables 2 and 3, respectively. Our observations are stated below.

Observation #1: The hypervisors exhibit different power consumptions, completion times, and energy consumptions on the same hardware running the same workload.

As can be seen in Figs. 2 and 3, different hypervisors have different powers, energy consumptions, and completion times on the same hardware. Here, the implications of the differences are two-fold. First, one hypervisor has different power consumptions and energy consumptions at different workload levels on the same hardware. Second, different hypervisors have different power consumptions and energy consumptions on the same hardware at the same workload level.

From Table 2, we observe that the power variation on HP DL380 G6 is 19.79% for very heavy computation-intensive workload by comparing the highest power (from Hyper-V) with the lowest power (from KVM). However, on another typical rack server, i.e., Intel S2600GZ, the power variation of the highest (from Docker) to the lowest (from Hyper-V) is 8.40% for very heavy computation-intensive workload. Similarly, on the Lenovo RD450 2U server, the power variation of the highest (from Hyper-V) to the lowest (from KVM) is 120.53% for very light computation-intensive workload. We also observe that, in general, the heavier the workload, the higher is the power variation for the three 2U rack servers, except for the 1/4 and 1/2 workloads, on Lenovo RD450 when running different hypervisors. However, the opposite results are obtained for the desktop server and laptop. Because we only ran KVM and Docker on APM X-C1, no significant trend in power variation is observed with respect to the workload level. It is worth noting that on the rack servers, the maximum power variation occurs when the workload is the heaviest, whereas on the desktop server, laptop, and ARM64 server, the maximum power variation occurs when the workload is the lightest. As can be seen in Table 3, the energy variations are larger than the power variations in Table 2. This is because the energy is the product of the power and the task completion time. Owing to space constraints, we only present the variations in completion time (ratio of the highest to the lowest) in Table 4. Moreover, compared to the power variations, the energy variations of the hypervisors on different platforms are more diversified and scattered across all the workload levels, except the 1/2 workload level.

Observation #2: Although the hypervisors have different energy efficiencies aligned with different workload types and workload levels, no single hypervisor outperforms the other hypervisors in terms of power, energy consumption, or completion time for all workload levels on all platforms.

Table 2. Power variation of computation-intensive workloads

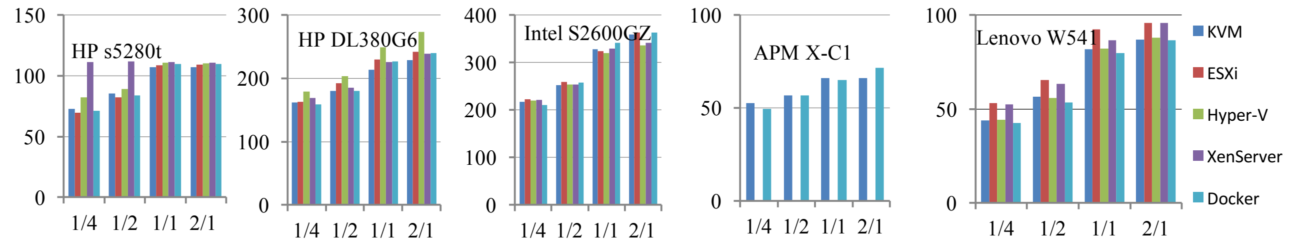
Platform	1/4 workload	1/2 workload	1/1 workload	2/1 workload
HP s5280t	**59.87**%	35.78%	3.52%	3.53%
HP DL380 G6	12.38%	12.98%	16.67%	**19.79**%
Intel S2600GZ	5.44%	3.03%	6.71%	**8.40**%
Lenovo W541	**25.03**%	22.71%	15.87%	10.81%
APM X-C1	**10.11**%	0.05%	1.18%	8.18%
Lenovo RD450	**120.53**%	102.74%	15.18%	29.03%

Table 3. Energy variation of computation-intensive workloads

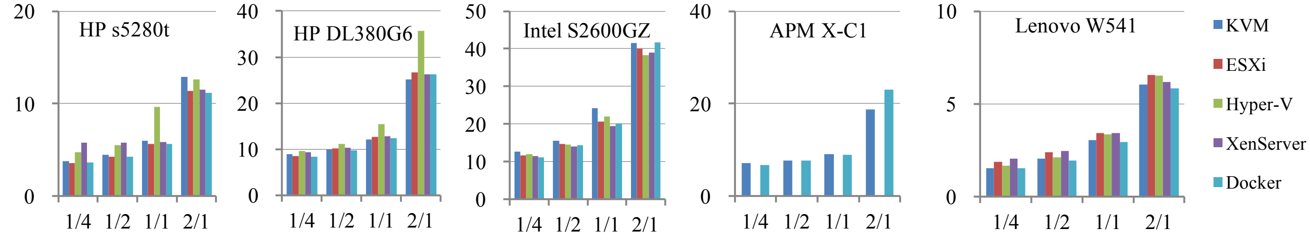
Platform	1/4 workload	1/2 workload	1/1 workload	2/1 workload
HP s5280t	61.86%	36.74%	**72.46**%	14.88%
HP DL380 G6	15.70%	14.05%	28.05%	**41.90**%
Intel S2600GZ	13.56%	11.06%	**25.06**%	9.26%
Lenovo W541	**35.34**%	25.90%	16.62%	12.77%
APM X-C1	10.04%	0.05%	2.48%	**22.60**%
Lenovo RD450	**833.06**%	670.40%	248.84%	13.25%

Table 4. Completion time variation of computation-intensive workloads

Platform	1/4 workload	1/2 workload	1/1 workload	2/1 workload
HP s5280t	14.06%	20.41%	**70.48**%	17.32%
HP DL380 G6	6.23%	4.71%	13.03%	**18.95**%
Intel S2600GZ	12.13%	11.85%	**24.69**%	5.01%
Lenovo W541	11.18%	6.69%	11.08%	**14.76**%
APM X-C1	0.06%	0.00%	1.28%	**13.33**%
Lenovo RD450	**323.10**%	282.93%	252.23%	13.94%



(a) Power consumption (x axis: workload level, y axis: power (watts))



(b) Energy consumption (x axis: workload level, y axis: energy (10^6 joules))

Fig. 2 Power and energy consumption of varying computation-intensive workloads

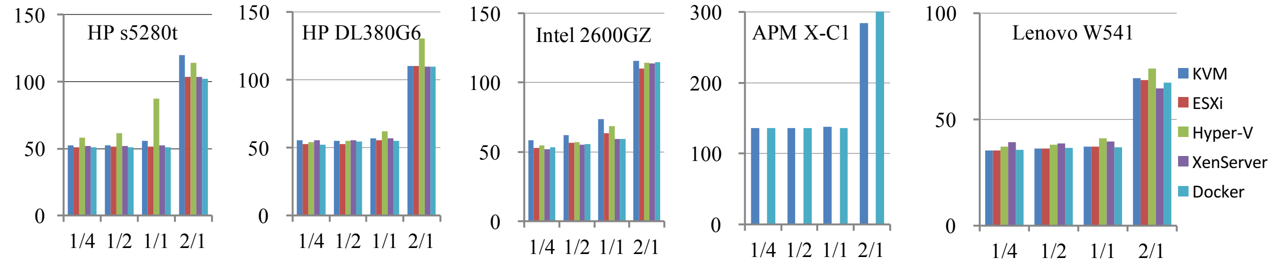


Fig. 3 Completion time of varying computation-intensive workloads (x axis: workload level, y axis: completion time (10^3 seconds))

The highest and lowest power consumptions at all workload levels on all platforms are summarized in Table 5.

