Evaluation of Disaster Recovery Through Virtualization

Jasime Fitt
Department of Computer Science
Colorado State University
Fort Collins, USA
jasminef@colostate.edu

Eldridge Harris
Department of Computer Science
Colorado State University
Fort Collins, USA
eldridge.harris@colostate.edu

Phillip Johnson
Department of Computer Science
Colorado State University
Fort Collins, USA
pdj1183@colostate.edu

***Abstract*— Virtualization is the method of duplicating virtual server infrastructure and data onto remote facilities. Many companies are transitioning to the use of virtualization as a form of disaster recovery due to cost saving and accessibility benefits. Disaster recovery refers to an organization's efforts to address technology-related accidents or disasters. These incidents range from natural disasters to human attacks. Such disasters can quickly devastate organizations' networks and databases if not properly taken into consideration when building these networks and databases. Virtualization and the use of cloud computing has piqued the interest of the IT industry. The cloud is a pool of virtualized computer resources, allowing for the computer resources to be accessible practically anywhere. Perhaps virtualization can become a valuable and efficient method in disaster recovery planning. Throughout this article, we discuss the importance of disaster recovery while also explaining the methodology and development of virtualization as a form of disaster recovery.**

Keywords—cloud computing, disaster recovery, virtualization, containers

I. Introduction

As more and more companies transfer to online systems and cloud-based technology, technology-related disasters and accidents have become more catastrophic for organizations and companies. This means that any period of system downtime can result in extreme financial loss, or even as extreme as putting human lives at risk. Within the last decade, there has been an explosion of system virtualization development and research.

Disaster recovery particularly raises concern due to the rising tension in Ukraine as of 2022. There are many companies based in Ukraine, if they have their data physically stored in Ukraine, their companies are very likely to fail. In a similar vein, 2012's Hurricane Sandy, one of the worst natural disasters to hit the United States, disrupted many communications and networks. Disasters like these cannot always be predicted, thus creating a disaster recovery plan is essential to protecting data and networks.

Generally, many disaster recovery services come at a high cost and do not guarantee full recovery of data, nor time required to restart operations. According to Wood, current disaster recovery methods include "periodic tape backups that are trucked offsite" and "continuous synchronous replication of data between geographically separated sites" [12]. Typically, disaster recovery services replicate application state between two different data centers. If the primary data site is unavailable, then the second site, also known as the backup site, takes over and creates a new copy of the application that uses the most recently replicated data [12].

Virtualization protects data in the case that the physical location where IT functions are operated is destroyed. Virtualization first appeared in the 1960s when IBM created virtual machines (VM) that could hold multiple instances of a single-user CMS operating system, all of which could run parallel [2]. Thus, since a VM is not dependent on physical hardware, multiple VMs can be installed on a single piece of hardware, along with multiple operating systems. However, in the 1990s, virtualization became more important. Companies discovered ways to virtualize the Intel x86 processor architecture, which was groundbreaking because the x86 processor was much less accommodating [2]. Before virtualization, a machine could only handle one OS image, could only run one application without conflict, and machines had costly infrastructure [9]. However, after virtualization, VMs could be used on just about any system and OS and applications could be managed by boxing them together into VM [9].

Virtualization has encouraged companies to re-evaluate their disaster recovery plans because it increases hardware utilization and also offers easy management of server infrastructure. As of 2020, 92% of businesses use some form of server virtualization [6]. This is an incredible increase from 2007, in which only 64% of companies invested in virtualization [1]. With the increase usage of virtualization, it is paramount to evaluate the efficiency, plausibility, and risks of virtualization as a method of disaster recovery in modern society.

II. Virtualization

Virtualization is a technology that allows for virtual instances of computers to be run separately from the physical hardware. According to Scroggins [8], virtualization is the "concept or process of separating the logical from the physical." With virtualization, the user can do more with less hardware and less physical technology. As Spinellis puts it, "virtualization means never having to beg for a server" [2].

Virtualization can be applied to all different types of technology, the most popular being server virtualization. However, some other classifications of virtualization technology include clustering, desktop virtualization, storage virtualization, network virtualization, data virtualization, and mobile virtualization [8]. While one might think of virtualization as a large-scale operation, it comes in many forms that are all important when considering virtualization as a form of disaster recovery.

