A Survey of Virtualization Technologies: Towards a New Taxonomic Proposal

Una revisión de las tecnologías de virtualización: hacia una nueva propuesta taxonómica

Luis E. Sepúlveda-Rodríguez1, Julio C. Chavarro-Porras2, John A. Sanabria-Ordóñez3, Harold E. Castro4, and Jeanna Matthews5

ABSTRACT

At present, there is a proliferation of virtualization technologies (VTs), which are part of the basic and underlying infrastructure of popular cloud computing. Those interested in VTs are faced with a non-unified volume of information and various approaches to modes of operation, classification structures, and the performance implications of these technologies. This makes it difficult to decide which type of VT is appropriate for a particular context. Therefore, this paper reviews the state of the art on VT taxonomic models. Methodologically, a literature review is carried out to identify VT classification models, recognizing their features and weaknesses. With this in mind, a new taxonomy of virtualization technologies is proposed, which responds to the weaknesses identified in the analyzed schemes. The new VT taxonomy combines the Abstraction Level and Virtual Machine Type approaches, providing the reader with a means to visualize VTs. In doing so, the reader can locate the level of abstraction at which each VT is developed, in addition to the type of machine projected, whether it is a complete system or an execution environment for processes. The proposed taxonomy can be used in the academic environment to facilitate teaching processes or in the business environment to facilitate decision-making when implementing VTs.

Keywords: container, taxonomy, virtualization, virtual machine, virtualization technologies

RESUMEN

En la actualidad existe una proliferación de tecnologías de virtualización (VTs), las cuales constituyen una parte de la infraestructura fundamental y subyacente al tan popular cloud computing. Los interesados en las VTs se enfrentan a un volumen de información no unificada y con enfoques diversos acerca de los modos de operación, estructuras de clasificación e implicaciones del desempeño de estas tecnologías. Esto hace difícil decidir sobre el tipo de VT adecuado para un contexto particular. Por lo anterior, este trabajo realiza una revisión del estado del arte acerca de los modelos taxonómicos de las VTs. Metodológicamente, se realiza una revisión de la literatura para identificar modelos de clasificación de las VTs, reconociendo sus características y debilidades. Considerando lo anterior, se propone una nueva taxonomía de las tecnologías de virtualización, que responde a las debilidades identificadas en los esquemas analizados. La nueva taxonomía de VTs combina los enfoques de Nivel de Abstracción y Tipo de Máquina Virtual, proporcionando al lector un medio para visualizar las VTs. Al hacerlo, el lector puede ubicar el nivel de abstracción en el que se desarrolla cada VT, además del tipo de máquina proyectada, ya sea un sistema completo o un entorno de ejecución para procesos. La taxonomía propuesta puede ser utilizada en el ámbito académico para facilitar los procesos de enseñanza o en el ámbito empresarial para favorecer la toma de decisiones a la hora de implementar VTs.

Palabras clave: contenedor, máquina virtual, taxonomía, tecnologías de virtualización

Received: July 20th, 2021
Accepted: June 9th, 2022

Introduction

In recent years Virtualization Technology (VT) has been used to obtain benefits such as isolation, resource splitting, consolidation, security, migration, and ease of management (Varasteh and Goudarzi, 2017). VT builds an abstraction of applications and hardware in a virtual view (AbdElRahem et al., 2016). This virtual view can be different from the physical

view of computing resources (Stallings, 2015). In addition, Silberschatz et al. (2014) note that VT allows an operating systems (OS) to run as an application within another OS.

VT includes emulation, which refers to the fact that there are differences between the physical and logical architectures used by virtualized processes. Thus, a virtual machine (VM) could use the same host architecture, a different emulated architecture, or a hybrid. In addition, the processes could use a physical architecture with modifications in order to make virtualization easier (paravirtualization).

VT allows creating one or several environments, i.e., many computers can look like a single large resource (resource aggregation) or, conversely, a single computer is considered as several instances of itself (resource splitting) (Hoopes, 2009; Silberschatz et al., 2014).

Unfortunately, the x86 computer architecture, despite being one of the most widely adopted architectures in the world, cannot be completely virtualized (Adams and Agesen, 2006). However, this situation can be solved through mechanisms and VT approaches that act at different levels of abstraction. The abstraction levels where VT takes place are the instruction set level, the hardware abstraction level (HAL), the OS level, the user library interface level, and the application level (Nanda and Chiueh, 2005).