Table 5 Hypervisors with the highest and lowest power consumptions for computation-intensive workload

Platform	Power	Workload level			
		1/4	1/2	1/1	2/1
HP s5280t	highest	XenServer	XenServer	XenServer	XenServer
	lowest	ESXi	ESXi	KVM	KVM
HP DL380 G6	highest	Hyper-V	Hyper-V	Hyper-V	Hyper-V
	lowest	Docker	KVM	KVM	KVM
Intel S2600GZ	highest	ESXi	ESXi	Docker	Docker
	lowest	Docker	KVM	Hyper-V	Hyper-V
Lenovo W541	highest	ESXi	ESXi	ESXi	ESXi
	lowest	Docker	Docker	Docker	Docker
APM X-C1	highest	KVM	Docker	KVM	Docker
	lowest	Docker	KVM	Docker	KVM
Lenovo RD450	highest	Hyper-V	Hyper-V	ESXi	Hyper-V
	lowest	KVM	KVM	Docker	Docker

From Table 5, we can see that no single hypervisor always has the highest or lowest power consumption for all workload levels on all platforms. Instead, the distribution of the hypervisors with the highest or lowest power consumption is neither platform-dependent nor workload-level-dependent. One hypervisor may have the highest power consumption for some or all workload levels on one platform, whereas it may have the lowest power consumption for some or all workload levels on another platform. In other words, there is no hypervisor affinity across different hardware and workload levels. For example, Hyper-V has the highest power and energy consumption on the HP DL380 G6 server running a computation-intensive application at all workload levels, whereas it has the lowest power consumption on the Intel S2600GZ server running fair and very heavy computation-intensive workloads.

More importantly, although the ARM64 server APM X-C1 has the lowest power consumption, it also has the highest energy consumption because it has the longest completion time. In addition, XenServer has the highest power consumption only on HP s5280t for all workload levels and Hyper-V has the highest power consumption only on HP DL380 G6 for all workload levels. Here, we consider ESXi as another example. ESXi has the highest power consumption on Lenovo W541 for all workload levels and on Intel S2600GZ for the 1/4 and 1/2 workload levels. However, ESXi has the lowest power consumption on HP s5280t for the 1/4 and 1/2 workload levels. It is worth noting that XenServer exhibits only the highest power consumption while KVM exhibits only the lowest power consumption if we exclude the APM X-C1 platform, because we did not run as many hypervisors on it as we did on the other platforms (i.e., we ran only KVM and Docker on it). In fact, the APM X-C1 ARM64 server has the lowest power variation for all workloads regardless of the hypervisor that is running on it.

Similarly, as shown in Table 6, the distribution of hypervisors with the highest or lowest energy consumption is more diversified than the power consumption distribution in Table 5 because the energy consumption is jointly affected by the power and task completion time.

Table 6 Highest and lowest energy consumption of computation-intensive benchmarks on all platforms

Platform	Energy	Workload level			
		1/4	1/2	1/1	2/1
HP s5280t	Highest	XenServer	XenServer	Hyper-V	KVM
	Lowest	ESXi	ESXi	Docker	Docker
HP DL380 G6	Highest	Hyper-V	Hyper-V	Hyper-V	Hyper-V
	Lowest	Docker	Docker	KVM	KVM
Intel S2600GZ	Highest	KVM	KVM	KVM	Docker
	Lowest	Docker	XenServer	XenServer	Hyper-V
Lenovo W541	Highest	XenServer	XenServer	XenServer	ESXi
	Lowest	Docker	Docker	Docker	Docker
APM X-C1	Highest	KVM	Docker	KVM	Docker
	Lowest	Docker	KVM	Docker	KVM
Lenovo RD450	Highest	Hyper-V	Hyper-V	Hyper-V	Hyper-V
	Lowest	KVM	KVM	Docker	Docker

Observation #3: Although container virtualization is considered as lightweight virtualization in terms of implementation simplicity and ease of maintenance, it is essentially not more power-efficient than conventional virtualization technology for computation-intensive workloads.

The power consumption and energy consumption of Docker are compared with the highest, lowest, and average power consumption and energy consumption of all the hypervisors (including Docker itself) in Tables 7 and 8, respectively.

Table 7 Comparison between Docker and the hypervisor with the highest, average, and lowest power

Platform	Power	Workload level			
		1/4	1/2	1/1	2/1
HP s5280t	Highest	**-36.17%**	-24.82%	-1.39%	-1.35%
	Average	**-12.59%**	-7.20%	0.03%	0.03%
	Lowest	2.04%	2.08%	2.08%	**2.13%**
HP DL380 G6	Highest	-11.01%	-11.41%	-9.10%	**-12.39%**
	Average	-4.31%	**-4.37%**	-1.03%	-2.02%
	Lowest	0.00%	0.09%	**6.05%**	4.95%
Intel S2600GZ	Highest	**-5.16%**	-0.86%	0.00%	0.00%
	Average	-3.44%	0.78%	**3.96%**	3.29%
	Lowest	0.00%	2.14%	6.71%	**8.40%**
Lenovo W541	Highest	**-20.02%**	-18.51%	-13.69%	-9.75%
	Average	**-9.76%**	-9.44%	-5.63%	-4.54%
	Lowest	0.00%	0.00%	0.00%	0.00%
APM X-C1	Highest	**-9.18%**	0.00%	-1.17%	0.00%
	Average	**-4.81%**	0.03%	-0.59%	3.93%
	Lowest	0.00%	0.05%	0.00%	**8.18%**
Lenovo RD450	Highest	-22.02%	-15.68%	-13.18%	**-22.50**%
	Average	-4.94%	-1.70%	-8.93%	**-14.18**%
	Lowest	**71.97**%	70.94%	0.00%	0.00%

Note: Positive values imply that Docker consumes more power.

