Current Development, Challenges and Future Trends in Cloud Computing: A Survey

Hazzaa N. Alshareef
College of Computing and Informatics, Saudi Electronic University, Riyadh, Kingdom of Saudi Arabia

***Abstract*—Cloud computing is a new paradigm in information and communication technologies (ICTs) that provides the ability to access shared pools of different computing resources that are related to many cloud users within a pay-per-use or on-demand approach. It has transformed the delivery model of ICT from a product to a service. This provides several different advantages for institutions, companies and users based on savings and reduced capital expenditure through lower operating expenses. This paper provides a comprehensive survey of cloud computing. It first develops an understanding of cloud computing in general and discusses its advantages, current development, challenges and future trends. Subsequently, a detailed discussion on the cloud computing architectures, services models, fault tolerance mechanisms, services selection methods, adoption by industry, and scheduling of cloud-based resources is also presented. Nonetheless, cloud computing has many obstacles which expose it to a number of limitations. Some of these challenges include security of data, fault tolerance, and load balancing. A number of techniques in literature are proposed to cope with these challenges which are discussed and analyzed. Experimental data and usage drift validates the popularity of cloud computing and its adoption in recent years. Future trends in cloud computing support the use of intelligent machine learning (ML) techniques and new technologies to cope with some of the challenges and making cloud computing more efficient, secure and commercially viable to be widely accepted.**

Keywords—Cloud computing; security challenges; machine learning; resource scheduling; information and communication technologies

I. INTRODUCTION

In the present digital age, computer systems and associated applications have become inextricable part of life. Concomitantly, the need for better, cheaper, more efficient and on demand application services and infrastructure is felt like never before. Cloud computing is an approach that provides on-demand access to a shared pool of customizable computing resources (e.g. applications, networks, storage, servers etc.) and services [1]. Service providers can disseminate these resources with only marginal interaction and little management effort. Obtaining dynamic computing resources within the cloud computing paradigm provides the ability to cooperate with and scale up/down the given services, taking the demands of clients into account as well as the cost of the leveraged resources. This effectively contributes to a decrease in the operational cost pertaining to IT services. The scalability of cloud services provides smaller businesses with the ability to take advantage of various state of the art expensive and computing-intensive facilities that were previously affordable by large companies only [2].

Cloud computing enables the provision of information services and network computing resources, such as applications, servers, and storage [3], over the internet without installing them or purchasing them on their own. In 2005, Intel, IBM, and various other enterprises (including universities) within the United States began to operate a cloud computing virtual laboratory enterprise. This type of enterprise began with several experiments at North Carolina State University, situated near the IBM headquarters. In 2007, Google and IBM cooperated to start the processing of a new network computing approach, called cloud computing [4]. The new conversional Intel and Microsoft computing method was tested and thereafter caught the attention of a considerable number of research organizations.

In terms of virtualization, computers, networks, storage, and databases can all be potential cloud computing resources according to certain rules and service agreements. Global giants in the IT industry such as Google, IBM, Amazon, Microsoft, Alibaba etc. are investing in advanced research and innovative ways of utilizing cloud computing most effectively and widely. After launching a cloud computing platform, a significant issue is demonstrating the operative distribution and management pertaining to the virtual sharing of resources based on user demand by improving the effectiveness of the resource usage.

Technology advances and market forces are two related and integrated factors that drive interest in cloud computing. Rapidly improving business cases are producing enhancements in computing infrastructure, which has motivated several enterprise applications and services to consider moving to the cloud. In terms of technology, the existence of lower-cost processors and lower-latency networks, integrated with significant progress in virtualization, has moved computation from local IT platforms to disseminated cloud infrastructures. Evidence shows that, although cloud computing is considered a main business path for the upcoming years, moving to the cloud paradigm is seen to encounter a range of challenges. Some of the prominent ones among these are the issues related to security of the cloud data, scheduling of resources, fault tolerance, load sharing and load balancing. For instance, financial institutions are motivated to shift on cloud computing due to many advantages. Howbeit, the intrinsic security issues and challenges related to resource acquisition for smooth services are still hindering the complete migration. In 2014, more than 50 million users' Dropbox accounts were hacked, which resulted in a wide trust-deficit in the security of cloud computing. When cloud computing is considered a viable alternative, it should offer a similar security level to that of the conventional systems. In order to fulfil this goal, a comprehensive awareness regarding attacks and their countermeasures is required in order to detect malicious activities [5].