The concept of virtualization was formalized in Goldberg’s thesis (1973) and published in other works (Goldberg, 1974; Popek and Goldberg, 1974). In these studies, VMs were defined as “an efficient and isolated duplicate of the real machine” (Goldberg, 1973, p. 12). In later works, the term VM was expanded to include other kinds of virtualization, including applications at user level such as libraries, system calls, interfaces/services, system configurations, processes, and state files (Nanda and Chiueh, 2005).

The term virtual machine monitor (VMM) was also established by Popek and Goldberg (1974). It is a software layer that supports infrastructure using the resources of a lower level to create multiple independent and isolated VMs (Cafaro and Aloisio, 2011; Nanda and Chiueh, 2005). Similarly, Stallings (2015) determined that a VMM acts as an intermediary between the real machine and VMs. VMMs are also called hypervisors (Hoopes, 2009).

VT also brings financial benefits regarding returns on investment and reductions in the total cost of ownership of computer systems hardware (AbdElRahem et al., 2016). Moreover, VT uses less energy, which is related to the so-called green computing (Jing et al., 2013; Ranjith et al., 2017; Thathera et al., 2015) and plays an essential role in safeguarding the environment. Other goals of VT include increasing the scalability and availability of computing environments, as well as improving the administrative and security structures of the existing computational infrastructure (Hui and Seok, 2014; Kusnetzky, 2011).

Kampert (2010) indicates that the benefits of VTs have revolutionized data centers in the last two decades and have motivated the development of many variations to suit different use cases. In response, several attempts have been made in the academic literature to establish classification schemes for these variations of VT.

This paper reviews VT classification schemes and proposes a new taxonomy that responds to several identified weaknesses. This taxonomy improves and unifies the previous works in the classification of VTs in three ways: first, it combines and unifies approaches that consider the VM type with those that consider the level of abstraction; second, it updates classification approaches to include examples of VTs that have emerged more recently; third, it introduces a taxonomic key diagram based on our unified classification, which can guide the selection of VTs in either academic or production environments.

The remainder of this document comprises the following sections: VT classification schemes, The need for a new taxonomy, Proposal for a virtual machine taxonomy, Taxonomic key diagram, and Conclusions.

**VT classification schemes**

This section presents the results of a literature review by means of a systematic process of combined database search and manual reference tracking using the Snowball technique (Samireh and Claes, 2012). In this way, 12 classification schemes for VTs were identified, and their characteristics were highlighted. A paragraph is added at the end of each case which highlights the strengths and weaknesses of the classification scheme.

**VT taxonomy by Nanda and Chiueh**

Nanda and Chiueh (2005) classified VTs according to the following five levels of abstraction of a computer system.

**Instruction set architecture (ISA) level**

VTs emulate an ISA, allowing VMs to run as if they were running on hardware. When the ISA offered by this layer differs from the real ISA, this is called emulation.

- **Hardware Abstraction Layer (HAL):** VTs use the same ISA as the host. Here, it is possible to perform independent OS installations, and its applications run as if they were executed in a real environment.

- **Operating System:** VTs work through an OS module to provide a virtualized system call interface.

- **Library Level:** User-level libraries control the communication between the applications and the rest of the system. VTs allow implementation as an Application Binary Interface (ABI) or an Application Programming Interface (API).
• **Programming Language Level:** VTs implement the virtualization layer as an application that can create a simplex or complex VM.

Although Nanda and Chiueh (2005) establish a way to classify VTs, they do not consider virtualization types at the same level of abstraction. Besides, it is necessary to include some VTs that have emerged in recent years.

**VM taxonomy by Smith and Nair**

![Figure 1. VM taxonomy proposed by Smith and Nair in 2005](source: Smith and Nair (2005))

Smith and Nair (2005) presented a taxonomy with two main categories: Process VMs and System VMs. Furthermore, these categories divide VTs according to whether the ISA supported in the VM is the same as the underlying hardware (Figure 1).

**Process VMs**

This category describes an environment in the ABI interface or at the API level. It is called a *Multiprogrammed System* when it uses the same ISA; otherwise, it is called *Dynamic Emulator* or *Binary Translator*. The subcategories are described below:

- **Multiprogrammed Systems** are multiprogramming OSs that implement the management of timeshare access to the available underlying hardware resources. These systems use the same ISA and can handle multiple user processes ‘simultaneously’. The OS delivers an individual VM for each user process that runs concurrently. One implementation in this context involves dynamic binary optimizers using the same ISA from the host system.