Virtualization through cloud computing "allows the coexistence of multiple service providers or tenants in an infrastructure formed by optical networks interconnecting datacenters" [4]. In this infrastructure, cloud networks formed by virtual networks that are connected to VMs, allow for the requirements needed for processing and storage. This allows for cloud providers to virtualize servers and storage.

III. Data Center Disaster

With any large-scale data operation, unexpected mistakes and errors are bound to happen. As stated by Murphy's Law, anything that can go wrong will go wrong, eventually. The main objective and goal of any data operation is to stay operating and functional through any event, whether that be a natural disaster, an event of corruption, or even just a small system failure. The main objective of having a recovery plan in place is to prevent a data disaster from having a negative effect on the systems impacted by the event.

In simple terms, a data disaster is any hiccup in the normal processes of a storage device that causes important files and documents to be destroyed or deleted unexpectedly. This alone can cause the normal operations of a company to be halted, and the repercussions of data loss are very serious. According to the National Archives and Records Administration [5], 43 percent of companies that do not have a disaster recovery plan will go out of business after experiencing data loss. In addition, 93 percent of companies that experience data loss for 10 days or more will file for bankruptcy within 5 months [5].

IV. Benefits of Data Disaster Prevention and Disaster Recovery Plans

Due to the potential of catastrophe when a data disaster occurs, there are many benefits to having a plan in place to fall back on in that scenario. According to Evolve IP "Cloud-based data storage and backups simplify the process of archive maintenance, enhance the effectiveness of backups, and reduce the cost of disaster recovery" [5]. These three benefits alone make having a backup recovery system in place essential for and operation, but there are three main benefits that will be addressed here. The first is cost efficiency: by minimizing the risk of a data disaster occurring in the first place by getting rid of as many vulnerabilities as possible and adding detective measures that have the goal of ending any unwanted event quickly, the company will be saved countless amounts of lost profits and headaches. Secondly, the company will retain customers much more consistently as a reliable company without hiccups keeps better customer relations and continues to meet customer expectations day in and day out [5]. Finally, having cloud-based backup services allows for company scaling to be completed much easier and with much less risk. With all data backed up and secure, there is much more flexibility in upscaling without the need for onsite or offsite maintenance [5]. Having a plan in place for data disaster not only minimizes large risk from unexpected events, but it also allows for a more streamlined IT process and minimizes the risk of simple human error.

There are a few requirements that are considered while deciding what an effective disaster recovery service does. Performance, consistency, and geographic separation all determine how effective a disaster recovery service is. Companies must also calculate their own recovery point objectives and recovery time objectives to determine when the most recent backup occurred and how long it takes for an application to come back online after a failure [12]. A disaster recovery service must have minimal impact on performance. That is, disaster recovery services can impact performance directly or indirectly, but an effective disaster recovery service will have minimal impact on the application. Furthermore, the disaster recovery service must also be consistent. When a disaster occurs, the disaster recovery service must be able to restore applications to a consistent state. An example of this is keeping a consistent copy on a disk, then using a disk replication scheme to make consistent copies at a backup site. Lastly, the disaster recovery service must have its primary and backup sites geographically separated in order to be affected. This means that a single disaster should not affect both sites.

A disaster recovery service must also be able to detect when a disaster has occurred in order to activate the backup site. This has been proven difficult due to "transient failures" or "network segmentation," which can trigger false alarms [12]. However, this has been proven to be the most efficient with virtualization.

V. Virtualization Solutions

There are multiple virtualization processes that have been created with the goal of data retention, recovery, or backup. These can be to create a system that will protect the data inside data centers. Each of these solutions target a certain data loss disaster and has its own methods to protect against loss of data. As explained previously, there are many ways that data loss can occur so there is no perfect solution to specifically combat data loss.

A. *Container Based Virtualization [3]*

For this section the device that needs to be protected will be called the Running System, and the system responsible for backing up and restoring will be called the Disaster System.

