Virtual machines (VM) have been introduced early in the 1960s by IBM to consolidate the hardware and decrease exploitation costs [7]. The mainframes were sold at about $2.9 million (equivalent to about $25 million in 2020) and rented for $63,500 (about $553,417 in 2020) per month in a typical configuration, making computing systems running on personal computers to enterprise and commercial systems with entirely different instruction set architectures than the underlying hardware. VMs typically run above a software layer called “Virtual Machine Monitor” (VMM) or simply hypervisor. It controls the run-time resources of the VMs and ensures proper execution of privileged instructions.

The x86 architecture separates processor privileges with a protection ring or levels [3]. It is a mechanism that protects data and restricts operations that programs can run. Each program that executes in an x86 system is assigned to a specific ring or level that defines the access privileges on system resources. Figure 1 shows the different privilege levels available in x86 architectures. Typically, level 0 is reserved for the operating system (OS) services that directly interface with the hardware (kernel mode). Levels 1 and 2 are mostly unused and are reserved for some drivers and middleware. User applications run at level 3 (user mode) [3]. In Figure 1(a), no virtualization is implemented. The user applications run at level 0 and the kernel of the OS handles privileged instructions at level 0. Executing at level 0 allow the kernel to directly access and control the underlying hardware. In reality, it was just a way to transparently share resources and run workloads from different users in an isolated way on the same hardware. A few decades later, researchers investigated models, challenges, and solutions to efficiently implement “virtual sub-environments” in physical machines [4]. The VM abstraction then provided concurrent and interactive access to the underlying hardware.

The continuous innovation in virtualization technology has led to the emergence of an ecosystem of products ranging from VMs running on personal computers to enterprise and commercial systems running in the cloud. Virtualization concepts are also applied beyond traditional hardware devices such as processors, memory, disk, and network cards. As example, some research propose to virtualize Field-Programmable Gate Arrays (FPGA) for cloud and data center applications [14–16]. Graphic Processing Units (GPU) are also provisioned as part of virtual resource pools [10, 11]. Among the most common virtualization softwares are VirtualBox, KVM, QEMU, Xen, VMware workstation, and container engines such as Docker and LXD. The emergence of multiple virtualization systems supporting hardware consolidation in personal computers, embedded systems, and cloud-scale deployments raise the need for architecture classification and performance evaluation. In the context of this work, we study the architectures of state-of-the-art virtualization systems and provide a quantitative evaluation of the performance that can be achieved against IO-intensive (such as applications intensively accessing the disk), memory-intensive (such as matrix-based applications), and compute-intensive benchmarks (such as high-performance applications). We also evaluate the overhead introduced by virtualization technologies against native executions.

1 INTRODUCTION

Virtual machines (VM) have been introduced early in the 1960s by IBM to consolidate the hardware and decrease exploitation costs [7]. The mainframes were sold at about $2.9 million (equivalent to about $25 million in 2020) and rented for $63,500 (about $553,417 in 2020) per month in a typical configuration, making computing systems only accessible to a small range of customers [12, 23]. A VM could be seen as an instance of the physical machine in which the users had the illusion of fully owning the hardware. In reality, it was just a way to transparently share resources and run workloads from different users in an isolated way on the same hardware.

2 BACKGROUND

2.1 Type of Virtual Machine Monitors

VMs have several advantages among which easy maintenance, fast recovery from fault, rapid provisioning and domain isolation [2]. They allow running multiple operating systems simultaneously on the same machine. Furthermore, they support the execution of systems with entirely different instruction set architectures that of the underlying hardware. VMs typically run above a software called “Virtual Machine Monitor” (VMM) or simply hypervisor. It controls the run-time resources of the VMs and ensures proper execution of privileged instructions.
the hardware. Depending on how far apart the VMM is from the actual hardware in the x86 privilege levels, we consider two types of hypervisors [8, 9]: (1) **Type-1 hypervisor** (bare metal): the VMM is installed directly above the hardware (see Figure 1(b)). Examples of such VMMs include Xen and Linux enabled by Kernel-based (KVM) [19]. The VMM is responsible for emulating the privileged instructions launched in the guest space. (2) **Type-2 hypervisor (hosted)**: in this configuration, the VMM is installed in the host OS (see Figure 1(c)). An example of this category is VMware Workstation. Privileged instructions in the guest space typically cause a "world switch" to the host kernel under the supervision of the VMM. In general, a set of applications or/and drivers implemented in the VMM are used to access kernel privileged instructions.

