The Impact of Cloud Computing on Technology and Society

Fall 2018

Aimun Khan Benson Huang
John Koelling Matthew Barondeau Nimay Kumar
Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, Texas

***Abstract*— Cloud computing is a model in which computational and storage resources are provided as utilities that can be leased to users through the Internet. Virtualization, resource pooling, resource provisioning, and data encryption define different layers of the cloud computing architecture. Various service models operate in these different levels, providing the requested infrastructure, platform, or software to end users.**

Cloud computing is a prevalent model for hosting and delivering services over the Internet. Various service models and deployments have been established to accommodate new IT demands. In addition, the rapid growth of cloud services has prompted specialization and the advanced the prevalence of the Internet of Things. Despite cloud computing offering many advantages, the technology still faces data ownership ambiguity, security challenges, privacy issues, and environmental consequences.

I. INTRODUCTION

The National Institute of Standards and Technology defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]. This definition encompasses the core functionality of cloud computing, and introduces some of the key components of the cloud computing model. The technical foundation for cloud computing is in virtualization and resource management.

Different cloud service providers offer infrastructure, platform, and software as utilities, where customers only pay for what they need. These resources are scalable and can grow to meet the needs of the user request. Furthermore, the cloud platform is accessible anywhere on the globe with an internet connection allowing expanded accessibility over conventional computing models. This is a major shift from traditional business applications that require localized hardware and software resources.

In this paper, we will first provide a technical description of the various characteristics and technologies that compose the cloud computing model. We will decompose the cloud computing architecture and examine each layer. This paper will then discuss the impact of cloud computing by describing the models and use cases of cloud computing. Cloud computing increases the power of devices by tying their processing to a remote server environment, which results in more accessible applications for users. Smart devices have created new access points to cloud services in previously unconnected items. Cloud computing is not without risks however, as data must travel quickly and securely between locations for use. This technology introduces security and privacy concerns for data and brings forth unforeseen environmental consequences.

II. CORE TECHNOLOGIES

This section describes how virtualization, resource management, and data encryption technologies are leveraged to make cloud computing viable.

A. Virtualization

Virtualization is a foundational component of cloud computing, providing the ability to partition hardware systems into virtual machines.

Virtualization is done using a type of software called a hypervisor, which interacts with the hardware to partition systems into separate software-based environments known as virtual machines. These are containers that emulate environments for programs using physical resources (such as CPUs or storage devices) provisioned by their hypervisor [2]. The hypervisor implements virtualization by allowing a single physical host to operate multiple virtual environments or allowing a single virtual environment to be spread across multiple hosts.

Virtualization has other advantages over traditional environments. Hypervisors augment scalability by providing additional capabilities such as pooling computing resources from clusters of servers and dynamically assigning virtual resources to applications on-demand. Hypervisors also provide reliability benefits by seamlessly moving virtual resources to different physical locations in the event of individual hardware failures [3]. These principles provide the necessary foundation for cloud computing by simplifying and abstracting the management of hardware.

B. Resource Pooling and Privisioning

Resource pooling is a network design principle that abstracts several distinct resources into a unified pool that behaves as a single resource. Hypervisors use resource

provisioning techniques that detect and allocate resources to virtual machines to match their volatile workload.

Resource pooling refers to the technique of flexibly managing, grouping, and unifying numerous resources. For each resource pool, the provider can specify permission and storage limitations. Resource pooling is similar to a folder hierarchy, where a resource pool can contain other child resource pools or virtual machines [4]. Cloud providers can delegate control over resource pools, child pools, or virtual machines to users. Generally, cloud resources refer to any or all levels of the resource pool that can be provisioned to users.

The objective of resource provisioning is to detect and allocate the requested level of resources on time, so that applications can perform optimally. Infrastructure providers make a large amount of resources available from their data centers, which enables service providers to scale and expand their service to handle an increase in demands. Static resource provisioning techniques allocate resources to cloud workloads prior to resource scheduling through user specified workload requirements. In addition, dynamic provision handles fluctuations in requirements to maximize resource utilization during application run time by monitoring usage of resources and reallocating them accordingly [4]. These resource provisioning techniques provide users the necessary amount of computing resources.

C. Data Encryption

Given the volume of data transferred on Cloud servers, data transfer requests must be secure. Only users who have permission to access a particular service should be able to make requests. In addition, the pipeline through which that data reaches the user must be secure. For these reasons, data encryption and user integrity are integral components of Cloud computing.