- **Dynamic Emulators** use process VMs to support compiled binary programs for an ISA different from the underlying hardware. This condition implies executing an emulation effort performed through interpretation, which can be relatively slow. However, this situation can be compensated when a software cache is implemented in order to deal with the overload.

**System VMs**

These are characterized by hosting one or several complete and independent OSs running simultaneously on the same hardware of the host computer, which results from the intermediation performed by the VMM. The subcategories of the system VMs are described below:

- **Classic System VMs** use the VMM and execute it directly on the bare hardware without an underlying OS. Thus, the VMM has real access to hardware resources and serves as an intermediary between the guest OSs and the hardware itself. In this case, the VMs are called *Hosted VMs*.

- **Whole-system VMs** provide virtualization of a complete environment, but guest systems use an ISA different from those used in the underlying hardware, unlike the previous category. In this case, the VMs are called *Codesigned VMs*.

Smith and Nair’s study (2005) can be considered an essential basis for classifying VTs that provide a virtual environment for a complete system or processes. However, this work does not contemplate what was established by Nanda and Chiueh (2005) regarding the levels of abstraction. Another important aspect is that this classification model does not have a high degree of detail; it uses very general descriptions, without even including specific technologies. It is essential to consider that this study was carried out in 2005 and does not include subsequently developed technologies.

**Virtualization taxonomy by the SCOPE Alliance**

The SCOPE Alliance (2008) proposed an extension of the work carried out by Smith and Nair in 2005. The proposal includes more branching of the main categories and more examples of VTs (Figure 2).

This classification places type I and type II hypervisors as distinctions of the Classic OS VM model of System VMs that support the same ISA as the underlying hardware.
Regarding the Process VMs category, this classification distinguishes between a Multiprogrammed System and Dynamic Translators. Multiprogrammed Systems are further classified depending on whether the OS provided by the underlying system is the same as the OS used by the application. If it uses the same OS, the category is called *Multitask OS*, which contains OS Virtualization. If the OS is different, it is called *OS Translator*. When the processes are based on a different ISA, they are called *ISA & ABI Translators*; otherwise, they are called *High-level Language*.

Although the SCOPE Alliance's study (2008) contributes to complementing the taxonomy of VTs, the research does not contemplate aspects such as the levels of abstraction indicated by Nanda and Chiueh (2005). This situation gives rise to problems of conceptual inference, in which, for example, type I and type II hypervisors are perceived to be at the same level of abstraction. Additionally, according to the date of publication of the study, it is necessary to expand concepts and update VTs that have emerged in recent years.

### Taxonomy of VTs by Kampert

Kampert (2010) presented his taxonomy of VTs using different virtualization techniques. This taxonomy uses the unified modeling language, as shown in Figure 3, where all elements are classes. For example, the class Domain is a superclass of the classes Server, Application, Desktop, Storage, and Network.

Kampert’s taxonomy (2010) aims to cover the domains in a complete way in which the concept of virtualization takes place, including storage and network virtualization not seen in previous taxonomies. However, this taxonomy itself does not offer the level of granularity necessary to identify VTs in each of the specified domains.

![Figure 3. VT taxonomy by Paul Kampert](source: Kampert (2010))

### Virtualization model by Kusnetzky

Kusnetzky’s virtualization model (2011) is composed of seven parts, five distributed in layers, and two arranged parallel to the layers above Kampert (2010). Each part is briefly described below:

- **Access virtualization**: Many users share the same system.
- **Application virtualization**: Many applications run transparently on different OSs and hardware platforms.
- **Processing virtualization**: Allows the division or aggregation of resources.
- **Network virtualization**: Presents a logical view of the physical network elements.
- **Storage virtualization**: Hides the location and type of physical storage devices in which applications store their data.
- **Security for virtual environment**: Controls the access to the various elements of virtual media in order to protect them from unauthorized actions.
- **Management of the virtual environment**: Controls the available physical resources and the generated virtual environments.

![Figure 4. Kusnetzky's model of virtualization](source: Kusnetzky (2011))

Kusnetzky presents a way to include categories for a range of virtualizable computational resources but does not provide details about the existing VTs in each layer of the model. In addition, the model does not differentiate between technologies of the same layer. For example, in Processing Virtualization, there is no evidence of a difference between the types of VMs present in type I or type II hypervisors.

### Taxonomy of VTs by Pessolani

Pessolani et al. (2012) proposed their taxonomy of VTs with five main categories: 1) Hardware or System Virtualization, 2) Para-virtualization, 3) Virtualization based on OS, 4)
Virtualization at the Process or Application level, and 5) Virtualization of OS. Additionally, the main categories include subcategories that suggest a level of abstraction (Figure 5). These main categories are described below:

- **Hardware or System Virtualization** puts the type I hypervisor on top of the hardware with its VMs and their respective guest OSs.

