METHODS

# Virtualization in distributed system: a brief overview

**Reeya Manandhar and Gajendra Sharma**[*]

Department of Computer Science and Engineering, Kathmandu University, Dhulikhel, Nepal

[*]**Correspondence:**
Gajendra Sharma,
gajendra.sharma@ku.edu.np

Virtual machines are popular because of their efficiency, ease of use, and flexibility. There has been an increasing demand for the deployment of a robust distributed network for maximizing the performance of such systems and minimizing the infrastructural cost. In this study, we have discussed various levels at which virtualization can be implemented for distributed computing, which can contribute to increased efficiency and performance of distributed computing. The study gives an overview of various types of virtualization techniques and their benefits. For example, server virtualization helps to create multiple server instances from one physical server. Such techniques will decrease the infrastructure costs, make the system more scalable, and help in the full utilization of available resources.

**Keywords:** distributed system, virtualization, hypervisor, technology, virtual machine

## 1. Introduction

The concept of virtual machines has been in the computing community since the early 1960s [1]. The recent advancement in virtualization technology pursues new opportunities in delivering services and represents a strategic approach for IT operators to increase their organizational performance. The virtualization technique offers an isolated, secure system with flexible deployment. Such mechanisms maximize the reliability, scalability, and fault tolerance of distributed systems [2].

A distributed system, also known as distributed computing, is a system consisting of a collection of independent components located on different machines that communicate and coordinate actions with each other and are linked together using the network. All of the nodes in this network interact and coordinate with each other to complete tasks, and it appears as a single coherent network to the end-user. The system is highly efficient as the workload is split and distributed to various nodes for fast completion. The task running in different nodes forms a common system where multiple machines can process the same function at the same time. As computation takes place separately on each node, adding additional nodes and functionality as needed is simple, affordable, and fault-tolerant.

Virtualization is the process of creating virtual instances of a computer system abstracted from real resources such as OS, storage devices, or network systems. With it, multiple instances can run independently and simultaneously on a single system. Technology advancements have propelled virtualization to the spotlight of the IT sector. This technology encompasses a variety of mechanisms and techniques that address various problems such as performance, reliability, security, and resource expenses, which are the reasons for the rise of this technology.

Distributed virtualization is the process of transparently sharing resources from several users [3]. Each node is isolated and does not interfere with the other. It helps the end user to access, store, analyze, and organize the distributed system components. Various types of virtualizations can be used to increase performance in distributed systems. Some of them are OS virtualization, storage virtualization, and network virtualization. Each of them has its own paradigm that offers a reliable, secure, and cost-effective distributed system.

This study gives an overview of virtualization, its type, and its benefits. The rest of the article is organized as a literature review in section "2. Literature review", different

types of virtualization techniques and their benefits in section "3. Different types of virtualizations and their benefits," and the final section with a concluding note.

## 2. Literature review

The study (4) presents an x86 virtual machine monitor, the Xen hypervisor, that enables multiple operating systems to share conventional hardware in a secure and resource-managed manner while preserving performance and functionality. It provides applications such as server consolidation, co-located hosting facilities, distributed web services, secure computing platforms, and application mobility. It provides a platform for the deployment of network-centric services and addresses the issue related to host transient servers. It is targeted to host up to 100 virtual machine instances simultaneously. A single virtual machine can host a real operating system, which may be securely multiplexed into thousands of unmodified operating systems. For example, Linux and Windows XP can be hosted simultaneously with a negligible performance overhead. Each guest OS handles paging on its own, with its own guaranteed memory and disk allocation. Guest OS refers to OSes that Xen can host. Using different virtualization strategies to assess performance in a variety of microbenchmarks and system-wide testing, Xen significantly outperforms rival commercial and open-source alternatives.

The study (5) presents a self-virtualized device approach for high-performance I/O virtualization in high-end systems, permitting guest virtual machines to more efficiently interact with the virtualized device. A self virtualized network interface (SV-NIC) is developed and tested, which targets the high-end NICs used in the high-performance domain. The SV-NIC offers virtual interfaces (VIFs) to guest virtual machines, manages the device's physical resource utilization, and delivers high-performance network connectivity to guest domains. The experimental results demonstrate that it is highly scalable and is about 2X better than the driver-domain approach. It offers 77% higher throughput and 53% latency reduction as compared to the driver-domain approach and virtual interfaces compared to the Xen hypervisor. The performance of the SV-NIC enables high performance as it reduces hypervisor involvement and eliminates driver-domain involvement in the I/O path.