This form of virtualization is implemented to act on the operating system layer which allows it to have greater knowledge of what the Running System is doing. This will allow the Disaster System to take finer control of the backup and recovery process (e.g. urgent data can be restored in advance so that the Running System can access it sooner [3]). This system also allows for a low amount of running time overhead allowing it to be efficient on large scales. The Running System and the Disaster System are isolated from each other in order to allow the Disaster System to control the Running Systems while creating the backup checkpoints.

This also means that the Running System and the Disaster System's disks will be completely disconnected giving further separation in the event of disaster. To initialize the backup the Disaster System would do a full disk backup and for the actual incremental backups the Disaster System uses backup checkpoints that contain recently changed data as well as the corresponding memory states in order to prevent data inconsistence. The actual architecture for the system will be a Physical to Virtual/Container (P2V/C).

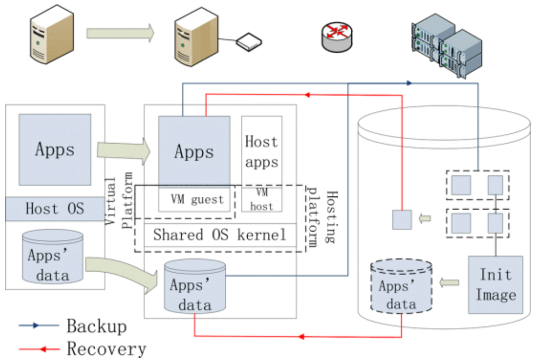

Fig. 1. Disaster System Architecture with backup and recovery paths [3]

The benefit to using this form of virtualization is that there is a completely sepreate system that is in charge of the backup, so in the event of failure on the Running System, it is unlikely that the Backup System will also be harmed. In order to keep this true the two systems should be seperated in different cities or states so that if there were to be a natual disaster or power outage the two systems would not both be taken out. It also seems to be a well thought out system that is optimized for permormance and data efficiency. Data backup is inheritly data costly as you need to have the exact the same amount of storage on the Backup System as the Running System. This can be costly on very large data systems, but the safety of having a constant backup of the whole running system could be worth that cost.

B. *Agent Based Cloud Virtualization [4]*

This version of virtualization utilizes a cloud network to connect the user to a data server. This allows for On-Demand Self-Service as a way for users to control the system remotely without the need for human interaction with each cloud service provider. This version of Cloud Virtualization is organized by the Cloud Manager in which the user interreacts with. It oversees the authentication and validation of the user. After verification it handles fulfilling the users' requests for data storage and retrieval. The Cloud Manager gets assistance from a hypervisor to create the virtualization. The hypervisor is in charge of creating virtual nodes using the system resources and keeping a list of those created nodes. There is then a Fault Handler that looks at the data that is accessed from the nodes and determines whether there are any data faults in that node. If there are then the system uses erasure code to scan the node and find which files are available and which ones are not. After this it uses those available files to recover the destroyed files. Artificial intelligence is used thought the system to scan for changes to the data that was made by malware or other security vulnerabilities.

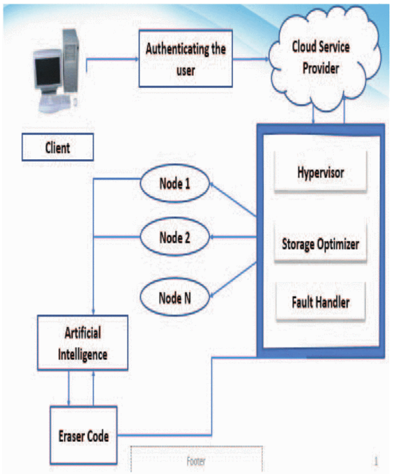

Fig. 2. Data Recovery System using cloud networks [4]

Erasure Code is a method of data loss protection that works by breaking data up into smaller sectors that are then all stored on different drives. This allows the system to keep the data in different areas in small chunks so that in the case of failure the amount of data for a file in each chunk is minimal. It would allow the system to recover the lost data in files easily. It is able to accomplish this by breaking the data down into chunks that can be used to recreate the whole of the data with only a few chunks. For example, if given the number 187 the erasure code would break the number into 18 and 7 and then create some equations that can be satisfied with those numbers, $x + y = 25$, $x – y = 9$, and $(x – 8) + y = 17$. With those equations you only need two of them to survive to solve what the original number is. This is different from the normal RAID solutions that are employed in data systems. RAID allows data to be stored in multiple places whereas the Erasure code encodes the data and then stores it in different locations. [13]