- **Paravirtualization** distributes its elements to Hardware or System Virtualization, but the guest OS is modified to be aware that it is virtualized.

- **Virtualization based on OS** is founded on using independent workspaces called containers, which are based on the host OS.

- **Virtualization at the Process or Application level** uses an application on the host OS to provide a VM that allows the execution of processes based on it.

- **Virtualization of OS** needs a host OS to carry out the functions of a hypervisor in order to support the guest OSs, which in turn have their own completely independent applications.

Pessolani’s taxonomy (2012) does not explicitly consider the levels of abstraction to which these technologies apply. In addition, it focuses only on the conceptual elements, leaving specific examples aside, nor does it establish a way to divide types of VMs within each main category.

**Taxonomy of virtualization concepts by Pék**

Pék et al. published a taxonomy of virtualization concepts in 2013. This work extends the studies by Smith and Nair (2005) and the SCOPE Alliance (2008) (Figure 6).

This taxonomy adds elements and several components, such as in the Hosted category, equivalent to type II hypervisors from the study by the SCOPE Alliance (2008). It also includes the Paravirtualization subcategory.

**Taxonomy of virtualization by Ameen**

Ameen and Hamo presented a taxonomy with three levels in 2013. (Figure 7). The first level contains the following ten categories:

- **Mobile** software that is embedded on a mobile phone to decouple the applications and data from the underlying hardware (VMware, 2022).

- **Data** abstracts the source of individual data items and provides a common data access layer for different data access methods such as SQL, XML, JDBC, File access, MQ, JMS, etc. (Mann, 2006).

- **Memory** adds an extra level of address translation to give each VM the illusion of having zero memory address space, as real hardware provides (Waldspurger, 2002).
A Survey of Virtualization Technologies: Towards a New Taxonomic Proposal

- **Desktop** is the ability to display a graphical desktop from one computer system on another computer (von Hagen, 2008).

- **Storage** creates logical abstractions of physical storage systems (B. Li et al., 2005).

- **Server** is a type of virtualization that allows running many OSs both in isolation and independence.

- **Network** provides an abstraction layer that can decouple the physical network equipment from the delivered business services over the network (Annapareddy, 2011).

- **Application** allows the user to run the application using local resources without installing the application in his system completely (Annapareddy, 2011; White and Pilbeam, 2010).

- **Grid** provides a way to abstract multiple physical servers from the application they are running (Mann, 2006).

- **Clustering** causes several locally connected physical systems to appear to the application and end-users as a single processing resource (Mann, 2006).

The following describes the virtualization types at the second level of the taxonomy, which are derived from the Server category, as indicated by Ameen and Hamo (2013):

- **Emulation** is a virtualization method in which you can create a complete hardware architecture in software (Ameen and Hamo, 2013).

- **Hosted OS** uses software-only. The hypervisor is over an OS (Ameen and Hamo, 2013; von Hagen, 2008).

- **Hardware** the hypervisor is assisted by processor hardware such as AMD-V or Intel VT-x processor virtualization technologies (von Hagen, 2008).

- **Paravirtualization**, according to Ameen and Hamo (2013, p. 7), is “a technique in which the guest OS includes modified (para-virtualized) I/O drivers for the hardware”.

- **Container** is a kernel-layer abstraction and refers to techniques in which the abstraction technology is built directly into the OS kernel rather than having a separate hypervisor layer (Ameen and Hamo, 2013; Q. Lin et al., 2012).

- **Hybrid** is a combination of Full Virtualization and Paravirtualization that uses input/output (I/O) acceleration techniques (White and Pilbeam, 2010).

At the third level of the taxonomy are the type I and type II hypervisor categories derived from Server/Hardware.

Ameen and Hamo’s taxonomy (2013) is closely related to the works by Kampert (2010) and Kusnetzky (2011). Furthermore, it presents a classification scheme through a three-level hierarchical structure. However, although this graphical representation is interesting, it is unbalanced, since it focuses only on detailing the Server category.

**Taxonomy of VTs by Abdulhamid**

Abdulhamid et al. (2014) presented a taxonomy focused on cloud computing (Abdekhoda et al., 2019; Fareghzadeh et al., 2019) and based on the work by Sahoo et al. (2010), which includes categories such as Full Virtualization, OS-Layer Virtualization, Hardware-Layer Virtualization, Paravirtualization, Application Virtualization, Resource virtualization, and Storage virtualization. In addition, this work adds the Grid Virtualization and Cloud Virtualization categories (Figure 8).