The study (6) proposes a distributed virtualization system (DVS) that allows several instances of a distributed virtual operating system (VOS) to operate on the same cluster and can be extended beyond the physical machine limits. It allocates a subset of nodes for each instance and shares nodes between them, i.e., supports resource aggregation and resource partitioning. Each VOS runs isolated within a distributed container (DC) and can be executed in different nodes using the same abstract resources as offered by the DVOS. Each single virtual resource provided by the host

OS on each node acts as a resource allocation unit. Such granularity allows better use of resources and provides greater elasticity and efficiency. It simplifies application deployment time and cost and benefits administrators from the process location transparency because it eliminates the need to deal with IP addresses, ports, URLs, and other details. A DVS is suitable for software development applications based on microservices architecture (MSA) or service-oriented architecture (SOA), as it is inherently distributed. The proposed model integrates virtualization and DOS technologies to give the benefits of both, making it appropriate for delivering high-performance cloud services, including high availability, replication, elasticity, load balancing, resource management, and process migration.

The study (7) explores container-based virtualization as a lightweight alternative to the hypervisor-based approach, which can be adopted on smart objects for enhancing the IoT Cloud service provisioning. It is also known as operating-system-level virtualization, which partitions the physical machine resources, creating multiple isolated user-space instances. It allows IoT Cloud providers to deploy and analyze its performance on smart objects in the perspective of an IoT scenario both in terms of cloud service management and business opportunities. The two general usage models for containers are the application container where a single application runs in the container and the system container where multiple isolated instances of user space run at the same time, each one with its own in it process, process space, file system, and network stack. This technology is also supported by SBC (single-board computer) devices equipped with a Linux Container Virtualization (LCV) layer. Some popular container engine solutions include Docker, LXC, lmctfy, and OpenVZ. An experiment is performed using Raspberry Pi and Docker container engine to investigate the response of the device when no container and several containers run according to several configurations that could be adopted in real use cases. The result shows that the overhead of container virtualization is minimum with respect to the overhead due to processing the requests by CoAP servers.

In this study, the Service Level Enforcement Discipline for Storage (SLEDS) system (8) provides statistical performance guarantees on a storage system built from commodity components. It is a distributed controller that manages client workloads to meet QoS goals. It guarantees that each client's storage performance is good as if a subset of physical resources were dedicated to it and unaffected by the behavior of other clients. It does so by collecting performance samples regularly and throttle (delaying) I/Os from overly demanding clients whenever other clients suffer poor performance. It will redirect resources away from applications that are receiving adequate QoS toward those with poor performance. The reliability and performance of I/O service are maintained by decoupling the gateway from communications with the server and are based on a storage area network system

for block-level storage service. Testing the prototype with heterogeneous clients, they concluded that the system will facilitate both consolidated and non-consolidated worlds where applications benefit from the reliability of high-end storage devices and no longer suffer performance failures due to resource limitations.

The study (9) proposes an architecture, Advanced Cloud Protection System (A), for monitoring the integrity of guest virtual machines and the cloud infrastructure components while remaining fully transparent to the service user and the service provider. It shows how virtualization can increase the security of cloud computing. It protects the guest's virtual machines from intruders and attacks such as worms and viruses. The study highlights various issues of cloud computing, such as user access, privacy, data segregation, and recovery mechanism, and develops a prototype to address those issues. The prototype of the proposed ACPS has been fully implemented on two open-source solutions: Eucalyptus and OpenECP, and is tested for its effectiveness and performance evaluation. The findings show that the system is resistant to attacks and incurs minimal overhead.

The above-mentioned reviews are mentioned in short in the above table.

# 3. Different types of virtualizations and their benefits

## 3.1. Storage virtualization

Storage virtualization is the process of grouping multiple physical storage arrays from a virtual area network (SAN) and making them look like a single virtual storage device. The process involves abstracting the disks and drives and presenting them as a single storage unit to facilitate the application and network-independent management of storage. It gathers information about the amount of storage capacity accessible across physical storage components, combines it, and makes it available to applications (10). It provides better utilization and management of storage in heterogeneous environments. Storage virtualization helps to address this problem by facilitating easy backup, archiving, and recovery tasks by consuming less time.

Some of the virtualizing storage are as follows:

- Simplicity:
  It simplifies the IT environment as it eliminates complexities related to data centers by reducing the amount of hardware needed to run applications. It enables dynamic storage utilization and virtual scalability of attached storage resources.
- Affordable
  It reduces costs related to hardware and other operational costs as it masks the complexities of

managing infrastructure and resources in memory, networks, servers, and storage.
- Storage utilization
  It helps in properly partitioning the storage. The most commonly used data are stored in its highest performing storage pool, and the least used data are put in the weakest performance storage pool. They can be quickly retrieved from virtual storage.