In summation, the cloud computing landscape has been considerably improved. Not only have further service offerings and providers packed the space, but improvements have also been made in the infrastructure and services. Cloud computing is considered to form an epitome that provides the ability to access shared pools of different computing resources in a pay-per-use or on-demand manner. With its incorporation into conventional systems, the weaknesses of traditional servers are overwhelmed in terms of efficiency, speed and scalability. It also offers savings in capital expenditure through reduced operating expenses. However, a few obstacles still exist that impose limitations when the technology is used. Lack of security, data consolidation [6], load balancing [7] and fault tolerance [8] represent some of the most important restraints of cloud technology. In addition to these, the adoption and absorption of cloud computing into existing systems, especially in small and medium enterprises (SMEs) [9], poses great challenges in terms of system conversion, change acceptance and embracing a myriad range of accompanying technologies which were previously unavailable. This paper provides a review of the cloud computing paradigm and its main computing and implementation options. It also discusses benefits and challenges of moving to the cloud from users' and companies' perspectives. Many of the aforementioned challenges are discussed in detail and a number of suitable ways are presented, based on literature, to cope up with these challenges. It summarizes the proposed solutions and techniques to deal with cloud computing risks and limitations, and discusses how future technologies, such as artificial intelligence [10], [11] and block chain [12], can embark a new era of prevalence of this technology. Gill et al. [10] has used the concept of "Triumvirate: IoT + AI + Blockchain" to describe the influence and interdependence of AI, IoT and Blockchain technologies that are anticipated to shape the future. They have also highlighted that future companies need to be well informed using Big Data Analytics and Data Science techniques to understand market trends, customer preferences and correlations.

The remainder of the paper is organized as follows: Section II provides a background and overview of cloud computing, including a definition, its structural layers, as well as service and delivery models. Section III then discusses benefits of moving to cloud computing. Challenges pertaining to cloud computing and the ways to cope with them are discussed in Section IV. Section V highlights a number of research trends and directions in cloud computing, whereas Section VI provides a discussion on cloud research trends. Section VII gives a conclusion of this work.

II. Cloud Computing Overview

The term 'cloud', or 'fog' in cloud computing signifies the existence of a remote virtual space. This section presents the definition, architecture, service models, delivery models, and the main characteristics of cloud computing.

A. Definition

There is, as yet, no standard definition of cloud computing. This is due to the dynamic nature of the term and its vast area of application. Nonetheless, industry and academic players are making essential strides towards agreeing on a standard definition ([13], [14], [15], and [16]). For example, as declared by the United States National Institute for Standards and Technology (NIST) [2], "cloud computing is 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" .

Madhavaiah and Bashir [17] analysed a number of definitions from both a business and research perspective and proposed a comprehensive definition: "Cloud computing is an information technology-based business model, provided as a service over the Internet, where both hardware and software computing services are delivered on-demand to customers in a self-service fashion, independent of device and location within high levels of quality, in a dynamically scalable, rapidly provisioned, shared and virtualized way and with minimal service provider interaction". Nonetheless, cloud computing is still an emerging technology with its ability to integrate and be integrated in new and associated technologies, which would significantly impact and form its true definition over a period of time.

B. Structural Layers

Majority of the researchers agree that the structure of cloud computing is based on four layers [1], [13]. The first layer is the Application Layer, which is located at the top of the architecture and is the layer that is the most visible for end-users. It consists of different applications and software packages related to the real cloud. As an example, office tools, storage management applications, email systems, and virus scanning and removal applications. The second layer is the Platform Layer, which provides the programming-level interface in accordance with different application approaches and operating systems (OS). This provision attempts to simplify the deployment of an application in the environment of the cloud. The third layer is the Infrastructure Layer, which is based on dynamically employing virtualization techniques for assigning the storage and computing resources needed. A well-known example of virtualization is VMware. The fourth layer is the Hardware Layer, which creates the data centers that contain various physical components, such as, cooling infrastructure, electrical power components, switches, routers, and servers. Fig. 1 shows a typical cloud computing ecosystem, whereby a number of hardware components and applications make the cloud, and network nodes comprising of different devices are the end users of cloud infrastructure and services.

C. Service Models

Cloud computing architecture represents a service-oriented approach, in which services are provided by every layer through to the top one. Accordingly, the services that are obtained in the cloud computing paradigm are classified based on three different types: the Software as a Service (SaaS) model, the Platform as a Service (PaaS) model, and the Infrastructure as a Service (IaaS) model.