![Figure 8. Taxonomy of VTs by Abdulhamid](source)

Some categories have already been described. Below is a brief description of the new ones.

- **Grid Virtualization** focuses on the virtualization of grid resources either for a virtual organization (VO) or for a Virtual Organization Cluster (Abdulhamid et al., 2014).

- **Cloud Virtualization** or **Cloud Computing** (Sehgal and Bhatt, 2018) enables on-demand provisioning of virtual resources through the Web, as well as applying the concept of pay-per-use. In this category, the VTs form cloud computing services, provisioning virtual resources to customers on demand (Abdulhamid et al., 2014; Aceto et al., 2013).

Although the taxonomy by Abdulhamid et al. (2014) shows two levels, only one level can be observed which comprises its nine categories from a hierarchical perspective. On the other hand, the description of each category lacks details and examples of VTs.

**Types of VMs by Li**

X.-F. Li (2016) presented his work with four types of VMs:

- **Type 1**: The Full ISA VM allows full ISA-level emulation or virtualization. The OS and its applications can run on top of the VM as a real machine (X.-F. Li, 2016).
- **Type 2**: The ABI VM allows ABI-level emulation of the processes in the guest OS. These applications can run in conjunction with native ABI applications (X.-F. Li, 2016).

- **Type 3**: The Virtual ISA VM provides a runtime engine for applications encoded in the virtual ISA to run on it (X.-F. Li, 2016).

- **Type 4**: The Language VM gives a runtime engine that runs programs written in a guest language (source). The runtime engine needs to interpret or translate the program.

Although the study by X.-F. Li (2016) presented a four-type classification scheme, it does not indicate a hierarchical structure that clarifies how they relate. It also does not have a supporting graph to facilitate understanding. This work does not contemplate many of the categories indicated in other previously presented taxonomies.

**Taxonomy of VMs by Bugnion**

Bugnion (2017) presented a structure with two levels that shows the concepts related to VMs. The first level is related to abstraction, and it includes the following categories: Language-based VM, System-level VM, and Lightweight VM. The second level is related to the platform, and it includes two categories derived from System-level VM, which are called Machine Simulator and Hypervisor. The latter is divided into Bare-metal Hypervisor (type I) and Hosted Hypervisor (type II) (Figure 9).

![Diagram of VM taxonomies](source)

**Figure 9. Taxonomy of VMs and the platforms that run them presented by Bugnion**

Source: Bugnion et al. (2017)

- **Language-based VM** refers to any managed language runtime environment such as the Java VM, Microsoft Common Language Runtime, and JavaScript engines embedded in browsers.

- **Lightweight VM** refers to software mechanisms to ensure that applications run directly on the processor as securely isolated from other environments and the underlying OS.

- **System-level VM** refers to the computer environment that resembles the hardware of a computer, so that the VM can run an OS and its applications in complete isolation from the other VMs and the rest of the environment. This category includes two Hypervisor types (type I and type II).

Bugnion’s work (2017) is less a taxonomy than a book focusing on the core architectural support provided by hardware to run VMs efficiently.

**The need for a new taxonomy**

The taxonomies described above have many elements that contribute to the classification of VTs. However, in each of these classification schemes, some aspects that need improvement have been identified. Each scheme offers a taxonomic approach, such as a) Abstraction Level, b) Type of VM, and c) Virtualization Domains. Table 1 summarizes the classification schemes analyzed in this paper by author, year, and taxonomic approach, which were published between 2005 and 2017. It is worth noting that we found no taxonomies published between 2018 and 2021.

The Type of VM is the most popular approach, as demonstrated by CS2, CS3, CS6, and ACS7. On the other hand, CS4 and CS5 take a different perspective; their objective is to consider, in a general way, the largest number of technological domains in which it is possible to carry out virtualization processes, hence the name Virtualization Domain. Some taxonomies have a dual approach; for example, CS8 and CS9 combine the Type of VM with the Virtualization Domain, and CS11 combines the Type of VM with the Abstraction Level. Lastly, CS1 and CS10 consider the Abstraction Level approach as fundamental for the categorization of VTs. These differences in viewing VTs can confuse the community interested in this field when reading different authors.