### 2.2 VMware Workstation

VMware Workstation is a Type-2 hypervisor that runs on x86 processors. It supports Windows and Linux hosts, and allows users to run multiple VMs on a single machine [1]. It virtualizes IO devices using a hosted IO model which consists in taking advantage of pre-existing support in the host OS. This approach has several advantages among which application portability and consistency. It also delivers near native performance for CPU-intensive workloads. Figure 2 summarizes the architecture of VMware workstation. Non-privileged instructions from the guest can run natively on the hardware without interference from the VMM. On the other hand, when guest applications issue privileged instructions, the VMM traps and emulates. Specifically, the VMM requests a "world switch" from the VM Driver. Next, the VMM provides data to the VM App. The VM App is then in charge of mapping the virtual requests to the hardware. QEMU internal data structure, thus ensuring a smooth execution of guest applications [3]. This whole emulation however comes with a considerable overhead. In a comparative study, Weber et al. reported that QEMU was up to 5x slower than native environment on some compute-intensive applications [22].

### 2.3 Kernel-based Virtual Machine

Kernel-based Virtual Machine (KVM) is a virtualization module present in Linux releases since kernel version 2.6.20. It represents the latest generation of open source virtualization utility. It transforms a Linux system into a Type-1 hypervisor and benefits from decades of innovation in Linux process scheduling, memory management, device drivers, etc — to manage VMs [19]. It requires processors with virtualization extensions such as Intel VT or AMD-V. To emulate processors and IO devices, KVM is combined with QEMU (Quick Emulator) [3]. IO communication between the virtual and physical system is done through VirtIO. VirtIO is an abstraction of IO devices implemented by Rusty Russel for communication interfaces between guests and host in paravirtualized architectures. KVM uses VirtIO as paravirtualized device drivers since kernel version 2.6.25 [3, 20]. Figure 3 highlights the key components of the KVM-QEMU virtualization. To execute guest applications on the physical hardware, QEMU creates POSIX threads that represent the virtual CPUs. It has the advantage of making virtual applications appear as processes in the host environment. The guest applications are run via KVM kernel modules that provide extension support for hardware virtualization such as Intel VMX [21]. Specifically, QEMU opens the device file /dev/kvm exposed by KVM kernel module and runs a set of ioctl() functions. These functions allow setting and updating the state of each virtual CPU in the QEMU internal data structure, thus ensuring a smooth execution of guest applications [3]. This whole emulation however comes with a considerable overhead. In a comparative study, Weber et al. reported that QEMU was up to 5x slower than native environment on some compute-intensive applications [22].

### 2.4 Containers: Docker

#### 2.4.1 Containers

Containers are virtualization technologies in which the virtual environment directly runs above the host OS. They run within a container engine instead of an hypervisor. They are not designed to run a complete systems, but focus at the application level. Containers are developed to reduce the footprint of systems, especially those that do not need heavy virtualization infrastructures. Figure 4(a) and (b) show the typical virtualization stacks for VMs. Next, Figure 4(c) illustrate the key difference between container and VM stacks. It resides in that containers only run applications on top of a container engine instead of a hypervisor. Containers only need application binaries and a run-time engine, while VMs require support to run entire guest OSes above the underlying OS or hypervisor. They implement "OS-level virtualization" as opposed to "hardware virtualization" with VMs. This particular feature makes them lightweight and very portable. Containers are well-suited for fast development and deployment of applications as
Table 1: List of the testing applications