Table 8 Comparison between Docker and the hypervisor with the highest, average, and lowest energy consumption

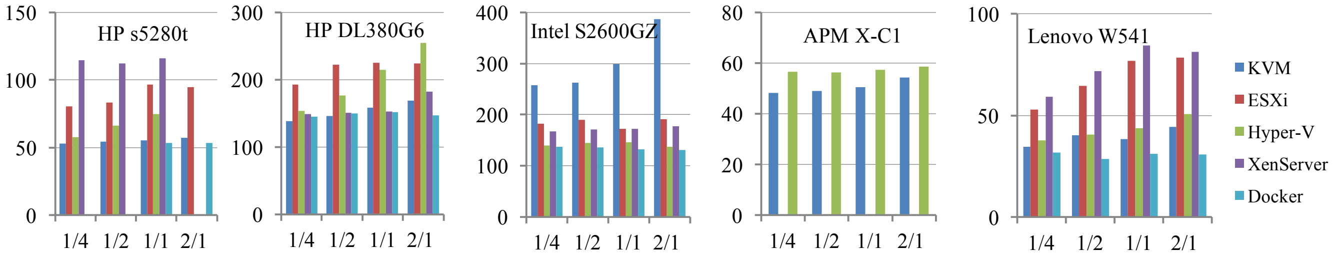
Platform	Energy	Workload level			
		1/4	1/2	1/1	2/1
HP s5280t	Highest	-37.34%	-26.12%	**-42.02%**	-12.95%
	Average	**-15.91%**	-11.78%	-14.32%	-5.93%
	Lowest	**1.41%**	1.02%	0.00%	0.00%
HP DL380 G6	Highest	-13.57%	-12.32%	-19.58%	**-26.35%**
	Average	**-7.25%**	-4.50%	-4.97%	-6.20%
	Lowest	0.00%	0.00%	2.98%	**4.51%**
Intel S2600GZ	Highest	-11.94%	-8.15%	**-16.69%**	0.00%
	Average	**-5.35%**	-1.93%	-5.07%	4.40%
	Lowest	0.00%	2.02%	4.19%	**9.26%**
Lenovo W541	Highest	**-26.11%**	-20.57%	-14.25%	-11.33%
	Average	**-12.42%**	-11.20%	-9.28%	-6.41%
	Lowest	0.00%	0.00%	0.00%	0.00%
APM X-C1	Highest	**-9.13%**	0.00%	-2.42%	0.00%
	Average	-4.78%	0.03%	-1.22%	**10.15%**
	Lowest	0.00%	0.05%	0.00%	**22.60%**
Lenovo RD450	Highest	**-76.43**%	-74.29%	-71.33%	-11.70%
	Average	-34.57%	-32.72%	**-35.19%**	-5.97%
	Lowest	**119.93**%	98.07%	0.00%	0.00%

Note: Positive values imply that Docker consumes more energy.

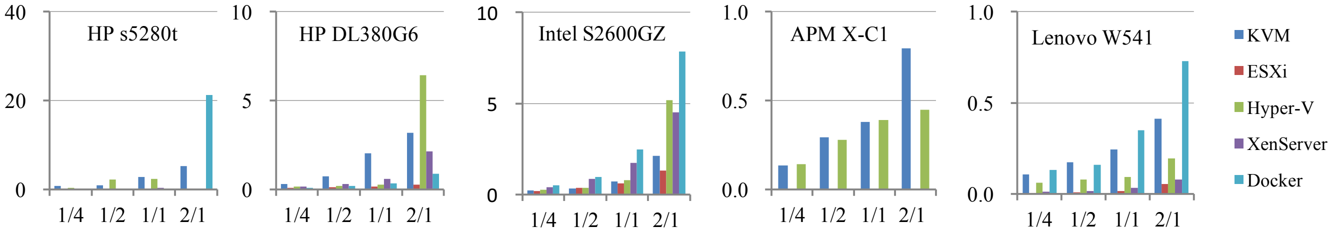
For example, Docker has the highest power consumption when running fair and very heavy computation-intensive workloads on Intel S2600GZ. Docker consumes 8.40% more power than Hyper-V (the lowest), and 3.29% more power than the average power, for very heavy computation-intensive workload on Intel S2600GZ. Further, Docker consumes 6.71% more power than Hyper-V (the lowest), and 3.96% more power than the average power, for fair computation-intensive workload on Intel S2600GZ. From Table 7, we can see that, on average, Docker consumes less power than the highest and average power consumed by all the hypervisors on all the platforms running all workload levels, except that it consumes 0.14% more power than the average power of all the hypervisors. Although Docker consumes less power, it does not consume the lowest power in all the experiments. The heavier the workload level, the higher is the power consumption of Docker (from 0.41% to 4.73%). On average, the power consumption of Docker is very close to the lowest power consumption for the 1/4 and 1/2 workload levels on all the platforms (only 0.41% and 0.87% higher than the lowest ones). However, for the 1/1 and 2/1 workload levels, Docker consumes 2.97% and 4.73% more power than the lowest power consumed by the other hypervisors. On the ARM64 server, Docker has the highest power consumption when running light and very heavy computation-intensive workloads. On Lenovo W541, Docker has the lowest power and energy consumption consistently for all workload levels; Docker outperforms all the other hypervisors in terms of both power and energy only on this platform.

B. Mixed Workloads

The power, energy, and completion time for mixed workloads (LAMP) are shown in Figs. 4 and 5. Because the LAMP workload is Web-server- and database-centric, it uses fewer CPU cycles than PrimeSearch. Although the power consumption of the hard disk and solid-state disks is much lower than that of the processors configured in the same server, the real-time power fluctuates at a higher magnitude than the computation-intensive workload. This means that although the average power of the mixed workload is less than the computation-intensive workload on nearly all the servers for all four workload levels, the power variation during LAMP execution is greater than that during PrimeSearch execution. The power and energy variations (ratio of the highest to the lowest) of different hypervisors running different levels of workloads on all the platforms are summarized in Tables 9 and 10, respectively. For example, ESXi has the highest standard deviation of the real-time power on HP s5280t, HP DL380 G6, and Intel S2600GZ, while KVM has the highest standard deviation of the real-time power on Lenovo W541 and APM X-C1. We present two examples of the power fluctuations in Fig. 6.