![Table of classification schemes](source)

**Table 1. Summary of classification schemes**

<table>
<thead>
<tr>
<th>ID</th>
<th>Author(s)</th>
<th>Taxonomy approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS1</td>
<td>Nanda and Chiueh (2005)</td>
<td>Abstraction Level</td>
</tr>
<tr>
<td>CS2</td>
<td>Smith and Nair (2005)</td>
<td>Type of VM</td>
</tr>
<tr>
<td>CS3</td>
<td>SCOPE Alliance (2008)</td>
<td>Type of VM</td>
</tr>
<tr>
<td>CS4</td>
<td>Kampert (2010)</td>
<td>Virtualization Domain</td>
</tr>
<tr>
<td>CS5</td>
<td>Kusnetzky (2011)</td>
<td>Virtualization Domain</td>
</tr>
<tr>
<td>CS6</td>
<td>Pessolani et al. (2012)</td>
<td>Type of VM</td>
</tr>
<tr>
<td>CS7</td>
<td>Pék et al. (2013)</td>
<td>Type of VM</td>
</tr>
<tr>
<td>CS8</td>
<td>Ameen and Hamo (2013)</td>
<td>Type of VM and Virtualization Domain</td>
</tr>
<tr>
<td>CS9</td>
<td>Abdulhamid et al. (2014)</td>
<td>Type of VM and Virtualization Domain</td>
</tr>
<tr>
<td>CS11</td>
<td>X.-F. Li (2016)</td>
<td>Abstraction Level</td>
</tr>
<tr>
<td>CS12</td>
<td>Bugnion (2017)</td>
<td>Type of VM and Abstraction Level</td>
</tr>
</tbody>
</table>

Source: Authors
Therefore, there is a need for a new taxonomy that provides a unified, organized, and current view of VTs. Therefore, this paper makes the following contributions:

- A review of the literature with the identification, analysis, and comparison of 12 classification VT schemes (Table 1).
- A proposal for a new VM taxonomy. This work identified, expanded, and combined different studies, offering a single view of multiple concepts such as the Types of VMs and their corresponding Abstraction Level. The taxonomy includes examples of older VTs in order to provide a reference factor to those who have some knowledge about them. It also includes examples of new VTs that have gained wide recognition in the industry and academia, such as those related to containers. The taxonomy is also intended to be an instrument to support the pedagogical processes within the academic community with interests in VTs (Figure 10).
- A taxonomic key diagram that facilitates the visualization of the technological ecosystem that surrounds this topic and consequently helps the academic and industrial community in the decision-making processes regarding the selection of VTs (Figure 11).

Proposal for a virtual machine taxonomy

This section presents a new taxonomic proposal for virtualization technologies. This taxonomy considers the 12 studies reviewed in this research, but it focuses mainly on studies such as CS1, CS2, CS3, CS7, and CS11 (Figure 10). The proposal presents an innovative contribution that integrates the Abstraction Level and Type of VM taxonomic approaches. In addition, it contributes by extending the examples of VTs, which are placed in the diagram representing the new taxonomy. The first approach considers the layers of the classical architecture of a computer system and makes it possible to visualize the VTs according to the level of abstraction they occupy at the time of execution. The second approach considers the types of virtual machines, be it complete systems or execution environments for processes. The description of the taxonomy is shown below, making a cross-analysis between the two approaches.

Approach 1: abstraction layers

The first approach of this taxonomy uses the abstraction layers in a computer system, such as the Hardware Abstraction Layer (HAL), the Operating System (OS), the Application Binary Interface (ABI), the Application Programming Interface (API), Type I/Type II Hypervisors, and Libraries. In Figure 10, the labels located on the left side indicate the abstraction level, and they are the title of rectangular structures with horizontal distribution in the taxonomy. With these layers, the taxonomy makes it possible to locate VTs depending on the level at which they take place. Thus, the reader can quickly infer aspects such as the dependence or not of an underlying OS, as well as determine the number of intermediaries involved in the virtualization process. Furthermore, this information allows inferring the possible performance of these technologies. The abstraction layers are described below from bottom to top.

Hardware Abstraction Layer (HAL)

HAL includes those VTs that are placed directly on top of the hardware. This arrangement is also known as BareMetal and is identified by the absence of intermediaries between the VMs and the underlying hardware, suggesting a higher performance for the set of VTs placed here. This layer contains the category called Type I Hypervisor and can have several types of VTs.