| Category | Benchmark [18] | Details | Scales with # CPU cores |
|----------|----------------|---------|------------------------|
| Processor | pts/aobench | It returns the average execution time in seconds. The lower the value is, the better it is. |✗|
| | pts/asmfish | It stresses the processors. It returns a score that represents the average number of nodes per second. The higher the value is, the better it is. |✓|
| Disk | pts/aio-stress | Performs asynchronous I/O operations on the disk. It returns the average throughput in MB/s. The higher the value is, the better it is. |n/a|
| | pts/blogbench | Mimics the load of real-world busy servers on the filesystem with random read, write, rewrite. It returns a score. The higher the value is, the better it is. The benchmark runs a write and a read test. |n/a|
| | pts/compilebench | Simulate common I/O operations on disk. It measures how well the filesystem can maintain directory locality as the disk fills up. It returns the throughput in MB/s. The higher the value is, the better it is. The benchmark runs 3 different tests: Compile, Initial Create, and Read Compile Tree. |n/a|
| | system/iozone | Tests the hard drive and file-system performance. It returns the average throughput in MB/s. The higher the value is the better it is. The benchmark assesses read and write performance. In our experiments, we tested with the following parameters: record size = 1MB and size = 512 MB. |n/a|
| System | pts/phpbench | Performs a large number of simple tests in order to bench various aspects of the PHP interpreter. It returns a score. The higher the value is, the better it is. |✗|
| | pts/stress-ng | It stresses a computing system in various ways. It returns a score in Bogo Ops/s that reflects how well the system reacted. The higher the value is, the better it is. We ran 5 stressors that are: Memory Copy, Matrix Math, Vector Math, Context Switching, and Crypto |✓|

Figure 4: Difference between containers and VMs. (a) VMs running on a bare-metal hypervisor. (b) VMs deployment on a hosted hypervisor. (c) Deployment architecture of containers.

In the next section, we will discuss our methodology for evaluating the performance of the different virtualization technologies that we study.

3 METHOD AND APPROACH

Our approach to carry out the comparative study can be summarized in 4 steps:

1. **Categorizing the major virtualization schemes**: this first step was accomplished in the previous section. The purpose is to limit the scope of our study to a well-defined set of tools implementing the selected virtualization architectures.

2. **Selecting the tools**: in this work, we compare the performance of KVM, VMware workstation, and Docker as they represent examples of Type-1 and Type-2 hypervisors, and container.

3. **Selecting the properties to assess**: we focus on evaluating how the different selected virtualization environments perform against some workloads. We will particularly observe I/O speed by measuring how the disk access time scales with different applications. We will also study the memory consumption when running the same applications across the different environments. Finally, we will look at the processor utilization.

4. **Recording and Analysing results**: we record observations from running the experiments. Next, we present the results and discuss the observed metrics.

2.4.2 Docker. Docker is one of the most popular container technology currently in use [5]. It mainly focuses on improving developer experience and enable the distribution of microservices as images for direct deployment.
4 EXPERIMENTAL EVALUATION

In this section, we present and elaborate on the experimental observations.

4.1 Evaluation Setup

In order to conduct our experiments, we installed the 3 virtualization software stacks in a Dell R7415l EMC server running on a 2.09GHz AMD Epyc 7251 CPU ×16 cores with 64GB of memory and 1TB of hard drive. We installed CentOS-7 64-bit with a kernel of version 3.10.0 to manage the resources of the server. To run virtual machines on KVM, we installed Virtual Machine Manager 1.5.0 and QEMU 2.11.50. We also installed VMware Workstation 15.5.2 or VMWare for brevity. Next, we created a virtual machine with 8GB of RAM, 4 processors, and 40 GB of hard drive (SCSI). We installed the same release of CentOS-7 that was installed on the server. Finally, to conduct experiments on containers, we installed Docker version 1.15.1 on the Dell EMC server.