The SaaS model is a cloud computing approach in which various applications remain on the service provider's cloud infrastructure and are provided to many users via web apps and interfaces. The major idea behind SaaS is based on eradicating the practice of applications residing locally on individual users' devices, where the computing power is insufficient to provide high computing performance and effectiveness to users [18]. The genesis of cloud computing can be traced to the SaaS approach [2]. Some examples of SaaS providers are Intercom [19], Trello[20], Hipchat, and Rackspace [21].

The PaaS model is a service approach for providing a platform that creates and operates different applications based on the programming interface that is obtained from and supported by the cloud provider [22]. Consequently, scalability issues, high server rapidity and storage capacities are all addressed under the PaaS approach. Therefore, PaaS users are able to create, operate and deliver their particular applications based on the use of remote IT platforms. Nonetheless, users are unable to monitor core cloud platforms (e.g., storage, OS or servers). An example of a PaaS provider is Microsoft Windows Azure [23].

The IaaS model is an approach whereby virtual infrastructures, such as virtual servers, storage and other fundamental computing resources, are offered by cloud service providers to users in order to enable them to disseminate and operate their particular applications or OS; and to download or upload files or software on the cloud. Using the IaaS approach, users can monitor their software, including applications, which is disseminated throughout the cloud. Nevertheless, such users have a limited ability to monitor the virtual infrastructure that is obtained from the cloud service provider. An example of an IaaS provider is Amazon EC2 [24].

D. Delivery Models

Cloud based services are disseminated to users via four main delivery models, based on, (i) the control required, (ii) the number of users, and (iii) privacy and security demands of the users [2], [25], [26]. These delivery models are categorized into: (i) Public cloud, (ii) Private cloud, (iii) Community cloud, and (iv) Hybrid cloud.

The Public cloud consists of a third party that possesses the physical resources in their entirety and delivers different cloud services to users via the Internet. Users who are supported by the cloud provider range from individuals to corporate institutions.

The Private cloud is provided exclusively to a particular institution. This model can assist in monitoring the performance of a system, security guidelines, and data. An institution is also able to disseminate its own particular cloud services, and a third-party institution can handle the model by itself.