Operating System (OS)

This layer contains the sublayers Application Binary Interface (ABI) and Application Programming Interface (API). In the ABI sublayer, the VTs use the OS as an intermediary to access the underlying hardware. The virtualization is carried by OS calls and uses Dynamic Binary Translation, Type II Hypervisors, or Libraries. This situation suggests that the VTs may present degradation in performance due to intermediation costs between the different environments. VTs implement virtualization based on high-level languages, offering portability in the API sublayer, as APIs support multiple hardware and software platforms. However, this sublayer has considerable degradation given the multiple interpreters between the VTs and the hardware functions.

Approach 2: Type of VM

The second approach of this taxonomy considers VTs according to their type: System VMs or Process VMs. System VMs contain a whole OS (guest OS) within their virtual environment. On the other hand, Process VMs use the host OS as an intermediary between the virtual environment and the actual hardware.

System VMs

This type of virtualization has two categories. The first is Classic System VMs and is characterized by the fact that the host and guest OSs have the same ISA. The second category is Whole-System VMs and is characterized by the host OS and guest OS having a different ISA.

- **Classic System VMs**: This category is known as Hardware Virtualization and includes two subcategories: the first one is Native VMs, and the second one is Hosted VMs. It is important to note that each subcategory takes place at different levels of abstraction.

- **Native VMs**: This VT is also known as Type I Hypervisor and corresponds to the HAL abstraction.
level. It uses a software layer directly on top of the hardware. It also presents a subdivision, as shown below:

- **Transparent** indicates that the OS inside the VM is unaware of its virtualization state and is divided into the following types:
  
  - **Hardware-Assisted** virtualization involves the use of physical components to facilitate the management of VMs. Examples of this are: KVM (2021), Microsoft Hyper-V (Jason et al., 2009; Syrewicze and Siddaway, 2018), Xen (Xen Cambridge, 2022), VLX (Armand and Gien, 2009), and VMware ESX/ESXi (Z. Li, 2021; VMware, 2022).
  
  - **Dynamic Binary Translation** implies that the Type I Hypervisor catches and inspects the code of each guest OS request to convert it into a proper request towards the underlying hardware, e.g., VMware ESX/ESXi (Z. Li, 2021; VMware, 2022) and XtratuM (Wessman et al., 2021; Xtratum, 2022).
  
  - **Para-virtualized** is also known as Operating System-Assisted Virtualization and refers to efficient communication between the guest OS and the hypervisor. This implies modifying the guest OS to be aware of virtualization and to take advantage of that condition. Examples of this are: Xen (Barham et al., 2003; Matthews et al., 2008; Xen Cambridge, 2022; Xen Project, 2022), VLX (Armand and Gien, 2009), KVM (Abeni and Faggioli, 2020; KVM, 2021), and VMware VMI (VMware, 2022).
  
  - **Hosted VMs**: This subcategory is also known as Type 2 Hypervisors, corresponds to the ABI abstraction level, and uses a layer of software on a Host OS. It presents the same subdivision and functions of the **Native VMs** category, so only examples of VTs will be listed below.

  - **Transparent**: VMware Workstation/Fusion (VMware, 2022), Parallels Desktop (Parallels, 2021), and Oracle VirtualBox (Oracle, 2021b).
  
  - **Dynamic Binary Translation**: VMware Workstation/Fusion (VMware, 2022; Z. Li, 2021), Microsoft Virtual PC (Honeycutt, 2003), Plex86 (2021), Parallels Desktop (Parallels, 2021), and Oracle VirtualBox (Oracle, 2021b).

  - **Para-virtualized**: VMware Workstation, with the addition of the corresponding para-virtualization driver to the network in the guest OS (El-Anani, 2021; VMware, 2022).

  - **Whole system VMs**: This category is called Hardware Emulation and presents an OS different from the underlying hardware. It takes place at the API abstraction level, evidencing a preexisting OS on which emulation can occur. The subcategory is called Dynamic Binary Translation and features VTs such as QEMU (Díaz et al., 2021; QEMU, 2021), Simcs (Magnusson et al., 2002), Bochs (Bochs, 2021), Rosetta (Apple Inc, 2009), and BIRD (Nanda et al., 2006).

**Process VMs**

This type of virtualization also has the same two categories as System VMs, depending on whether the host OS and guest OS have the same ISA. When the ISA is the same, the category is called Multiprogrammed Systems; otherwise, the category is called Dynamic Translators. Both categories are located at the OS layer.

- **Multi-programmed systems**: In this category, the VTs share the OS among many processes, generating independent execution spaces for each one. This generates the illusion that, for a moment, a process is an exclusive executor in the system. This category is then divided into two, depending on whether there is an OS. When the same OS is projecting, the category is called Multitasking OS; otherwise, it is called OS Translators.