(a) Power consumption (x axis: workload level, y axis: power (watts))



(b) Energy consumption (x axis: workload level, y axis: energy (10^6 joules))

Fig. 4 Power and energy consumption of mixed workloads

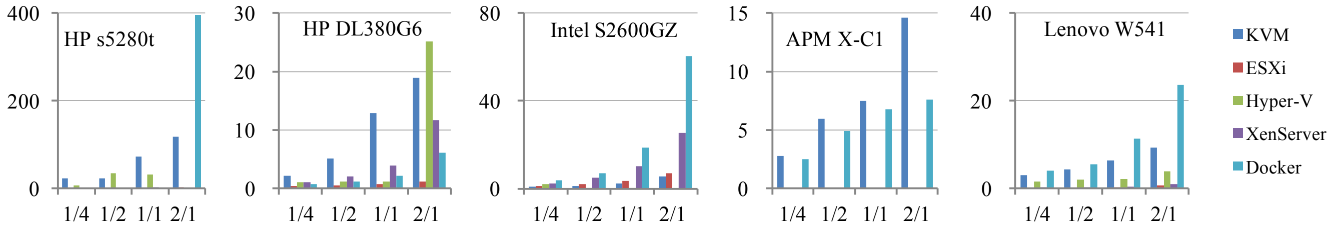


Fig. 5 Completion time of mixed workloads (x axis: workload level, y axis: completion time (10^3 seconds))

Table 9. Power variation of mixed workloads (ratio of highest to lowest)

Platform	1/4 workload	1/2 workload	1/1 workload	2/1 workload
HP s5280t	**116.51%**	107.00%	110.24%	65.35%
HP DL380 G6	38.95%	52.22%	47.74%	**73.44%**
Intel S2600GZ	88.62%	92.58%	126.10%	**196.85%**
Lenovo W541	85.86%	118.48%	**171.63%**	164.87%
APM X-C1	**17.11%**	15.21%	13.45%	7.98%
Lenovo RD450	**33.08%**	27.35%	25.22%	26.49%

Table 10. Energy variation of mixed workloads (ratio of highest to lowest)

Platform	1/4 workload	1/2 workload	1/1 workload	2/1 workload
HP s5280t	779.35%	**2702.96%**	2060.54%	2423.03%
HP DL380 G6	279.90%	592.76%	1093.76%	**2303.57%**
Intel S2600GZ	152.20%	200.32%	305.40%	**490.85%**
Lenovo W541	1699.02%	1814.92%	**2288.77%**	1256.78%
APM X-C1	5.17%	5.09%	3.25%	**77.83%**
Lenovo RD450	276.28%	323.48%	436.09%	**486.52%**

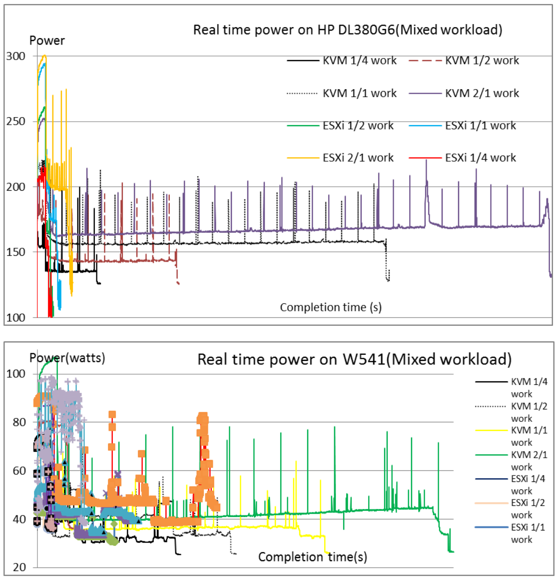


Fig. 6 Real-time power (incl. 2 minutes after completion)

As we have already seen in Table 2, the power variation on a typical rack server is 8.4%~19.79% for very heavy computation-intensive workloads, compared with the lowest power consumption. However, from Tables 9 and 10, we can see that the power and energy variations of mixed workloads are much greater than those of computation-intensive workloads in Tables 2 and 3. Except for the much higher power and energy variations among all the hypervisors on the same platform running mixed workloads, the power variations during the execution of the mixed workloads are greater than those of the computation-intensive workloads.

From Tables 9 and 10, we also observe that the three 2U servers have the largest power and energy variations for the very heavy 2/1 mixed workload level, except for the power variation on Lenovo RD450.

The average power and energy standard deviations of different hypervisors running different levels of workloads on different platforms are listed in Tables 11 and 12, respectively.

Table 11. Power standard deviation of computation-intensive workloads (compared to average power of the whole execution)

	KVM	ESXi	Hyper-V	XenServer	Docker
HP s5280t	1.72%	0.95%	2.60%	0.49%	0.80%
HP DL380 G6	1.65%	1.16%	1.22%	0.95%	1.39%
Intel S2600GZ	3.83%	5.05%	1.02%	2.66%	4.82%
Lenovo W541	3.25%	0.78%	2.96%	2.76%	1.81%
APM X-C1	0.66%	N/A	N/A	N/A	1.22%
Lenovo RD450	6.25%	4.59%	1.50%	4.51%	2.86%

Note: Smaller values imply that the power fluctuation when running the benchmark is lower.

Table 12. Power standard deviation of mixed workloads (compared to average power of the whole execution)

	KVM	ESXi	Hyper-V	XenServer	Docker
HP s5280t	6.54%	11.26%	7.64%	N/A	N/A
HP DL380 G6	6.24%	19.66%	6.35%	7.64%	1.61%
Intel S2600GZ	7.89%	29.44%	2.24%	3.64%	2.79%
Lenovo W541	19.98%	5.22%	16.95%	8.87%	7.04%
APM X-C1	2.13%	N/A	N/A	N/A	1.11%
Lenovo RD450	5.43%	3.48%	0.70%	1.60%	2.34%

Note: Smaller values imply that the power fluctuation when running the benchmark is lower.

We also observe that different hypervisors may drop mixed workloads owing to the job scheduling timeout threshold. The success rates of 1 million database insertion operations in mixed workload conditions are listed in Table 13.