4.2 Evaluation Applications and Platforms

4.2.1 Benchmark Suite: After the selection of the virtualization tools, one of the most critical tasks consists in selecting the set of testing applications. Because our first concern was to find applications that can run on all of our virtualization tools, we selected the Phoronix Test Suite v9.6.0 (Nittedal) [18]. We downloaded the stable release for Linux and pulled the corresponding Docker image from Docker Hub. It features more than 200 individual test profiles and more than 60 test suites. It provides an interactive command line interface (CLI) that allows running testing applications with well-defined attributes. Results can be saved under multiple formats such as HTML, PDF, and plain text. Table 1 provides the list of benchmarks we use in our experiments. We specifically check IO, processor, and system performance.

4.2.2 Testing Platforms: In all the experiments, we evaluate each of the benchmarks listed in Table 1 on four environments that are: native system, Docker, VMWare, and KVM. It evaluates the performance of the different virtualization systems and provides an overview of the virtualization overhead compared to applications running without virtualization.

4.3 Experiment Analysis

4.3.1 Disk Access Performance. In this section, we discuss experiments related to comparing disk access performance. Our first test focuses on stressing the file system using the pts/aio-stress and system/iozone benchmarks. The results from the executions are summarized in Figure 5. We observed that in all of the three test applications, Docker achieved the highest throughput. Next comes the native execution, followed by VMware and KVM. One explanation is that containers do not need to "trap and emulate" as they run directly above the host OS, giving them a clear advantage over the virtualization with a hypervisor. Containers also outperformed the native execution consistently in the three experiments. This may result from the resource isolation implemented by containers, limiting interference from other processes in the system.

Figure 5: Combined disk performance metrics. "aio-stress" refers to the aio-stress test. "iozone_wr" indicates the write performance test of the iozone benchmark. "iozone_rd" specifies the read performance test of the iozone benchmark.

To assess how the four testing platform would perform against real-world server operations, we use the pts/blogbench benchmark. The results are recorded separately for read and write operations (see Figure 6a). A higher "Score" is equivalent to better performance. In the previous experiment, Docker had the best disk access throughput. Consequently, the initial assumption was to expect Docker to come on top again. However, this time it obtained pretty much the same score as the native execution, both far above VMware and KVM. This trend is repeated when executing the pts/compilebench benchmark (see Figure 7). On the compile phase, Docker has a throughput of 1320 MB/s while the native execution achieves 1230 MB/s. This nevertheless does not necessarily mean that Docker runs faster than the native as the recorded values are averaged by the benchmark suite. When studying the standard deviation returned by each runs, the native shows a deviation of 40.71 MB/s and Docker comes with 22.89 MB/s, which put the two execution in a fairly close range. On the same test, VMware obtained an average throughput that is 804.74 MB/s and KVM achieved 161.3 MB/s. Overall, these IO stressing benchmarks show that the native execution and Docker perform quite similarly. These results were expected as Docker directly runs on the host. However, the experiments also showed that IO-intensive applications hurt VMs. It is explained by the fact that VMMs consume significant run-time resources to handle context switches as the benchmarks continually attempt accessing the hardware through system calls. Overall, managing IO device accesses degrades the performance of the applications running in the VMs.
Table 2: Stress-ng execution results

|               | Memory Copy | Matrix Math | Vector Math | Context Switching | Crypto |
|---------------|-------------|-------------|-------------|-------------------|--------|
|               | Value       | Usage (MB)  | Value       | Usage (MB)        | Value  |
| No Virtualization | 1273.38    | 2841        | 32714.08    | 2539              | 38477.6 | 2507      | 0.04%       | 2812132.28   | 28.58   | 0.06%       | 1135.37     | 2842   | 0.05%       |
| Docker        | 2071.79     | 2791        | 32233.27    | 2902              | 51522.74 | 2791      | 0.05%       | 2229713.99   | 2902   | 0.05%       | 1581.11     | 2902   | 0.05%       |
| VMware        | 1180.92     | 1337        | 12860.3     | 1350              | 13167.09 | 13167.09  | 1.10%       | 482.88       | 1540   | 3.90%       |
| KVM           | 1406.08     | 451         | 12905.73    | 434               | 13317.28 | 440       | 4.50%       | 988668.05    | 440    | 4.50%       | 491.74      | 440    | 4.50%       |

Figure 7: Compilebench results. "Compile" indicates the Compile test. "Create" refers to the Initial Create test. "Read" specifies the Read Compiled Tree test.