## 3.2. Operating system virtualization

OS-level virtualization is a technology that partitions the standard operating system to create multiple isolated virtual machines (VM) so that it can run different applications handled by multiple users on a single computer at a time. The kernel of an OS allows more than one isolated platform instance. Such instances are called containers or virtualization engines. Such a technique is installed on a pre-existing operating system, and that operating system is called the host operating system (11). The operating system in OS-level virtualization is configured in such a way that it behaves like numerous other independent systems. The virtualized environment accepts instructions from different users running different applications on the same machine and does not interfere with each other. It provides application-transparent virtualization to users by isolating apps from the OS and offers granular control for individual applications.

The advantages of operating system virtualization are as follows:

1. Save time and cost as the user does not require to install extra hardware support. Fewer machines mean minimal maintenance, less power, higher electrical saving, and less cooling system requirements.

2. As every instance of this system is virtual, no extra space is occupied by it.

3. It is not only compatible with small businesses but also useful for large businesses.

4. By integrating server hardware, OS virtualization distributes resources across individual servers.

## 3.3. Network virtualization

Network virtualization is the process of transforming network functions by combining the available resources in a network to consolidate multiple physical networks into one virtual, software-based network (12). It abstracts network resources from the underlying hardware and enables virtual network deployment. Each of the networks is secure and independent from the others and can be allocated to a particular server or device in real-time.

**TABLE 1 |** Table representation of above-mentioned related work.

| References | Objective | Algorithm/Mechanism | Tools/Protocol |
|---|---|---|---|
| Barham et al. (4) | Enable users to dynamically instantiate an operating system to execute whatever they desire. | Borrowed Virtual Time (BVT) scheduling algorithm. | 2.4 GHz Xeon server with 2GB RAM, a Broadcom Tigon 3 Gigabit Ethernet NIC, and a single Hitachi DK32EJ 146GB 10k RPM SCSI disk. |
| Chambliss et al. (8) | Statistical performance guarantees a storage system and improves its effectiveness. | Leaky bucket mechanism. | IBM xSeries 345 servers, with a 2.4 GHz Intel Xeon proces- sor and 512 MB (client)/2 GB (gateway) RAM, FC SAN fabric. |
| Raj and Schwan (5) | Investigate the implications of virtualization for the high-performance domain, virtualization of peripheral resources. | Round-robin scheduling algorithm. | IXP2400-based ethernet board. |
| Lombardi and di Pietro (9) | Identify the security issues of cloud computing and provide a solution to the issue. | Based on (KvmSec: A security extension and KvmSma: Security management architecture) for the Kernel Virtual Machines. | Python, livibirt, Mysql, Turbogears. |
| Celesti et al. (7) | Explores container virtualization to analyze its performance on smart objects from the perspective of a IoT scenario. | Container based virtualization, Linux Container Virtualization (LCV). | Raspberry Pi and Docker container engine. |
| Pessolani (6) | Integrates Virtualization and DOS technologies to take the benefits of both and making proposed system suitable to deliver provider-class Cloud service. | Load balancing, process migration, consensus, fault detection mechanism. | POSIX APIs, RPC protocol. |

Every client has shared access to all the resources on the network from a single computer. There are two types of network virtualization: internal virtualization which provides network-like functionality to software containers to replicate the functionality of a single network server and external virtualization which combines multiple local networks into a single virtual unit. It intends to optimize network speed, improve productivity, and manage the network security of an organization.

Some of the benefits of network virtualization are as follows:

- Faster delivery of application:
  Network virtualization automates network configuration, which reduces the deployment time, allowing for faster delivery of product rollouts or major application updates.
- Enhances security
  Sensitive data is isolated from one virtual device to another. So, access to nodes is restricted, making data more secure.
- Improves manageability
  Functional grouping of nodes in a virtual network is eased. Virtual networks can share the same physical network, which enhances the utilization of network resources and allows communication between nodes.

## 3.4. Application virtualization

Application virtualization, also known as process-level virtualization, is a technology that allows users to access and utilize an application from a computer or separate device other than the one on which it was installed as if it were installed on a user's local device. It is useful when you simply want to virtualize only an application. The server-based technique is the most common method to virtualize applications (13). Remote applications can be set up on a server by a service provider and sent to an end user's computer. The user's instructions are sent back to the server to be executed. User experience is the same as that of an app installed on the real system.

Some of the benefits of virtualized applications are as follows:

- Simplified management
  Instead of manually installing apps on each user's workstation, IT administrators simply install an app once on a central server and then distribute it as required on user devices via app virtualization, making it easier to manage and maintain the application.
- Scalability
  Virtual applications can be deployed on any connected device regardless of that device's operating system as the app is actually running on the centralized server. This will lower the organization's hardware computing cost. As remote services, many companies also share application access. It results in a reduction in software and infrastructure costs. Such cost reductions are obtained by reducing organizational needs for several servers, which benefits in decreased maintenance and power costs.