Table 13. Average success rate of database insertion operations in mixed workload conditions for all workload levels

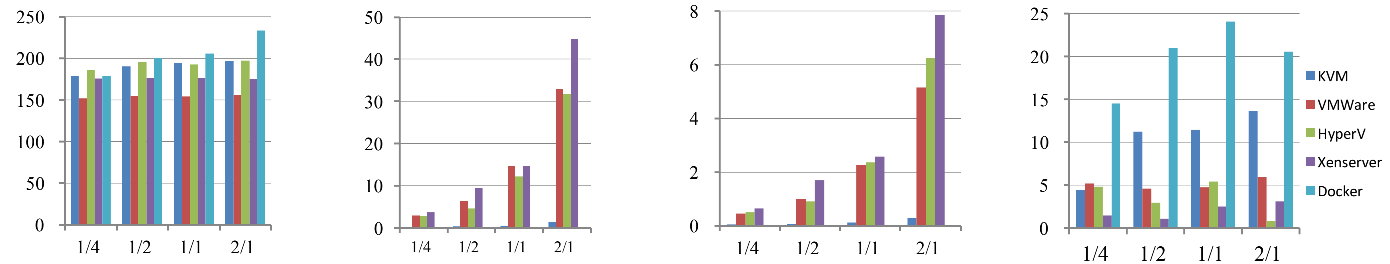
	HP DL380G6	Intel S2600GZ	Lenovo RD450
KVM	1.000	0.979	1.000
ESXi	0.902	0.891	1.000
Hyper-V	0.701	0.767	0.434
XenServer	1.000	1.000	1.000
Docker	0.811	0.720	1.000

Lenovo RD450 is equipped with four HDDs, which help improve the success rate. However, as can be seen in Table 13, Hyper-V has nearly the lowest success rate among the three 2U servers. Moreover, Docker has a lower success rate than the other hypervisors except Hyper-V. The success rate plays an important role if we only consider the power and energy efficiencies of the hypervisors during hypervisor selection for system deployment in data centers.

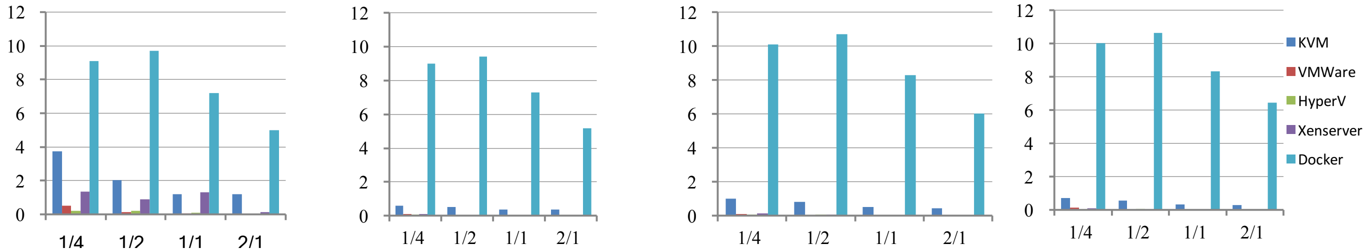
C. Memory-intensive Workloads

We selected the Lenovo RD450 server as a testing platform for the STREAM benchmark because this server is the most recently released server in our server list. The STREAM results are shown in Fig. 7. From Fig. 7(1), we can see that Docker has the highest power consumption for all workload levels except the 1/4 workload level. For example, Docker consumes 3%, 12%, 15%, and 29% more power than the average power of the other hypervisors. Because Docker does not use memory virtualization, it is the first to complete all memory-intensive operations. For example, the average completion times of KVM, VMware ESXi, Hyper-V, and XenServer are 116, 220, 210, and 375 times the completion time of Docker. If we only consider KVM, VMware ESXi, Hyper-V, and XenServer, then KVM has the shortest completion time. For example, the average completion times of VMware ESXi, Hyper-V, and XenServer are 10, 18, 24, and 26 times the average completion time of KVM.

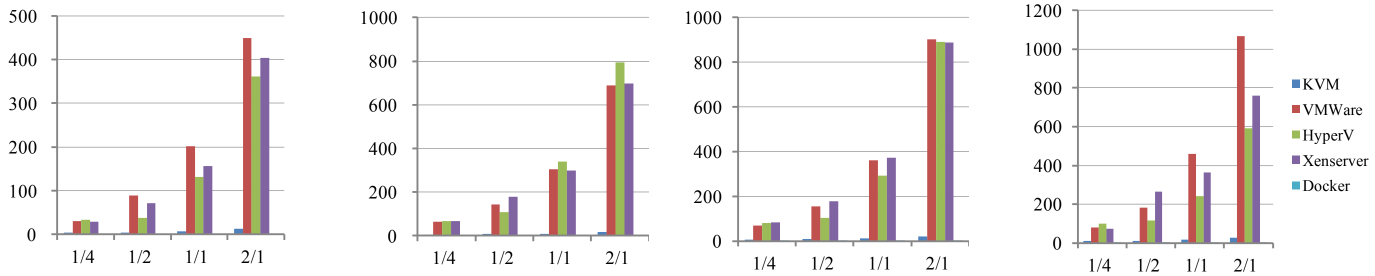
Moreover, Docker has the highest power deviation because it has the shortest completion time for all workload levels. Although Docker has the highest power consumption, it outperforms the other hypervisors in terms of the best rate and completion time. The achieved bandwidth of Docker is compared with that of the other hypervisors in Table 14. We can see that Docker can achieve 6–75 times the bandwidth of the other hypervisors for all memory-intensive operations.



(1)power consumption(watts) (2)completion time(x 1000s) (3)energy consumption(x million joules) (4) power deviation(watts)



(5)best rate of copy (GB/s) (6)best rate of scale (GB/s) (7)best rate of add (GB/s) (8)best rate of triad (GB/s)



(9)average time of copy(s) (10)average time of scale(s) (11)average time of add (s) (12)average time of triad(s)

Fig.7 Results of memory-intensive workloads

Table 14. Ratio of achieved memory bandwidth of Docker to average value of other hypervisors (times)

load levels / operations	1/4	1/2	1/1	2/1
copy	6	12	11	14
scale	42	59	66	50
add	29	44	57	50
triad	39	60	75	74

For convenient comparison, we summarize the highest and lowest power and energy consumptions of mixed workloads on all platforms in Tables 15 and 16, respectively. We can see that on the newer servers, Docker has nearly the lowest power consumption, especially on the 2U servers, except for the 1/4 and 1/2 workload levels on HP DL380 G6. However, ESXi always has the lowest energy consumption because it has the shortest completion time, except for the 1/4 workload level on Lenovo RD450 and the 1/2 workload level on Intel S2600GZ.