- **Multitasking OS** is divided into Operating System Virtualization and Same-ISA Dynamic Binary Optimizer.

- **Operating System Virtualization** happens at the ABI abstraction level and uses system calls for interaction with the underlying hardware. It uses the preexisting OS, and it allows generating independent workspaces for the processes. This type of virtualization is booming and is often known as lightweight virtualization, container-based, or simply containers (Tirifonov, 2018). For example: FreeBSD Jails (Biederman, 2006; Kamp and Watson, 2000) (Ryding and Johansson, 2020), Solaris Zones/Containers (Oracle, 2021a), OpenVZ (2021), Linux-VServer (Linux-VServer, 2018), AIX Workload Partitions WPAR (Gibson, 2007), Parallels Virtuozzo Containers (Virtuozzo, 2022), Denali (Whitaker et al., 2002), Google Native Client (Yee et al., 2009), Vx32 (Ford and Cox, 2008), User-Mode Linux (Dike, 2006; User-Mode Linux, 2022), Minix Over Linux (Pessolani and Jara, 2011), Ensim (2022), LXC (Canonical Ltd., 2021), Docker (Docker, 2022; Ryding and Johansson,
• **Same-ISA Dynamic Binary Optimizers** are translators implemented in software that perform optimized translations of binary code with an equal ISA. Their operation is transparent, and even the system’s native binaries can be optimized. An example of this is the Dynamo project (Bala et al., 2011).

• **Operating System Translators** allow the execution of applications built for OSs different from the system host, e.g., WINE (Jones et al., 2018; Wine, 2022), WABI (Oracle, 2018), Lxrun (2022), Visual MainWin (Fisher et al., 2006), and Vcuda (Balis et al., 2021; S. Lin et al., 2009).

• **Dynamic Translators**: Dynamic ISA translators can support processes that use the same host OS, e.g., FX!32 (Chernoff et al., 1998). It can also be the case of dynamic ISA translations for processes that use a different OS than the host, such as Transitive (eWeek, 2008; IBM, 2008). For the above cases, the translation occurs at the Library level. It can also be the case of dynamic...

---

**Figure 10.** Proposal for a new virtual machine taxonomy

*Source: Authors.*
ISA translators for processes using a different OS and acting through high-level languages such as Java Virtual Machine (JVM) (Lindholm et al., 1997; Beronic et al., 2021), the Microsoft .NET common language infrastructure (CLI) (Thai and Lam, 2003), and Parrot (2022).

**Taxonomic key diagram**

This work also proposes a taxonomic key diagram to guide decision-making about the technologies related to VMs, as indicated in the proposed taxonomy (Figure 11). The diagram uses a set of questions, which, depending on each possible answer, establishes a path that leads to identifying a VT defined in the aforementioned taxonomy. For example, the diagram can be used by asking the question ‘Do you need to virtualize the entire system or just some of its processes?’ If the complete system needs to be virtualized, the following question will inquire about the specific need. If the desired virtual system needs an ISA different from the underlying hardware, the answer from the taxonomic key is the Dynamic Binary Translation category, e.g., QEMU, Simics, and Bochs.

**Conclusions**

A review of literature on the different classification schemes for virtualization technologies proposed since 2005. These schemes have been introduced using a timeline that has allowed the identification of the following taxonomic approaches: Abstraction Level, Virtual Machine Type, and Virtualization Domains.

When performing the analysis of each classification scheme, it was possible to identify weaknesses. These include the presence of a single taxonomic approach in each scheme and the lack of topicality considering the date of publication, as well as the absence of the details on the inclusion of technologies.

The proposed taxonomy responds to the needs identified in the analyzed classification schemes. As a result, the proposal combines the Abstraction Level and Virtual Machine Type approaches, giving the reader a means of visualizing the virtualization technologies relating to virtual machines. By doing so, the reader is always aware of the level of abstraction at which each technology takes place, in
addition to the type of machine projected, be it a complete system or an execution environment for processes.

The proposed taxonomy can be used in academic contexts to facilitate teaching and learning or in the business field to favor decision-making when implementing technologies related to virtual machines.

The taxonomy allows for the classification of VTs present in more than one conceptual branch, as these tools evolve, meeting the needs of more than one approach by themselves or using extensions.

Finally, a taxonomic key diagram has been created for use by the industry in order to aid the selection of virtualization technologies.

References