This form of virtualization allows for a system of data recovery that does not require having to run your own backup system and pay for the storage requirements that entails. This also has security protections already baked into the system to keep malware from destroying the entire system. In addition to that, it allows the user to access the cloud network without having to physically travel to the server. One large downfall is that this system does not protect well against full outages. Since the nodes are all virtualized from the cloud server, if the cloud service provider were to have issues, then there is a chance that the data in the cloud network could be lost entirely

C. *Cloud-Network Cascading Failure Protection [5]*

This is a continuation of the previous virtualization method however, it provides utility to protect against failures that would arise from disaster, or other cloud network disasters. Cloud Computing systems tend to be made up of a large number of datacenters that are connected though virtual networks or by optical cabling. This would be a system that is implemented in all of the data centers. This would allow the different centers to communicate during disasters and work together to save the data.

Cloud Network Disaster Recovery reacts either a loss of communication between the Cloud Networks in a system and the Datacenters or a node failure in the Cloud Network. For loss of communication, it creates a new virtual link or remaps an existing virtual link to reconnect the Cloud Network to the datacenter. In the case of node failure, the Cloud Network will select an available backup node and migrate the failed node to the backup node. It will then create new virtual links to the rest of the network allowing access to the new node.

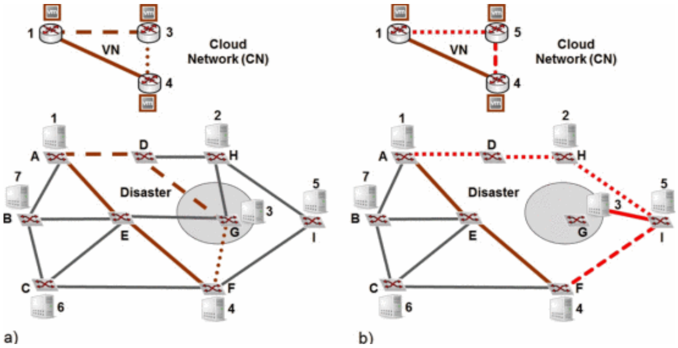

Fig. 3. Cloud Network Data Recovery where the connections to server G have been destroyed [5]

The figure above shows a scenario where the connections to a server are destroyed. The network can recovery by creating a new virtual link between servers G and I. This allows the servers to still access the server under disaster by passing through other servers.

Datacenter disasters rarely are caused by isolated issues, and there are usually cascading failures the build and affect more services than the original disaster. To protect against this, we can make the Cloud Network Recovery system aware of the possibilities of cascading failure. This system works the same as the one explained above, but it has added objectives to help reduce the total amount of damage sustained from cascading failures. This works by adapting as more systems begin to fail and analyzing which new virtual link would provide the most bandwidth to all of the affected servers. It was found that this way of looking for cascading failure reduced the expected bandwidth loss by 90% [5].

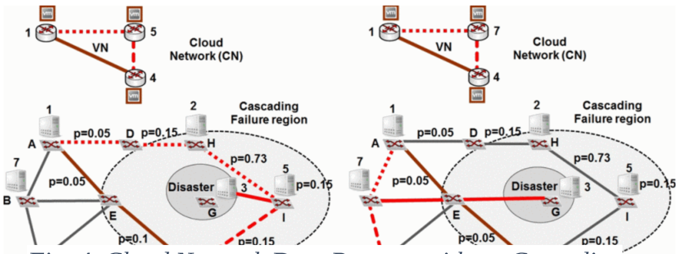

Fig. 4. Cloud Network Data Recover without Cascading Failure (a) vs. with Cascading Failure Awareness (b)

In figure 4 it shows how the different ways the two systems will react to a cascading failure in the area around server G. The system with Cascading Awareness selects a different virtual link to communicate with the affected servers allowing higher speed communication in the system.