- Security
  Security is maintained across several users. The IT admin is responsible for managing resources and controlling user access to the particular application. If a user's device is lost, then the IT admin can revoke sensitive data through remote access.

## 3.5. Server virtualization

Server virtualization is the process of creating multiple unique server instances from one physical server, each running its own operating system independently of multiple small, isolated virtual servers. It is a cost-effective way to provide web hosting services and has boosted resource utilization to the fullest without having to invest more in hardware. For creating virtual server instances, a virtualization software layer called a hypervisor is added to a host machine (14). The hypervisor separates the host's physical resources from virtual instances. Then virtualization software can be used to emulate the physical resources and create a new virtual server on top of it.

Some of the server virtualization benefits are as follows:

- Server consolidation
  Virtualization enables the partition of a single server onto multiple servers. Due to this, the total number of servers in the organization can be reduced. This process is called server consolidation.
- Cost reduction
  Server consolidation minimizes costs by increasing the utilization of existing resources. This eliminates management costs for those servers, as well as reduces the energy requirements needed to run the servers.
- Saves space
  We can potentially save a lot of physical space by hosting multiple virtual servers on fewer physical servers.

## 4. Conclusion

The study presents an overview of virtualization in distributed systems. Various papers related to virtualization like Xen hypervisor, container-based virtualization, security of cloud computing, and storage performance are discussed in this study. Additionally, different types of virtualization techniques with its benefits are described. In conclusion, virtualization in distributed systems enhances performance, reliability, fault tolerance, flexibility, and scalability, and minimizes the infrastructure cost of an organization. As resources are shared transparently among the nodes, security, privacy, and network management are some of the key concerns that need to be optimized to make the virtualization more secure and organized.

# References

1. Daniels J. Server virtualization architecture and implementation. XRDS: Crossroads. *ACM Magaz Stud.* (2009) 16:8–12. doi: 10.1145/1618588. 1618592

2. Kaul A, Gupta M, Kaur K. *Using virtualization for distributed computing.* Newton, MA: Techtarget (2015).

3. Shewandagn S. *Virtualization in Distributed Systems - Sertsedengle Shewandagn.* San Francisco, CA: Medium (2022).

4. Barham P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, et al. Xen and the art of virtualization. *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles – SOSP '03.* New York, NY (2003). doi: 10.1145/945445.945462

5. Raj H, Schwan K. High performance and scalable I/O virtualization via self-virtualized devices. *Proceedings of the 16th International Symposium on High Performance Distributed Computing - HPDC '07.* New York, NY (2007). doi: 10.1145/1272366.1272390

6. Pessolani PA. An architecture model for a distributed virtualization system. *J Comput Sci Technol.* (2019) 19:e17. doi: 10.24215/16666038. 19.e17

7. Celesti A, Mulfari D, Fazio M, Villari M, Puliafito A. Exploring Container Virtualization in IoT Clouds. *Proceedings of the 2016 IEEE International Conference on Smart Computing (SMARTCOMP).* Nashville, TN (2016). doi: 10.1109/smartcomp.2016.7501691

8. Chambliss D, Alvarez G, Pandey P, Jadav D, Jian X, Menon R, et al. Performance virtualization for large-scale storage systems. *Proceedings of the 22nd International Symposium on Reliable Distributed Systems.* Manhattan, NY (2003). doi: 10.1109/reldis.2003.1238060

9. Lombardi F, di Pietro R. Secure virtualization for cloud computing. *J Netw Comp Applic.* (2011) 34:1113–22. doi: 10.1016/j.jnca.2010.06.008

10. Castagna R, Brown R. *Storage virtualization. SearchStorage.* Newton, MA: Techtarget (2021).

11. GeeksforGeeks. *Operating system based Virtualization.* Noida: GeeksforGeeks (2021).

12. Shtil D. *What Is Network Virtualization?.* Santa Clara, CA: Gigamon Blog (2020).

13. Citrix. *What is Application Virtualization? App Virtualization Definition - Citrix.* Fort Lauderdale, FL: Citrix (2023).

14. Bigelow SJ, Gillis AS. *What is server virtualization? The ultimate guide.* Newton, MA: Techtarget (2021).

15. Brush K, Kirsch B. *Virtualization. Search Server Virtualization.* Newton, MA: Techtarget (2021).

16. Figueiredo R, Dinda P, Fortes J. Guest editors introduction: resource virtualization renaissance. *Computer.* (2005) 38:28–31. doi: 10.1109/mc. 2005.159

17. Shukur H, Zeebaree S, Zebari R, Zeebaree D, Ahmed O, Salih A. Cloud computing virtualization of resources allocation for distributed systems. *J Appl Sci Technol Trends.* (2020) 1:98–105. doi: 10.38094/jastt1331