Encryption involves obfuscating packets of data by modifying the bits. Only someone who knows the exact order of operations made to the packets can undo those modifications and read the original data [5]. One common standard of encryption used by many Cloud services is the Advanced Encryption Standard. AES is an algorithm for encrypting data that involves chopping up packets into 16-byte blocks and applying a series of substitutions and permutations to each block of data based on a predetermined key value [6]. Unlike linear operations like adding and multiplying, these bitwise operations are extremely hard to undo, making it nearly impossible to bruteforce the data into its original configuration. Only a user with access to the encryption key will be able to undo these transformations and read the data in its original form. This ensures that even if packets are intercepted by a third party during transfer, the third party will be unable to do anything with the stolen packets.

In order for this encryption to be safe, the keys to decrypt the data must also be securely shared between users. DiffieHellman Key exchange is a popular algorithm in which two users are able to agree on a key without ever sending the key over the Internet [7]. This is accomplished through data-to-key mapping algorithms known as hash functions, which are composed of basic functions like exponentiation and logarithms. Instead of authenticating data transfers by directly sending each other their keys, users apply their privately held hash functions to the data and verify that the result is the composition of the two hash functions. This method allows the data to be decrypted without exposing the privately chosen hash functions. Even if a malicious user intercepts this composition, the irreversible nature of hash functions makes it impossible to determine the keys from the composition alone. This type of authentication is necessary to safely transfer data at the large-scale necessary for Cloud computing.

III. Architecture

This section separates the cloud computing architecture into four layers: the hardware, infrastructure, platform, and application layers. These layers build on top of each other as shown in Figure 1.

A. Hardware and Infrastructure Layer

The hardware layer, as seen at the bottom of Figure 1, is composed of the physical servers, routers, switches, and power and cooling systems. This layer is implemented in data centers, which often contain thousands of servers. The hardware layer provides the physical resources necessary for virtualization [4]. The infrastructure layer creates a pool of storage and computing resources by using virtualization to partition the physical resources into virtual machine containers. The aforementioned technologies of hypervisors, resource pooling, and resource provisioning compose this layer. This layer provides the foundation for the platform and application layers.

B. Platform and Application Layer

The platform layer consists of operating systems and application frameworks. This layer simplifies application development for virtual machines by removing the need for hardware configuration. For example, Google App Engine operates at the platform layer by providing support for web applications by easily allowing users to read and write to data stores. The application layer consists of the cloud applications that users interact with. At the top of the hierarchy, the application layer is the most visible layer, acting as an interface to end-users on the cloud [8]. These layers are strongly interconnected and are essential to cloud computing systems.

IV. Cloud Computing Models

There are three primary ways that users interface with cloud computing: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The key advantages of these service models are agility and convenience, providing on demand resources to users. With IaaS, users outsource their hardware demands to the cloud provider. This can be contrasted with traditional hardware ownership where all aspects of hardware operation were

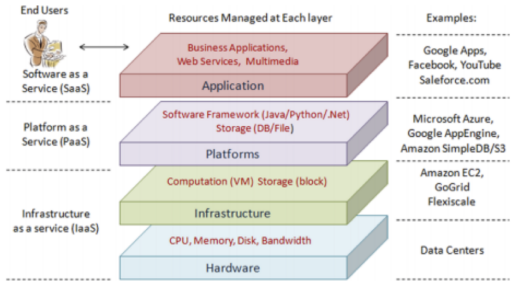

Fig. 1. Launching an app to open a URL

owned and controlled by the user. PaaS cloud providers bundle hardware with suites of tools for software development. Finally, with SaaS, the cloud provider delivers applications to users over the Internet, accelerating software distribution.

A. Infrastructure as a Service

IaaS providers offer storage and computing resources on a pay-per-use basis. Typically, IaaS providers also supply a range of services to accompany these infrastructure resources, such as resource monitoring, security, backup, and recovery. As a result, customers can lease the infrastructure rather than maintain their own local hardware. The advantage of faster and more cost-efficient operations makes IaaS an effective and widely used model.

IaaS provides users with scalable hardware resources, alleviating the problem of acquiring computation resources and leasing access to large scale computing power. This model is effective for workloads that are temporary or change unexpectedly [9]. This can be contrasted with non-cloud infrastructure in which a user has access to only as much hardware as they can support. For growing companies, IaaS removes the need to commit to upfront hardware costs and dynamically provisions resources as those hardware demands evolve.