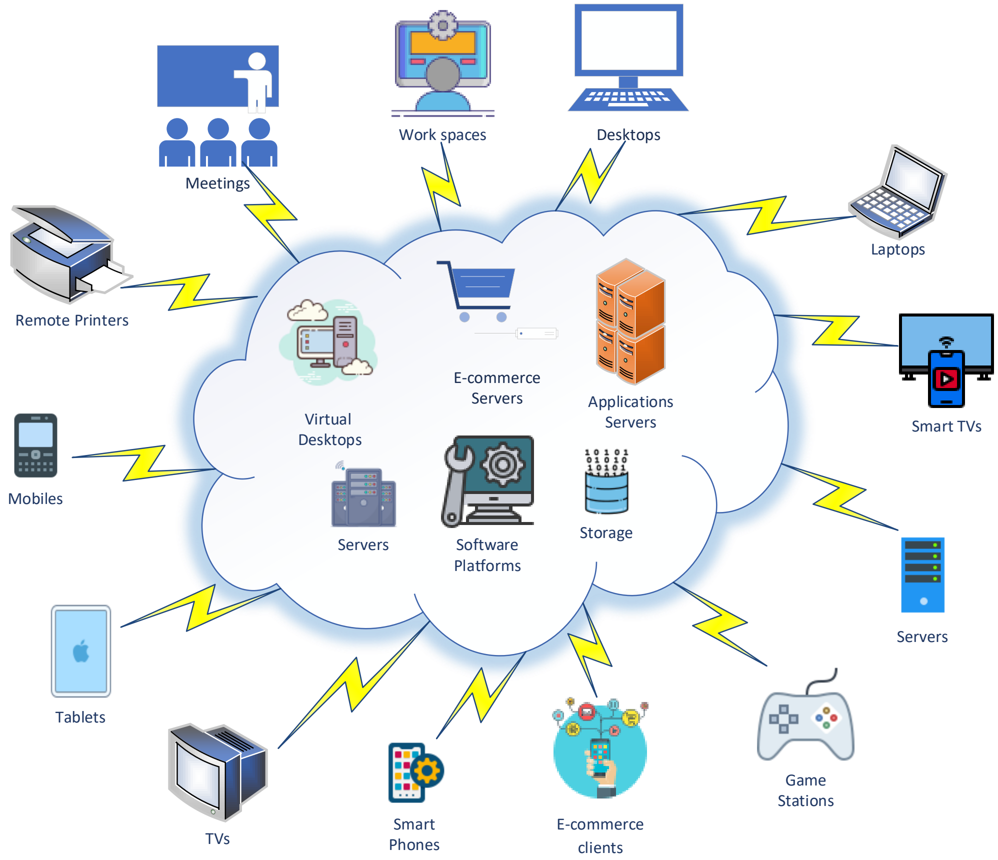

Fig. 1. Cloud computing technology ecosystem.

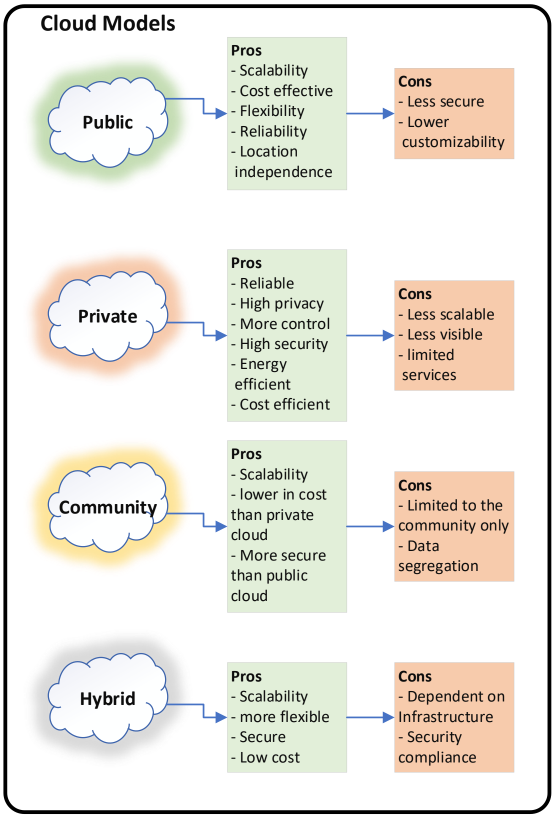

Fig. 2. Cloud delivery models.

The Community cloud represents the provision of different cloud services based on a particular group of institutions with the same mission, compliance conditions, policies, and security demands [22]. A community cloud demonstrates a generalization of the private cloud and, therefore, includes further institutions in each realization.

The merger and combinations of various cloud models (e.g., community, public and private cloud models) represents a hybrid cloud model. They have become very popular in recent years primarily due to popularity and wide usage of cloud services, which now faces complex dynamics of the corporate infrastructures and new business markets. Despite these various models being grouped with each other, they remain distinct and are included in exclusive standards and have distinctive standards and technology with respect to data operability and various applications. The Hybrid cloud inherits the advantages and disadvantages of community, public and private clouds. Consequently, it represents an optimal approach that makes a delicate balance between price and control, which are strong considerations for economic viability as well as user satisfaction with cloud services and applications. Fig. 2 demonstrates the advantages and disadvantages of the cloud delivery models.

E. Cloud Computing Characteristics

Cloud computing is a relatively recent term and the technology has emerged from the usage and trends of computer networks and its associated services and business models. In studies [14], [27]–[29], a number of cloud computing characteristics are discussed. More characteristics may emerge as the technology grows. We present some of the essential characteristics of cloud computing in this section, which define and advocate the core technology and its acceptance specifications.

1) Dynamic (Flexible): cloud resource platforms are dynamically scalable, meaning that, they are able to be enlarged or reduced in size based on user demand, which minimizes the investment risk related to the user and can satisfy the demands of many users. Cloud computing provides users with the sense that infinite computing resources are available to them.

2) Virtualization: cloud computing applications and platforms are created according to resource virtualization concept. Virtualization performs a significant task in developing the effectiveness of resource efficiency and raising the level of service security and reliability.

3) Economies of scale: cloud computing is dominated by large companies, such as IBM, Microsoft, Google, and Amazon, which have the ability to employ large-scale resources that enable them to minimize rental and exploitation costs. As a result, cloud computing companies can recruit as many users as possible. Since there is a large number of potential users involved in any cloud-based service, it becomes financially viable for the service provider to offer the service at a very low cost.