Table 15 Highest and lowest power consumptions of mixed workloads on all platforms

Platform	Power	Workload level			
		1/4	1/2	1/1	2/1
HP s5280t	Highest	XenServer	XenServer	XenServer	ESXi
	Lowest	KVM	KVM	Docker	Docker
HP DL380 G6	Highest	ESXi	ESXi	ESXi	Hyper-V
	Lowest	KVM	KVM	Docker	Docker
Intel S2600GZ	Highest	KVM	KVM	KVM	KVM
	Lowest	Docker	Docker	Docker	Docker
Lenovo W541	Highest	XenServer	XenServer	XenServer	XenServer
	Lowest	Docker	Docker	Docker	Docker
APM X-C1	Highest	Docker	Docker	Docker	Docker
	Lowest	KVM	KVM	KVM	KVM
Lenovo RD450	Highest	Hyper-V	Hyper-V	Hyper-V	Hyper-V
	Lowest	Docker	Docker	Docker	Docker

Table 16 Highest and lowest energy consumptions of mixed workloads on all platforms

Platform	Power	Workload level			
		1/4	1/2	1/1	2/1
HP s5280t	Highest	Docker	Docker	Docker	Docker
	Lowest	ESXi	ESXi	ESXi	ESXi
HP DL380 G6	Highest	KVM	KVM	KVM	Hyper-V
	Lowest	ESXi	ESXi	ESXi	ESXi
Intel S2600GZ	Highest	Docker	Docker	Docker	Docker
	Lowest	ESXi	KVM	ESXi	ESXi
Lenovo W541	Highest	Docker	Docker	Docker	Docker
	Lowest	ESXi	ESXi	ESXi	ESXi
APM X-C1	Highest	Docker	KVM	Docker	KVM
	Lowest	KVM	Docker	KVM	Docker
Lenovo RD450	Highest	Hyper-V	XenServer	XenServer	XenServer
	Lowest	Docker	ESXi	ESXi	ESXi

D. *Energy Efficiency of Hypervisors on Typical 2U Servers*

In our experiments, we selected three typical 2U servers, namely HP DL380 G6, Intel S2600GZ, and Lenovo RD450. Such 2U servers are widely used in existing data centers. To compare the energy efficiencies of the hypervisors on these servers, we compared their power and energy consumptions for computation-intensive, memory-intensive, and mixed workloads.

From Tables 5 and 15, we can see that no single hypervisor consumes the lowest power among all the hypervisors under the same workload level on the same 2U server.

Because these 2U servers have been developed in different years, we calculated the average energy consumption of one virtual machine under 1/1 and 2/1 workload levels to investigate whether there is a trend in energy efficiency improvement with time. The average energy consumptions of one virtual machine for computation-intensive and mixed workloads are listed in Tables 17(for the first execution interval from 1 to 1000000) and 18, respectively. The values in bold are the lowest values and the values in bold italics are the highest values.

Table 17 Average energy consumption of one VM running computation-intensive workload (unit: joules)

Platform	HPDL380		INTEL S2600GZ		Lenovo RD450	
Load level	1/1	2/1	1/1	2/1	1/1	2/1
KVM	**4561**	**4650**	***4777***	***3885***	2501	**2136**
ESXi	4788	5006	**3445**	**2153**	2433	2203
Hyper-V	***6255***	***7262***	4331	3342	***8041***	***2318***
XenServer	4929	4769	3705	3157	2504	2179
Docker	4639	4850	3727	3815	**2305**	2047

Table 18 Average energy consumption of one VM running mixed workload (unit: joules)

Platform	HPDL380		INTEL S2600GZ		Lenovo RD450	
Load level	1/1	2/1	1/1	2/1	1/1	2/1
KVM	***256344***	200136	44037	66538	15719	12901
ESXi	**21474**	**16693**	**20918**	**21223**	**8465**	**7955**
Hyper-V	32231	***401234***	49946	***161722***	23278	16909
XenServer	75714	133409	109498	141368	***45381***	***46660***
Docker	41339	55734	***155131***	245829	8900	8169

From Tables 17 and 18, we can see that although the newer server tends to be more energy-efficient owing to technological improvements, such energy efficiency improvement may be masked by different hypervisors under different workload levels. However, our experimental results show that more energy-efficient servers consume less (but not always the least) energy for the same VM workload, compared with the older servers. Thus, we need to carefully select the hypervisors for specific workloads to achieve further energy reduction even on newer hardware platforms.

E. *Insights on Energy Efficiency of Hypervisors*

From the above-mentioned observations, we gain the following insights for hypervisor selection in data centers in terms of power and energy consumption.

Insight #1: The hypervisor, hardware, and workload type are coupled with each other, and such complication requires system designers to be mindful of virtualized infrastructure and cloud data centers to carefully select hypervisors.

From the experiment results presented above, we can see that different hypervisors exhibit different energy efficiencies on different platforms when they run different types of workloads.

For example, Hyper-V has the highest power and energy on the rack server HP DL380 G6 for computation-intensive workloads, whereas ESXi has the highest power for mixed workloads and KVM has the highest energy consumption on it (except 2/1 workload level).

Furthermore, all the hypervisors have similar power and energy consumption on the rack server Intel S2600GZ for computation-intensive workload where the power variation is 3.03%–8.40% and energy variation is 9.26%–25.06% (see Tables 2 and 3). However, KVM has the highest power consumption and Docker has the highest energy consumption on it for mixed workloads (see Fig. 4). In addition, the power variations are much larger than those of the computation-intensive workloads (see Tables 9 and 10).

Insight #2: The power and energy efficiencies change with the workload level.

As can be seen in Tables 5, 6, and 15, a hypervisor that consumes the lowest power or energy for light workload may consume the highest power or energy if the workload level increases drastically and vice versa.

For example, consider computation-intensive workload. Hyper-V has highest power consumption only on HP DL380 G6 for all computation-intensive workload levels. However, Hyper-V also has the lowest power consumption on Intel S2600GZ for the 1/1 and 2/1 computation-intensive workload levels. Similarly, ESXi has the highest power consumption on Lenovo W541 for all computation-intensive workload levels and on S2600GZ for the 1/4 and 1/2 workload levels. However, ESXi has the lowest power consumption on HP s5280 for the 1/4 and 1/2 workload levels.

Insight #3: ESXi should be deployed in a non-power-sensitive environment to achieve high computing performance, while KVM should be deployed in a power-sensitive environment in order to deploy as many virtual machines as possible and achieve reasonable computing performance.

Insight #4: Typical 2U servers have a higher idle power percentage and should always be running with heavy virtual machine workload because their idle power percentage decreases when the workload increases.

Insight #5: The newly manufactured highly energy-efficient servers consume less (but not always the least) energy for the same VM workload, compared with older servers. Thus, we need to carefully select the hypervisors for specific workloads to achieve further energy reduction even on newer hardware platforms.

Insight #6: Although the ARM64 server has lower max power than a laptop, it has a higher idle power percentage and a less dynamic range. The ARM64 server should be deployed in a low-contending environment, where virtual machines have light computation-intensive workload and the power supply is very precious, as the surplus power capacity for typical 2U servers is insufficient. In other words, the ARM64 server should be deployed in a steady power usage server room, while a laptop (customized mobile server) should be deployed for highly dynamic workloads to leverage and complement the power fluctuation due to workload variations. For example, the ARM64 server configured with the KVM virtualization environment consumes 71.1% less power and 25.4% less energy than the HP DL380 G6 rack server (both packaged with eight processor cores) for the same computation-intensive workload, while the execution time of the ARM64 server is longer by 158%.