In the IaaS market, relatively few providers have emerged due to the demands required to supply a reliable IaaS product. With IaaS, a customer can request computing resources and have their request fulfilled quickly and reliably. As a result, the IaaS provider must have a large amount of available hardware. The economic costs associated with hardware has resulted in a small number of providers in the IaaS market: Amazon EC2, Google Compute Engine, Microsoft Windows Azure, and IBM Blue Cloud [10].

B. Platform as a Service

PaaS providers offer more of the application stack than IaaS providers, delivering a suite of hardware and software tools needed for application development. PaaS is primarily designed to support the entire software development lifecycle.

Like IaaS, PaaS providers deliver infrastructure servers, storage, and networking but also middleware, development tools, and database management systems. PaaS supplies an environment that frees users from having to install local hardware and software. As a result, customers can focus on creating and running applications rather than constructing and maintaining the underlying systems that host these applications [9]. Many modern applications are hosted on PaaS models such as Amazon Web Services (AWS), Microsoft Azure, and Google App Engine [11].

PaaS provides key tools that improve agility and flexibility to the entire software development lifecycle: building, testing, deploying, managing, and updating. With a PaaS setup, cloud apps and services can share data generated and stored within the platform, allowing easy data integration. Solutions residing outside the platform can be easily configured to share information with the PaaS setup. This enables a company to connect any legacy software to the platform without heavily modifying their codebase [12]. In addition, the platform setup can support change by allowing users to continually produce new services quickly that can be easily built into the application. These tools are building blocks that empower users to quickly create unique and sustainable solutions.

C. Software as a Service

SaaS is a software distribution model in which a service provider hosts its applications for customers to use over the Internet. The key benefit of SaaS is ease of access for consumers, as all the underlying infrastructure, middleware, software, and data are managed by the service provider.

SaaS is typically licensed on a subscription basis and accessed via a browser, becoming an alternative to traditional on-premise software installations. Some popular examples of SaaS solutions are Outlook, Salesforce, and Netflix [9]. Customers can access these services, which are often computationally intense, without installing them locally. In addition, the SaaS model allows the user to ignore hardware and software management concerns, maintaining only client side hardware (i.e. laptop, cell phone, or other device). This high accessibility has enabled new kinds of software products that would be impossible to deploy locally.

SaaS gives many key advantages that make this an appealing software delivery model for businesses. SaaS enables sophisticated enterprise applications affordable for users that lack the resources to deploy and manage the infrastructure and software themselves. Since these solutions reside in scalable cloud environments, costs are reduced based on level of usage. In addition, SaaS vendors can rapidly update their solution on a centrally maintained code base, which becomes immediately available for their customers. This architecture enables quicker innovation, saving valuable time that would be spent on maintaining various versions of outdated code [9]. These advantages enable quick and convenient software distribution for businesses as well as painless application access for their customers.

D. Cloud Deployments

There are different ways in which a cloud can be deployed depending on which functionality users desire. Public and private clouds are the most prevalent deployments, but others

exist in order to cater to different criteria. The differences between these deployments are primarily ownership and the level of control given to the user.

In a public cloud, administrator access and control is relinquished to the cloud provider. This provides maximum convenience for the user, since they only need to specify their required resources. The cloud provider becomes responsible for all reliability and security concerns. The quality of service given by the leading cloud providers means that failures are rare, but the impact of security and system failures is widespread because it could affect many clients of a compromised cloud provider [13].

A private cloud is managed and hosted inside an organization's own data center, which leaves access choices to the organization itself. Private clouds keep resources behind local firewalls and leaves hardware management to the organization. This allows setting higher security standards, as organizations have direct control and resources can be isolated for security.

While a typical cloud deployment is either public or private, other forms exist, such as Hybrid and Community clouds. Hybrid clouds employ a conjunction of private and public services. Hybrid clouds provide ownership advantages of a private cloud with the higher scalability of a public cloud to accommodate rapid computation changes. Community clouds are owned by the users, like a private cloud, but shared across multiple users with common interests. This allows the same benefits of a private cloud, but enables a larger group of users to take advantage of the enhanced security and split the cost of setting up their cloud.

The availability of different cloud deployments creates a spectrum of options for a cloud user. For example, one can use a private cloud to maximize security or a public cloud for maximum convenience. The variety of cloud models and deployments creates a range of options that lends itself to diverse business models built around different types of cloud computing.

V. Specialization

The accessibility of Cloud computing via tools like AWS and Google Cloud Platform allows consumers and companies to quickly purchase and use resources. This allows for smaller companies like startups to develop software products that require intensive computational and storage power without worrying about the overhead of maintaining their own internal server [10].