4) On-Demand service: cloud services and platforms are obtained and billed based on users' actual demands. Cloud computing eliminates the risk of a one-time large investment and permits users to acquire only the resources they need. Accordingly, services depend on short-term costs (e.g., on an hourly basis), whereby users release resources once they are no longer required.

5) Dynamic customization: cloud rental resources should be customizable to a considerable extent. For example, in the IaaS delivery model, users are permitted to disseminate virtual and specialized devices. Further services deliver low flexibility and are not applied to general purpose computing. However, it is expected that such services will continue to provide a particular level of customization.

6) High reliability: cloud computing platforms are required to ensure that customer data are secure, so that the application platform is seen to be reliable. In general, platform backups and multiple data are both applied in order to raise platform reliability. Dynamic network management approaches are also applied by cloud computing platforms in order to verify the effectiveness and status of every resource node. The reason for this is that nodes could be dynamically migrated when failure or low effectiveness is encountered. Another reason is to ensure that the performance of the entire system is unaffected in case of a fault.

III. Benefits of Moving to the Cloud

Owing to the popularity of cloud computing in recent years, the technology incorporates a number of benefits which ascertain the next level of networks and applications sharing, and distribution of services and resources on an economically viable and efficient manner. A number of previous studies [9], [30]–[33] have highlighted some of these advantages. Cloud computing enables on-demand network access and a host of associated applications and infrastructure to a number of customizable computing resources such as servers, software applications, storage spaces, services and other networks. In line with the characteristics identified in the previous section, this section highlights the main benefits and challenges of shifting to the cloud, which mainly include:

A. *Optimum Resource Utilization*

Since most of the cloud computing is based on use-per-pay model, therefore resources are released after every use. This results in the overall utilization of all the computing resources in an efficient and optimized manner, leading to green computing. It is in line with the United Nations' (UN) Sustainable Development Goal (SDG) number 12 [34] that relates to responsible consumption and production of goods and services.

B. *Rent on Demand*

Another major benefit of cloud computing is the availability of all types of computing and infrastructure resources and services for everyone, anywhere, and at any time. Industry 4.0 [35], [36] provides opportunities for growth and sustainability for all kinds of businesses, whether large or small. Future businesses lie in innovation which is the key for success for agile companies of today. Being able to access and utilize state of the art infrastructure and computing resources, as and when needed, startups can easily compete with giants of the industry resulting in better and cheaper products and services for the masses.

C. *Minimized IT Staff*

Moving to cloud computing results in reduction of inhouse IT staff to maintain the existing systems. Some technical staff is, however, still required to work with pre-existing vendors, including specialized vendors, in order to manage particular outsourced applications.

D. *Minimized Infrastructure*

Relocating resources to the cloud, or accessing platform as a service, means that it is possible to maintain a smaller inhouse hardware infrastructure.

E. *Managed Costs*

Since most of the cloud computing service providers gain income on the basis of economies-of-scale, and try their best to cut down the costs, increase the customers and have the latest and updated hardware and software applications. Therefore, prices and licensing can be minimized when adopting cloud computing. The latest costs are based on predetermined services and is derived from the costs model used by the vendor.

F. *Enhanced Vulnerability Control*

Vulnerability control is the ability to track system activities and logs to provide greater control and minimize the risk of attacks by detecting and preventing its occurrence before happening. In the cloud computing ecosystem, the service provider provides services to numerous customers concomitantly and a large revenue is generated as a result. Therefore, the service providers make sure that all the systems are up to date with state-of-the-art technologies in place, and no vulnerabilities or security threats are there.

IV. Cloud Computing Challenges

A number of benefits related to cloud computing are discussed in the previous section. However, it is neither an optimal solution nor risk free, particularly when data is out of users' reach. Other disadvantages of moving to the cloud are reliability issues and system performance, since users are fully dependent on cloud resources. For instance, when accessing the cloud in order to seek a service, the time needed to execute the task (the round-trip time, or RTT) could be an issue for users. This can be exacerbated if the cloud is busy serving other instances or traffic is already congested. Privacy and security are other limitations that are widely known to render cloud computing which is a challenge for users. Some of the major challenges faced by cloud computing are provided as follows:

A. *Reliability*

A cloud computing platform must guarantee the reliability of the application platform and the integrity of customer data. When a large-scale system is experienced, an effective solution is expected in order to receive a high level of reliability. In addition, a dynamic network management system controls the effectiveness and status of the resource nodes, whereby ineffective or failed nodes are dynamically migrated. Consequently, the entire performance of the system is not influenced by these nodes. Ensuring that all systems are working perfectly and reliably poses continuous challenges for the cloud service providers.