VI. Optimal Solution

Each of the systems above all have negatives associated with their operation so there is no one perfect answer to protecting data using virtualization. The various solutions all have a large price requirement that comes with them, whether it be a large amount of storage required to back up the whole Running System, or the infostructure required to implement a large-scale Cloud Computing Virtual Network and develop systems that are capable of virtualizing communication links or virtual nodes. The cost to implement Container Based Virtualization solution on a large-scale commercial data center would be too large so it would be recommended to use the Cloud Network Disaster Recovery system as it would be more cost efficient and is much easier to scale up to larger data server system. For smaller companies or individual use then the Container Based Virtualization would be more practical to implement and the cost for adding more storage would be a lot lower than creating a whole Cloud Network.

VII. Benefits of Virtualization

There are several benefits to virtualization, beyond just disaster recovery. According to Kedia [11], virtualization allows for cost reduction (server consolidation and less hardware), better hardware utilization (many operating systems can be hosted at once), reduced down time, dynamic load balancing, high availability of resources, scalability, reduction in power consumption, server consolidation, and much more. When compared, “the costs of running disaster recovery services using public cloud or privately owned resources… shows cost reduction of up to 85% by taking advantage of cloud resources” [12]. While that just a small list, there are many other benefits to virtualization. A company may choose virtualization as a form of disaster recovery due to efficiency and ease.

VIII. Restrictions

A. *Security Challenges of Virtualization*

Because VMs are hosted in a physical machine, VM security and physical machine security are important to take into consideration. Research [10] suggests that establishing security systems in VMs is complicated.

However, there are some threats posed to these systems, which target vulnerabilities in cloud computing. Due to the nature of virtualization, which has multiple VMs running on the same physical machine, attackers can access information of other physical machines [10], which is also known as VM Theft. Another security vulnerability is VM Escape, in which the attacker can take control over administrative tasks in order to modify user access and create VMs [10]. This causes concern because that means that hackers have the ability to breakdown the security of an entire system through using only an exploitation [10].

Furthermore, the usage of public cloud computing may also raise concern for security. Companies cannot guarantee that their data and information will be safe in a public cloud. Thus, cloud service providers must ensure that there is proper security for data. Clouds service providers must also guarantee the performance of applications running on their service in the case that a disaster occurs within their own services.

B. *Transitioning to Virtualization*

While the bells and whistles of technology are enticing to IT departments, it is difficult to get an industry-wide standard. Technology develops exponentially, therefore making development of technology extremely rapid in the 21st century. This means that IT departments of companies will never keep up with the new and shiny technology. People do not like change, thus changing the architecture of a company's way of doing things is difficult. To get a company's whole working force to adapt to new systems is challenging, thus making virtualization tricky.

IX. Future Projections

The implementation of virtualization in the current day and age is a non-negotiable for any company with a need for consistent data access and storage. As stated throughout this paper there are many cons that stem from unreliability and the slightest possibility for a failure in a system and having a strong, well-developed solution and plan for any accidents or unforeseen events that could occur is a barebones requirement for a successful business. With this being stated, virtualization in companies that have not already implemented a system for data protection will quickly become the new normal, regardless of the initial difficulties that will stem from adding another layer of technology onto the current of the company in question.

Traditional hypervisor virtualization and container-based virtualization are becoming more and more mixed due to the increase in DevOps incorporation. While many companies are working on switching to cloud-based virtualization, there are still many applications that containerization does not support. However, this is likely to change, due to the gradual migration to new technologies that better support companies' financial and security goals.

The other main reason that virtualization will become so prevalent, is the fact that data is much more secure from tampering attempts when distributed throughout multiple locations and servers. This means that the failing or hacking of a single instance of storage would not cause the entire system to be compromised and would just be restored by a convenient backup. Overall, with the implementation of a reliable form of virtualization, the inevitability of data loss is made very unlikely and the ability to restore any lost data is simple and easy with these systems. This reasoning alone will make virtualization the data protection method of choice for nearly all companies in the very near future.