Specialization is beneficial in the context of cloud computing the problems that require cloud services are best solved when different actors focus on different parts of the solution. Machine learning problems necessitate large storage of training data and massive computing power to create high-dimensional prediction models. For smaller companies that lack large internal servers, working on such machine learning problems is made possible by setting up a cloud service to offload computing requirements. Cloud providers can focus on making their cloud services a robust foundation for solving machine learning problems, and companies that use cloud services can focus on optimizing the product instead of spending time on the computing resources.

Smaller companies tend to focus on solving big data problems, while larger companies tend to focus on solving broad machine learning problems using data they generate from their large-scale proprietary services. Startups and mid-sized companies tackle problems at a local level for specific business needs in which they have extensive domain knowledge. This is only possible because cloud computing companies create a platform on which to solve these problems, while the companies that use this platform develop a deep understanding of their dataset and how to apply cloud services to solve their business problems. Large companies often have terabytes of user data from their widely used services. Difficult machine learning problems such as speech recognition, sentiment analysis, and video analysis require access to these types of massive datasets. This puts large companies in a unique position to solve these problems using their proprietary data.

VI. Evolution of Computing Devices

Cloud computing enables any device with an Internet connection access to large amounts of computing power and storage. Design of user devices has shifted its focus from increasing processing power to making devices lightweight and portable. These lightweight devices have become prevalent in daily life and constitute the Internet of Things. The ability to offload processing has redefined the role of devices as user interfaces to cloud services. As a result, many developers have focused their efforts on increasing the surface area of that interface, creating devices that give users new ways of accessing their cloud services.

One consequence of this trend has been increased growth in the Internet of Things (IoT) by allowing small devices to offload tasks to the cloud. A barrier to having many lightweight devices perform "smart" functions in the past has been the lack of onboard processing power. With the development of the cloud model, a device can have a small sensor and feed that data to the cloud where the processing will take place. The offloading of processing prolongs device battery life and reduces the cost of connected devices, allowing technology to play a larger role in appliances [14].

The emergence of IoT has led to a shift in how devices are used to consume information. Mobile devices have become pervasive, always staying connected to the Internet [15]. An example of this is the Amazon Echo which users interact with daily. While IoT devices take different physical forms, these smart devices all function as different entry points for users to seamlessly connect to Cloud services, interacting with them as if they were tangible parts of everyday life.

VII. Drawbacks

There are issues that hinder widespread adoption of the cloud. Many enterprises are reluctant to deploy their businesses to the cloud due to concerns about data ownership and trade offs between security and privacy. Additionally, cloud

computing has considerable environmental implications that are hidden from the typical user.

A. Ownership and Control

Cloud computing involves storing and accessing user information on remote servers operated by others through the Internet, which raises ownership concerns. The data owners must relinquish all decisions about infrastructure, security, and rights to the cloud provider. As a result, users lose control over their data, and owner rights cannot be guaranteed. For example, the Stored Communications Act gives the government the right to seize data stored by an American company, even if it is hosted elsewhere [13]. As a result, data ownership in the cloud is not always retained and can be ambiguous due to both government and company policies.

B. Trade-Offs

Just as in a normal computer, the cloud computing model introduces trade-offs when it comes to security and accessibility. In addition, the environment in which cloud computing operates creates opportunities for security flaws as well as data loss. Data encryption is an example of this: stronger encryption schemes make data availability less efficient, but weaker encryptions create more opportunities for security breaches [16]. Compromises between efficiency, cost, and security are a important factors to be balanced to optimize security and user privacy.

In cloud computing, independent customers often share physical infrastructure. This allows for reduced costs, but poses several inherent security flaws. Virtual machines co-existing on the same physical machine can give rise to threats such as Cross-VM attacks, which exploits shared cache memory between VMs [17]. Loss of physical control of data center leaves customers unable to directly prevent attacks and accidents, as well as unwanted data modification and loss.

C. Environmental Concerns

Centralization of computing power in the cloud model has allowed user devices to use less energy but has increased server energy demands. The transition of power hungry devices into one location has magnified heating problems and created a large, yet mostly overlooked, environmental impact.

The servers that enable cloud computing generate large amounts of heat due to their high performance processors and large user demand [18]. The high performance CPUs are able to run faster to meet the needs of the clients, but in the process generate more heat than a normal phone or laptop processor. In addition, they are usually kept at a higher utilization rate than any user processor which results in wasted energy and more heat.