B. *Resource Provisioning and Scheduling*

The dynamic deprecation and expansion of resources relies on users' demands, thus presenting new challenges for management systems and cloud platforms. In terms of provisioning of cloud resources, an effective cloud resource provisioning algorithm is highly needed that makes better resource utilization and allocation, reduces response time, and has robustness as well as fault tolerance capabilities. Similarly, scheduling for on-demand resource requirements, or long-term resource reservation is a challenging task, especially when the number of resources and the number of users are extremely large.

C. *Management Issues*

The process of managing a cloud computing platform is extremely complicated. Especially, resource consolidation is one of the key areas of research and have gained substantial attention from the research community in recent years. It involves managing the means of controlling the system's resources effectively, deploying and scheduling different resources in a dynamic manner, and managing clients, their billing systems and service agreements. Nonetheless, applying the approach of having one service provider creates obstacles, such as the following: (i) a lot of much energy is exploited via an enormous data centre in order to have it operational; (ii) centralized cloud data centres are affected by many single point failures; and (iii) data centres are geographically remote from their users, and data need to be moved from their source in order to be processed. This implies that personal or sensitive data generated through the use of different applications are kept in a location other than where they were produced.

D. Fault Tolerance

Fault tolerance refers to the continuity of cloud services even in the existence of any hardware or software malfunction. In case of failure of such components, it is a major challenge to keep all the system running and without performance degradation in presence of a fault.

E. Privacy and Transparency

Privacy and transparency of users' data and cloud services respectively is very crucial in any cloud computing system. In order to gain trust in a cloud-based service, where the users' data and all related credentials are stored in a virtual environment, privacy of data is very crucial. Similarly, transparency of cloud services and virtualization of all systems and infrastructure is of prime importance. Cloud service providers must inform the customers about how their data will be held, stored and transmitted. Things like what security and privacy schemes are deployed, and what internal policies and technologies are in place, are very important from the point of view of a client, especially, when the client is a big organization.

F. Security

Security of cloud computing is far most the biggest challenge in this technology. It deals with all kinds of challenges related to data security, information security, data integrity and confidentiality. Security is one of the main challenges and hurdles when cloud computing implementation and adoption is concerned. Therefore, a detailed discussion on it and the related issues is inevitable for the completeness of this work.

The following section provides a detailed discussion on works related to security challenges, including possible threats, attacks and their countermeasures.

V. Security Challenges and Cloud Computing

One of the potential hindrances in cloud computing adoption is that users are not informed of the physical location of their sensitive data. Since service providers locate cloud data centres in many geographical locations, it leads to a range of security issues and risks. The conventional security approaches, such as intrusion detection systems (IDSs), host-based antivirus software and firewalls, are not able to provide appropriate security through virtualized systems. Fast dissemination of risks derived from the virtualized environments produces different risks [37]. Subramanian and Jeyaraj [37] identify the top 12 threats to the cloud according to the Cloud Security Alliance (CSA), which include compromised credentials and broken authentication, and denial-of-service (DoS) attacks, as presented in Table I. Most of the threats identified are related to data breaching, representing the principal security problem that needs to be addressed.

Kamara and Lauter [38] indicate that many different risks emerge depending on the use of public clouds. Data integrity and confidentiality represent the highest risks and produce different but related issues. The authors propose a crypto-cloud architecture that contains three main features: the cloud storage service provider (CSSP), the consumer of the data, and the data authority (the user that possesses the related data). Encrypted files are uploaded by the data authority and the CSSP permits access to the files. The demanded file is then downloaded and decrypted based on the use of suitable credentials and tokens. This type of architecture faces various security issues at the service-level agreement (SLA), computation and communication levels. For instance, issues occur within the communication level because the same infrastructures and resources are shared through a virtual machine (VM), which increases the possibility of attacks.