Insight #7: ESXi uses power more actively to achieve high performance and high throughput, especially in highly contending conditions (for very stressful workload). ESXi and XenServer consume power more actively than Hyper-V and KVM on customized mobile servers.

Insight #8: ESXi-based virtual machines are migration candidates for power shifting or capping conditions because ESXi uses power more actively during the early stage of mixed workload experiment execution on a typical 2U server.

We calculated the energy per database insertion for all the hypervisors. On all the platforms and at all workload levels, ESXi consumes the least energy among all the hypervisors. For example, ESXi uses 84.57% less energy than KVM for each insertion on HPDL380. On Lenovo W541, ESXi and XenServer use 92.02% and 86.35% less energy than KVM for each insertion.

We calculated the accumulated energy consumption of each experiment for each workload level. The average power of ESXi, Hyper-V, and XenServer is 72.52%, 9.38%, and 88.39% higher than that of KVM. The significant time reduction results in energy savings of 92.2%, 52.49%, and 86.35% for ESXi, Hyper-V, and XenServer compared with KVM, respectively. However, ESXi cannot complete all the database insertions under heavy workload, i.e., **1/1** and **2/1** workload levels. ESXi completes 88.78% insertions for the **1/1** workload level and 72% insertions for the **2/1** workload level.

IV. RELATED WORK

To the best of our knowledge, this is the first study to compare the energy efficiencies of four mainstream hypervisors as well as a container engine across multiple platforms with different workload types and workload levels. We briefly review some related studies in this section.

Several studies have compared the performances of hypervisors [28, 29, 30, 32, 33, 34, 35]. All these studies have compared the performances of different hypervisors or the native physical performances on a single platform. By contrast, in the present study, we compared the energy efficiencies of different hypervisors and a container engine on different platforms. Some studies have compared the power and energy efficiencies of hypervisors [36, 37, 38, 39]. In [36, 37], the energy overheads of Xen and KVM were compared when running three virtual machines on a small server with 2 GB memory and a 500 GB hard disk. The results reflect the energy overheads of Xen and KVM compared with a physical machine. In [38], the power consumptions of Xen, KVM, Docker, and LXC were compared when running up to eight virtual machines on a desktop server with four processor cores and 12 GB memory; the emphasis was on the network traffic benchmark and the workload level was fixed. In [39, 40, 41], the energy efficiencies of Xen, KVM, and VMware were compared when running dedicated high-performance computation-intensive workloads. In [42, 43], the power consumptions of KVM, Xen, and OpenVZ were compared with those of non-virtualized environments when running network transactions.

What distinguishes our work from the above-mentioned studies is that we have compared the energy efficiencies of XenServer, KVM, VMware ESXi, Hyper-V, and Docker on various platforms, including high-end rack servers, a desktop server, a laptop, and an emerging ARM64 server. We ran computation intensive, memory intensive and mixed Web server benchmarks with varying workload levels on these platforms to investigate the differences in the energy efficiencies of these hypervisors. Moreover, we tried to mimic real multi-tenant cloud computing environments with massive virtual machines at different workload levels.

In summary, our study is more comprehensive, as it is based on fine-grained power samples from six platforms. In addition, we have presented the power variations for different workload levels on different platforms.

V. CONCLUSIONS

Understanding the energy efficiency of hypervisors on different servers under different workloads can facilitate the tasks of data center designers and system operators in many ways, including system capacity planning, power shifting, virtual machine placement, migrations, and resource scheduling. In this study, we conducted extensive experiments with different workload types and levels to emulate a realistic virtualized multi-tenant cloud environment. We used power and energy measurements to investigate the power and energy characteristics of different mainstream hypervisors on different types of servers. Our results showed that hypervisors exhibit different power and energy characteristics on the same hardware with the same workload. Moreover, different hypervisors exhibit different attributes and align with different workload types and workload levels. In addition, they may be deployed for different workload levels in different power situations. Our results also showed that container virtualization is a lightweight technique in terms of system implementation and maintenance, but essentially not more power-efficient than conventional virtualization technology. Finally, although ARM64 servers have low power consumption, they require long execution times to complete computing jobs and sometimes consume a large amount of energy as well. Thus, laptop processors and motherboards are strong competitors of the ARM64 server in terms of both power and energy consumption.

In the future, we would like to investigate the instruction execution of different hypervisors for power and energy characterization and profiling.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of China (No. 61472109), Natural Science Fund of Zhejiang Province (LY14F020045), and Fund of Department of Education of Zhejiang Province (Y200907182). This work was also supported in part by National Science Foundation (NSF) grants CNS-1205338 and CNS-1561216, and by the Introduction of Innovative R&D Team Program of Guangdong Province (No. 201001D0104726115).

This paper is a significantly extended version of the conference paper presented at the IEEE 7th Green and Sustainable Computing Conference (IGSC 2016) [45].

REFERENCES

[1] S. Subramanya, Z. Mustafa, D. Irwin, P. Shenoy, Beyond energy-efficiency: evaluating green datacenter applications for energy-agility, ACM/SPEC ICPE 2016, pp. 185–196.

[2] T. Mastelic, A. Oleksiak, H. Claussen, et al. Cloud computing: survey on energy efficiency, ACM Comput. Surv. 47(2) (2014) 1–36.

[3] P. Pande, F. Clermidy, D. Puschini, I. Mansouri, P. Bogdan, R. Marculescu, A. Ganguly, Sustainability through massively integrated computing: are we ready to break the energy efficiency wall for single-chip platforms?, Proc. Des. Autom. Test Europe (DATE), 2011, pp. 1656–1661.

[4] Z. Zhang, S. Fu, Characterizing power and energy usage in cloud computing systems, IEEE CloudCom 2011, pp. 146–153.

[5] S. Deb, K. Chang, X. Yu, et al., Design of an energy efficient CMOS compatible NoC architecture with millimeter wave wireless interconnects, IEEE Trans. Comput. 62(12) (2012) 2382–2396.

[6] D. Fan, Z. Tang, H. Huang, Hailin Huang, G. R. Gao, An energy efficient TLB design methodology, Proc. 2005 Int. Symp. Low Power Electron. Des. (ISLPED), 2005, pp. 351-356

[7] R. Rodrigues, A. Annamalai, I. Koren, et al., Reducing energy per instruction via dynamic resource allocation and voltage and frequency adaptation in asymmetric multicores, IEEE ISVLSI 2014, pp. 436–441.

[8] M. Bi, S. Chandrasekharan, C. Gniady, IAMEM: interaction-aware memory energy management, USENIX ATC 2013, pp. 267–278.