Cloud providers must not only provide power to these higher performance processors, but also must cool the servers so that they function properly. The combined factors of excessive heat generation and a requirement to cool the operating environment make running servers very expensive in comparison to spreading out the processing load to devices.

VIII. Conclusion

While all cloud computing services aim at giving their users distributed access to large amounts of hardware and storage, there are a multitude of implementations that share this goal. There are also many implementations of these ideas, including the public, private, and hybrid clouds. All of these service models share their use of virtualization to abstract the user from the particular hardware and location they are accessing at any given time. Regardless of the type of service, cloud computing has transformed the way users consume their data, giving users access to large amounts of computing power from lightweight devices.

Cloud computing, through the models of IaaS, PaaS, and SaaS, fulfills a variety of computing needs for different kinds of users. Cloud technology has shifted the design of devices towards a lightweight user interface, where infrastructure and software demands are offloaded to the Internet rather than locally. This shift has prompted specialization in the market and has accelerated the growth of the Internet of Things. Still, there are several considerations regarding data ownership, as well as security and privacy trade offs, that must be considered. Beyond the impact to the user, cloud computing also generates an often overlooked environmental impact. Despite these drawbacks, cloud computing has transformed how the IT industry develops and deploys software and how users interact with technology.

IX. References

[1] T. Grance, P. Mell, "The NIST Definition of Cloud Computing", pp. 800-145, Sep. 2011 [Online].

[2] D. Kapil et al., "Cloud Computing: Overview and Research Issues," in Int. Conf. on Green Informatics, Aug. 2017 [Online].

[3] L. Malhotra et al., "Virtualization in Cloud Computing," in Journal of Information & Technology Software Engineering, vol. 4, 2014 [Online].

[4] Q. Zhang et al., "Cloud computing: state-of-the-art and research challenges", in Journal of Internet Services and Applications, vol. 1, pp 7-18, May 2010 [Online].

[5] S. Singh and I. Chana, "Cloud resource provisioning: survey, status and future research directions", in Knowledge and Information Systems, vol. 49, pp. 1005-1069, Dec. 2016 [Online].

[6] C. Lu and S. Tseng, "Integrated design of AES (Advanced Encryption Standard) encrypter and decrypter", in Application-Specific Systems, Architectures and Processors, 2002 [Online].

[7] E. Bresson, et al, "Provably authenticated group Diffie-Hellman key exchange", in Proc. of the 8th ACM conf.on Computer and Communications Security of ACM CCS, pp. 255-264, 2001 [Online].

[8] F. F. Moghaddam et al., "Cloud computing: Vision, architecture and Characteristics," in IEEE 6th Control and System Graduate Research Colloq., Aug. 2018 [Online].

[9] D. Kapil et al., "Cloud Computing: Overview and Research Issues," in Int. Conf. on Green Informatics, Aug. 2017.

[10] K. C. Haug et al., "Cloud adaptiveness within industry sectors Measurement and observations", in Telecommunications Policy, vol. 40, pp. 291-360, Apr. 2016.

[11] Q. Zhang et al.,"Cloud computing: state-of-the-art and research challenges", in Journal of Internet Services and Applications, vol. 1, pp 7-18, May 2010.

[12] F. F. Moghaddam et al., "Cloud computing: Vision, architecture and Characteristics," in IEEE 6th Control and System Graduate Research Colloq., Aug. 2018.

[13] S. Shubashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," in Journal of Network and Computer Applications, vol. 34, pp. 1-11, Jan. 2011.

[14] A. R. Biswas and R. Giaffreda, "IoT and cloud convergence: Opportunities and challenges," 2014 IEEE World Forum on Internet of Things (WF-IoT), pp. 375-376, 2014.

[15] Zambonelli, Franco, et al. "Algorithmic Governance in Smart Cities: The Conundrum and the Potential of Pervasive Computing Solutions." in IEEE Technology and Society Magazine, vol. 37, pp. 80-87, 2018.

[16] F. Shaar, "DDoS Attacks and Impacts on Various Cloud Computing Components," in Int. Journal of Information Security Science, vol. 7, pp. 26-48, Mar. 2018.

[17] Z. Xiao and Y. Xiao, "Security and Privacy in Cloud Computing", in IEEE Communications Surveys & Tutorials, 2013.

[18] R. K. Sharma et al., "Balance of power: dynamic thermal management for Internet data centers," in IEEE Internet Computing, vol. 9, no. 1, pp. 42-49, Feb. 2005.