TABLE I. Cloud Security Alliance's Top 12 threats

Threat Number	Threat Name
1	Compromised credentials and broken authentication
2	Malicious insiders
3	Denial-of-Service (DoS) attacks
4	Account hijacking
5	Inadequate diligence
6	Permanent data loss
7	The Advanced Persistent Threat (APT) parasite
8	Cloud service abuses
9	Data breaches
10	Exploited system vulnerabilities
11	Hacked interface and App. Program Interfaces
12	Shared technology, shared dangers

Bhadauria and Sanyal [39] categorize these problems into (i) network, (ii) application, and (iii) host levels. Attacks are determined based on the communication levels indicated. The main security risks at the network level are data integrity and confidentiality, where problems related to network security include reused IP addresses, sniffer attacks, Domain Name System (DNS) attacks, and prefix hijacking in the Border Gateway Protocol (BGP). At the application level, security is required in order to prevent attackers gaining control and compromising the application. Problems at this level include hacking, dictionary attacks, hidden field manipulation, CAPTCHA breaking, cookie poisoning and distributed DoS (DDoS) problems. Finally, the main host-level risks are foot printing, Trojan horses, unauthorized access, DoS, password cracking, profiling, worms, and viruses. Applying the aspect of cloud virtualization represents the largest of the computational-level issues.

Data is considered the key source of item related to any crypto-cloud approach. CSA regards a data breach as the highest security risk. Multi-tenancy and maintaining data storage on a remote place (i.e., out of your control) can cause data leakage. Chen and Zhao [40] produced a data life cycle: Generation => Transfer => Use => Share => Storage => Archival => Destruction, which requires protection during all the phases. Data-level security is categorized as data in rest and data in transit. Data in transit does not cause extra security threats in comparison with data in rest, as transmitting data can be performed based on a secured data transfer method. From the hacker's perspective, data in rest poses a high level of attraction.

As regards the cloud services are concerned, these are given by providers to consumers based on appropriate SLAs. Thus, major items related to the crypto cloud include the accountability of ensuring that SLAs are maintained. In practice, there is no exclusive standard for SLAs that are applied for all the requirements of security management. However, a few standards, such as the European Network and Information Security Agency (ENISA) and the European Commission Secure Provisioning of Cloud Services (SPECS), offer security by maintaining SLAs. Applying an SLA assists in obtaining an adequate level of quality of service.

Previous research has indicated that the above issues are encountered via three major attack vectors [5]: the network, the computing hardware, and the hypervisor (the computer software, firmware or hardware that creates and runs VMs). There also exist three kinds of attacker map of the three vectors: the cloud provider, internal users, and the external users. It is also possible for a cloud provider to act as an attacker. Permission and authority is given to employees working on the could, for example, might be exploited in order to steal sensitive user information based on either logical or physical manipulation of the hardware platform. External users can also have an impact on data integrity and confidentiality by interfering with communication channels or through lying dormant within the system in order to attack it later. Internal users, such as the owners of a VM instance, could use the hypervisor to attack other VM instances.

A. Major Cloud Attacks

The following subsection contains a discussion of the different attacks that have been discussed in previous research [41]–[44]. These types of attacks, which can be launched over a cloud infrastructure, are examined and presented along with the countermeasures that can be taken to control them.

1) Network-based attacks: The network represents a major vehicle of attack against applications that are being performed within a cloud platform. The majority of such attacks are closely related to the types of attacks typically recognized in conventional technology, although there are some network-based attacks that specifically relate to cloud computing.

2) Hardware-based attacks: Confidential data is protected from illegal access by being maintained within an encrypted form, interacting through different encrypted channels. However, data has to be decrypted sometimes for performing different computations from time to time. Attackers benefit from a multi-tenant environment in which they can simply access various physical resources (e.g., disk buses, memory buses, and instruction and data caches [L1, L2, L3]). Attackers explore and exploit decrypted data and the secret keys related to different common algorithms (e.g., RSA, DES, and AES) and VM instances.

3) Hypervisor-based attacks: The hypervisor is defined as the software layer which is located between the physical hardware and VMs in order to identify the fundamental architecture. The hypervisor is essential for ensuring the characteristics of cloud multi-tenancy. Moreover, it assigns several physical resources to guest VMs (e.g., peripherals, CPU and main memory). On the security side, hypervisors are the most important layer of protection within the cloud stack, since this is the highest privilege level. If attackers can gain control at the hypervisor level and compromise VM isolation, they can control any resource related to the host system.