According to Woods [12], his team is developing a prototype disaster recovery system that will explore the tradeoff between cost, recovery point objective, and recovery time objective. This service, known as Dr. Cloud, will be interesting to develop more research and hopefully create leads in order to optimize the relationship between cost, recovery point objective, and recovery time objective.

X. Conclusion

While data loss is inevitable, it is suggested that companies and organizations create data recovery plans to prepare and accommodate for such disasters. The IT industry has had a sweeping push towards virtualizing within the past decade.

Virtualization allows for cost effectiveness, hardware-minimization, accessibility, flexibility, scalability, and simplified administration, which is alluring for many companies. Companies, such as Android, have been utilizing virtualization to create technology that costs less [8]. Within the next couple of years, it can be projected that virtualization will continue to be the IT-industry standard, as well as the creation and advancement of more security features to support virtualization.

There are many different types of virtualizations, as well as many different types of virtualization solutions. Companies and organizations can determine what the best solution for their company to protect their data that fits within their budget, as well as achieves their goals for disaster recovery.

References

[1] M. Murukutla, “Virtualization: Disaster recovery for the hospitality industry?,”*ScholarWorks@UMass*.[Online].Available: https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1020&context=gradconf_hospitality. [Accessed: 31-May-2022].

[2] D. Spinellis, “Virtualize me,” *IEEE Software*, vol. 29, no. 5, pp. 91–93, Aug. 2012.

[3] . Xu, H. Yu and W. Zheng, "A Consistent Backup Mechanism for Disaster Recovery that Using Container Based Virtualization," 2012 Seventh ChinaGrid Annual Conference, 2012, pp. 95-100, doi: 10.1109/ChinaGrid.2012.10.

[4] S. P. Carolin and M. Somasundaram, "Data loss protection and data security using agents for cloud environment," 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), 2016, pp. 1-5, doi: 10.1109/ICCTIDE.2016

[5] C. Colman-Meixner, M. Tornatore and B. Mukherjee, "Cloud-Network Disaster Recovery against Cascading Failures," 2015 IEEE Global Communications Conference (GLOBECOM), 2015, pp. 1-5, doi: 10.1109/GLOCOM.2015.7417558.

[6] “4 Benefits of Disaster Recovery Planning,” *EvolveIP*, 16-Jun-2018. [Online]. Available: https://www.evolveip.net/blog/4-benefits-disaster-recovery-planning. [Accessed: 31-May-2022]

[7] “Why a virtual environment may be your best bet for backup and disaster recovery,” *Arcserve*, 11-Jun-2020. [Online]. Available: https://www.arcserve.com/blog/why-virtual-environment-may-be-your-best-bet-backup-and-disaster-recovery. [Accessed: 31-May-2022].

[8] R. Scroggins, “Emerging virtualization technology,” *Global Journal of Computer Science and Technology*, pp. 11–16, 2017. https://globaljournals.org/GJCST_Volume17/3-Emerging-Virtualization-Technology.pdf

[9] N. Jain and S. Choudhary, "Overview of virtualization in cloud computing," *2016 Symposium on Colossal Data Analysis and Networking(CDAN)*,2016,pp.1-4,doi: 10.1109/CDAN.2016.7570950.

[10] Manjeet Gupta, Devesh Kumar Srivastava, and Durg Singh Chauhan. 2016. Security Challenges of Virtualization in Cloud Computing. In Proceedings of the Second International Conference on Information and Communication Technology for

Competitive Strategies (ICTCS '16). Association for Computing Machinery, New York, NY, USA, Article 101, 1–5. https://doi.org/10.1145/2905055.2905315

[11] P. Kedia, R. Nagpal, and T. Pal Singh, "A survey on virtualization service providers, security issues, tools and future trends," *International Journal of Computer Applications*, vol. 69, no. 24, pp. 36–42, 2013.

[12] Wood, T., Cecchet, E., Ramakrishnan, K. K., Shenoy, P., Van Der Merwe, J., & Venkataramani, A. (2010). Disaster recovery as a cloud service: Economic benefits & deployment challenges. In 2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10).

[13] S. F. Inc, "Understanding erasure coding and its difference with RAID," StoneFly, 10-Jun-2022. [Online]. Available: https://stonefly.com/blog/understanding-erasure-coding.