4.3.2 Processor Performance. In this section, we discuss experiments related to comparing processor performance. The main goal is to assess how fast the compute-intensive applications will run under the four test environments (native, Docker, VMware, and KVM). We start with the pts/aobench benchmark. The results are presented in Figure 8. Here again, Docker achieves the best performance as the benchmark execution completes within 51 seconds. The native execution, VMware, and KVM respectively terminate after 72 seconds, 74 seconds, and 73 seconds. The experiment also reveals that VMs perform like native execution. The similar performance observed between the native execution, VMware and KVM are expected as compute-intensive workloads typically use less privileged instructions than IO-intensive applications. VMs running on modern VMMs directly execute on the underlying CPU. The VMM is only invoked when a VM issues privileged instructions.

When we run the pts/asmfish benchmark the native execution comes at the top followed by Docker. As opposed to the results observed with pts/aobench, the VMware and KVM performances are quite similar but far lower than that of Docker and the native execution. After investigation, it appears that the performance of pts/aobench depends on the available processing cores [17]. The benchmark description also highlights the fact that there is a significant performance improvement when increasing the number of AMD processors instead of Intel processors. In our testing setup, the native execution and Docker ran with 16 AMD cores while KVM and VMware only used only four cores, justifying the performance difference.

Overall, the lesson here is similar to what was observed with IO-intensive benchmarks. The native execution and Docker generally come at the top, and KVM and VMware have lower performance, but remain close in execution profile.

4.3.3 System Performance. Finally, we discuss experiments related to comparing system-level performances. We start by running pts/phpbench. It runs tests that assess several aspects of the workloads typically observed in PHP servers. Figure 10 summarizes the finding. The native execution, VMware and KVM have similar score while Docker performs way better. Table 2 summarizes execution results from running pts/stress-ng. Overall the results show that once more the native execution and Docker have similar performances that are above that of VMware and KVM. Nevertheless, KVM and VMware seems to be less memory hungry.
As a summary, throughout the experiments that we carried out, we observed that Docker and the native execution performed better on compute-intensive, I/O-intensive, and system benchmarks. In general, VMware and KVM had lower performance regardless of the stressor we used. The fact is VMs are more suited for deploying systems as they can run OSes, nested VMs, and even containers. However, they incur significant performance degradation compared to containers and bare-metal systems. Containers showed near-native performance as they are similar to the processes that execute directly above the host OS but are mostly limited to running specific applications.

5 CONCLUSION

We studied three types of virtualization environments in this work: containers, Type-1, and Type-2 virtual machine monitors. We evaluated the performance achieved by applications running in Docker, KVM, and VMware Workstation. One of the main observations is that containers appear to be the best platform to run applications if the target is fast execution with low overhead. However, they are not suited for deploying complete systems as they can only run applications. We also observed that VMware and KVM tend to have similar execution performances with minor differences. Future work will extend the study to other virtualization systems such as Xen and explore the performance achieved when virtualization technologies are nested.