[9] Q. Deng, D. Meisner, A. Bhattacharjee, et al., CoScale: coordinating CPU and memory system DVFS in server systems, IEEE/ACM MICRO-45, 2012, pp. 143–154.

[10] D. Tiwari, S. Vazhkudai, Y. Kim, et al., Reducing data movement cost using energy-efficient active computation on SSD, USENIX FAST 2013, pp. 119–132.

[11] Y. Zhang, J. Liu, M. Kandemir, Software-directed data access scheduling for reducing disk energy consumption, Proc. 32nd IEEE Int. Conf. Distrib. Comput. Syst. (ICDCS), 2012, pp. 596–605.

[12] R. Krishnaswamy, V. Nagarajan, K. Pruhs, et al., Cluster before you hallucinate: approximating node-capacitated network design and energy efficient routing, ACM STOC 2014, pp. 734–743.

[13] Y. Ohsita, M. Murata, Optical data center networks; architecture, performance, and energy-efficiency, Handbook on Data Centers, 2015.

[14] T. Banerjee, S. Ranka, A genetic algorithm based autotuning approach for performance and energy optimization, IEEE IGSC 2015, pp. 1– 8.

[15] C. Wang, N. Nasiriani, G. Kesidis, et al. Recouping energy costs from cloud tenants: tenant demand response aware pricing design. ACM e-Energy 2015, pp. 141-150

[16] Y. Guo, D. Zhu and H. Aydin, Reliability-aware power management for parallel real-time applications with precedence constraints, IEEE IGCC 2011, pp.1-8.

[17] D. Tiwari, S. Boboila, S. Vazhkudai, et al., Active flash: towards energy-efficient, in-situ data analytics on extreme-scale machines, USENIX FAST, 2013, pp. 119–132.

[18] C. Jiang, L. Duan, C. Liu, et al., VRAA: virtualized resource auction and allocation based on incentive and penalty. Cluster Comput., 16(4) (2013)639-650.

[19] H. Sun, P. Stolf, J.M. Pierson, et al., Energy-efficient and thermal-aware resource management for heterogeneous datacenters, Sust. Comput.: Inform. Syst. (2014) 1–15.

[20] W. Huang, M. Allen-Ware, J.B. Carter, et al., TAPO: Thermal-aware power optimization techniques for servers and data centers, IEEE IGCC 2011, pp. 1–8.

[21] S. Yeo, M. M. Hossain, J.-C. Huang, et al., ATAC: ambient temperature-aware capping for power efficient datacenters, ACM SOCC 2014, pp. 1–14.

[22] J. Carter, K. Rajamani, Designing energy-efficient servers and data centers, Comput. 43(7) (2010) 76–78.

[23] A. Gandhi, V. Gupta, M. Harchol-Balter, et al., Optimality analysis of energy-performance trade-off for server farm management, Perform. Eval. 67(11) (2010) 1155–1171.

[24] I. Goiri, W. Katsak, K. Le, et al., Designing and managing datacenters powered by renewable energy, IEEE Micro 34(3)(2014) 8–16.

[25] L. Li, W. Zheng , X. Wang, et al., Data center power minimization with placement optimization of liquid-cooled servers and free air cooling, Sustain. Comput.: Inform. Syst. (2016).

[26] A. Khosravi, S. Kumar Garg, R. Buyya, Energy and carbon-efficient placement of virtual machines in distributed cloud data centers, LNCS 8097, pp. 317–328.

[27] M. A. Islam, S. Ren, A new perspective on energy accounting in multi-tenant data centers, USENIX CoolDC 2016.

[28] I. Ahmad, J. Anderson, A. Holler, An analysis of disk performance in VMware ESX server virtual machines, IEEE WWC 2003.

[29] T. Deshane, Z. Shepherd, J. Matthews,et al., Quantitative comparison of Xen and KVM, XenSummit 2008.

[30] R. Morabito, J. Kjallman, M. Komu, Hypervisors vs. lightweight virtualization: a performance comparison, IEEE IC2E 2015, pp. 386–393.

[31] Natural Resources Defense Council, Data center efficiency assessment: scaling up energy efficiency across the data center industry: evaluating key drivers and barriers, 2014.

[32] J. Hwang, S. Zeng, F. Wu, Timothy Wood, A component-based performance comparison of four hypervisors, IFIP/IEEE IM 2013, pp. 269–276.

[33] J. Li ,Q. Wang,D. Jayasinghe, et al., Performance overhead among three hypervisors: an experimental study using Hadoop benchmarks, IEEE Big Data 2013, pp. 9–16.

[34] VMware. A performance comparison of hypervisors. https://www.vmware.com/pdf/hypervisor_performance.pdf

[35] W. Felter, A. Ferreira, R. Rajamony, et al., An updated performance comparison of virtual machines and linux containers. IBM Research Report, RC25482, 2014.

[36] Y. Jin, Y. Wen, Q. Chen, et al., An empirical investigation of the impact of server virtualization on energy efficiency for green data center, Comput. J. (2013) 1–14.

[37] Y. Jin, Y. Wen, Q. Chen, Energy efficiency and server virtualization in data centers: an empirical investigation, INFOCOM WKSHPS 2012, pp. 133–138.

[38] R. Morabito, Power consumption of virtualization technologies: an empirical investigation, IEEE/ACM UCC 2015, pp. 522–527.

[39] S. Varrette, M. Guzek, V. Plugaru, et al., HPC performance and energy-efficiency of Xen, KVM and VMware Hypervisors, SBAC-PAD 2013, pp. 89–96.

[40] M. Guzek , S. Varrette, V. Plugaru, A holistic model of the performance and the energy-efficiency of hypervisors in an HPC environment, LNCS 8046, pp. 133–152.

[41] M. Guzek , S. Varrette, V. Plugaru, et al. A holistic model of the performance and the energy-efficiency of hypervisors in an HPC environment, Concurr. Comput. 26(15) (2014) 2569–2590.

[42] R. Shea, H. Wang, J. Liu, Power consumption of virtual machines with network transactions: measurement and improvements, IEEE INFOCOM 2014, pp. 1051–1059.

[43] C. Xu, Z. Zhao, H. Wang, Energy efficiency of cloud virtual machines: from traffic pattern and CPU affinity perspectives, IEEE Syst. J. PP (99) (2015) 1–11.

[44] https://www.cs.virginia.edu/stream/

[45] C. Jiang, D. Ou, Y. Wang, X. You, J. Zhang, J. Wan, B. Luo, W. Shi, Energy efficiency comparison of hypervisors, Proc. 7th IEEE Int. Green Sustain. Comput. Conf. (IGSC 2016), 2016, pp. 1–8.