REFERENCES

[1] Edouard Bugnion, Scott Devine, Mendel Rosenblum, Jeremy Sugerman, and Edward Y. Wang. 2012. Bringing virtualization to the x86 architecture with the original vmware workstation. ACM Transactions on Computer Systems (TOCS) 30, 4 (2012), 1–51.
[2] Edouard Bugnion, Jason Nieh, and Dan Tsafrir. 2017. Hardware and software support for virtualization. Synthesis Lectures on Computer Architecture 12, 1 (2017), 1–206.
[3] Humbles Devassy Chirammal, Prasad Mukhedkar, and Amil Vettathu. 2016. Mastering KVM virtualization. Packt Publishing Ltd.
[4] Susanta Nanda Tzi-cker Chiueh and Stony Brook. 2005. A survey on virtualization technologies. Rpe Report 142 (2005).
[5] Theo Combe, Antony Martin, and Roberto Di Pietro. 2016. To docker or not to docker: A security perspective. IEEE Cloud Computing 3, 5 (2016), 54–62.
[6] Michael J. De Lucia. 2017. A survey on security isolation of virtualization, containers, and unikernels. Technical Report. US Army Research Laboratory Aberdeen Proving Ground United States.
[7] Peter J Denning. 2001. Anecdotes [virtual machines]. IEEE Annals of the History of Computing 23, 3 (2001), 73.
[8] Ankit Desai, Rachana Oza, Pratik Sharma, and Bhaai Patik Patel. 2013. Hypervisor: A survey on concepts and taxonomy. International Journal of Innovative Technology and Exploring Engineering 2, 3 (2013), 222–225.
[9] US Air Force. 2000. Analysis of the Intel Pentium’s ability to support a secure virtual machine monitor. In Proceedings of the... USENIX Security Symposium. USENIX Association. 129.
[10] Cheol-Ho Hong, Ivor Spence, and Dimitrios S Nikolopoulos. 2017. FairGV: fair and fast GPU virtualization. IEEE Transactions on Parallel and Distributed Systems 28, 12 (2017), 3472–3485.
[11] Cheol-Ho Hong, Ivor Spence, and Dimitrios S Nikolopoulos. 2017. GPU virtualization and scheduling methods: A comprehensive survey. ACM Computing Surveys (CSUR) 50, 3 (2017), 1–37.
[12] IBM. 2020. 7090 Data Processing System. Retrieved on March 2021 https://www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP7090.html.
[13] Beng-Hong Lim. 2001. Virtualizing the FC platform. Retrieved on March 2021 https://www.usenix.org/legacy/publications/library/proceedings/usenix01/sugerman/sugerman.html/node2.html.
[14] Joel Mbongue, Festus Hategekimana, Danielle Tchinkou Kwadjo, David Andrews, and Christophe Bobda. 2018. FPGAVir: A Novel Virtualization Framework for FPGAs in the Cloud. In 2018 IEEE 11th International Conference on Cloud Computing (CLOUD). IEEE, 862–865.
[15] Joel Mandebi Mbongue, Festus Hategekimana, Danielle Tchinkou Kwadjo, and Christophe Bobda. 2018. FPGA Virtualization in Cloud-Based Infrastructures Over Virtio. In 2018 IEEE 36th International Conference on Computer Design (ICCD). IEEE, 242–249.
[16] Joel Mandebi Mbongue, Alex Shaping, Pankaj Bhowmik, and Christophe Bobda. 2020. Architecture Support for FPGA Multi-tenancy in the Cloud. In 2020 IEEE 31st International Conference on Application-Specific Systems, Architectures and Processors (ASAP). IEEE, 125–132.
[17] openbenchmarking. 2020. asmfish. Retrieved on March 2021 https://openbenchmarking.org/test/pts/asmfish-1.1.1.
[18] Phoronix. 2020. Phoronix Test Suite Download. Retrieved on March 2021 https://www.phoronix-test-suite.com/?k-downloads.
[19] RedHat. 2020. What is KVM? https://www.redhat.com/en/topics/virtualization/what-is-KVM.
[20] Rusty Russell. 2008. virtio: towards a de-facto standard for virtual I/O devices. ACM SIGOPS Operating Systems Review 42, 5 (2008), 95–103.
[21] Rich Uhlig, Gil Neiger, Dion Rodgers, Amy L. Santoni, Fernando CM Martins, Andrew V. Anderson, Steven M. Bennett, Alain Kagi, Felix H. Leung, and Larry Smith. 2005. Intel virtualization technology. Computer 38, 5 (2005), 48–56.
[22] Chris Weber, Azhar Saiyed, Maaz Kamani, and Pirasanth Sivalingam. 2013. A scientific review on virtual machine performance. Faculty of Business and IT 2, 2335 (2013), 2–466.
[23] Ian Webster. 2020. Inflation Calculator. Retrieved on March 2021 https://www.in2013dollars.com/us/inflation/1960.