B. Countermeasures against Cloud Attacks

Cloud providers apply well-known approaches to protect against network-based attacks (e.g., antivirus gateways, IDSs, and firewalls). Such approaches are currently being extensively disseminated within edge networks in order to protect end systems from various forms of attack and to control and check outgoing and incoming traffic. For hardware-based attacks, newly arising techniques (e.g., Arm TrustZone technology and Intel Software Guard Extensions [SGX]) could prevent side-channel attacks. Another prospect is based on protecting hardware using cryptography. A commonly agreed instruction by cloud providers is the Intel Advanced Encryption Standard New Instructions (AES-NI), which proceeds towards different cache-based software side-channel attacks. In the case of hypervisor-based attacks, several software and hardware isolation mechanisms offer resources secure separation, whereby some hardware mechanisms, such as AES-NI, can have an impact on security within this level and are considered a part of hardware-based hypervisor protection.

Senyo et al. [18] also present a valuable tool for navigation, which could be applied by IT personnel in order to gain further insight into security threats that are based on the use of cloud computing. Personnel can then weigh the advantages and disadvantages of any improved resolutions.

VI. Discussion on Cloud Research Trends and Future Technologies

Cloud computing technology is said to be one of the biggest revenue generators for the software companies in recent years. Fig. 3 shows the spending in public cloud computing IaaS hardware and software worldwide in US billion dollars as depicted by Forbes [45]. It also shows that share of this spending by PaaS and SaaS/ The projected values up to 2026 shows that the spending in this technology will remain on the rise.

One key weakness of cloud technology is indicated by the low level of control over data that is disseminated to the cloud provider. This weakness is a major hazard, causing various problems, such as DoS, malicious insiders, and account or service traffic hijacking, which represents an essential impairment to massive cloud adoption. When users are not able to enter different physical systems, they must rely solely on the infrastructure provider in terms of addressing issues that are incurred with regard to data security. Previously, a capable method was Trusted Computing (TC), by which Trusted Platform Module (TPM) characteristics were applied to offer integrity for the software stack (i.e., the VM layer). The Intel Trusted Execution Technology (TXT) is one example of this method. The Intel SGX also represents a promising technique and is defined as an instruction set architecture (ISA) extension, which provides the ability to execute a number of different instructions within a secure memory location, known as the secure enclave. Accordingly, SGX allows users to apply effective security for their own data and applications without having to trust the cloud operator. A similar characteristic is provided by a further research domain, called homomorphic cryptography [46], which permits different calculations to be performed on encrypted data without acquiring a secret key or any decryption of the data.

Another research trend relies on the use of containers for the purpose of abstracting different applications based on the underlying OS, thus providing the ability for more rapid improvement and simpler deployment. Docker [47] represents a well-known container, whereby spreading the container term depends on the support provided by large providers, such as Amazon, Google, and OpenStack. Containers can also be utilized for improving user application security.

For institutions that require an increased level of data security (e.g., banks, the military, and trading and insurance companies), the requirements with regard to customer information security are very high. Ensuring data security in cloud computing is a common issue of concern for such institutions. Presently, service providers and researchers offer several different resolutions. Within the new application domain, there exist several security concerns that should be resolved. Although a few protection methods have reached a particular level of practical maturity, other related methods remain in their infancy and are inappropriate for dissemination to an operating setup. Several different methods exist for addressing vulnerabilities to data breaches and shared technology risks; however, they require further improvement. It has been seen that solutions to attacks related to the network-level have led to improvements. In contrast, approaches at the application level to addressing DoS risks and account or service traffic hijacking need to be improved, and several researchers are still investing time in these areas. There are also several tasks involving identifying new resolutions that would provide protection against the various hypervisor-based attacks encountered. Furthermore, applications attempt to leverage the infrastructure of the cloud based on utilizing heterogeneous resources derived from several providers.

The broad trend is to aim at using infrastructure from several providers and compared with conventional cloud offerings from single providers, disperse computing apart from resources. Subsequently, new computing approaches that aim at fulfilling market needs are evolving. In practice, many different security problems have been identified in relation to SLAs, computation, and communication. As referred to earlier, there have recently been large and significant security problems, presenting a range of opportunities for hackers to break service cryptosystems. Cloud computing has been demonstrated to be inadequate when it comes to security problems, and cloud service providers need to take into account that security should form an inevitable and important factor, and not be an afterthought.

Recently, a lot of effort is made by the machine learning (ML) community to scale up and enhance the existing data mining and ML algorithms to meet the challenges of handling large amounts of data [48], [49]. The work by Kim et al. [50] demonstrates a network threat detection and classification method based on ML, thus paving a path to the intelligent threat analysis technology. Similarly, the use of artificial intelligence (AI) and ML algorithms for the analysis of cloud services, for security, and for predictive and prescriptive analytics is very promising.

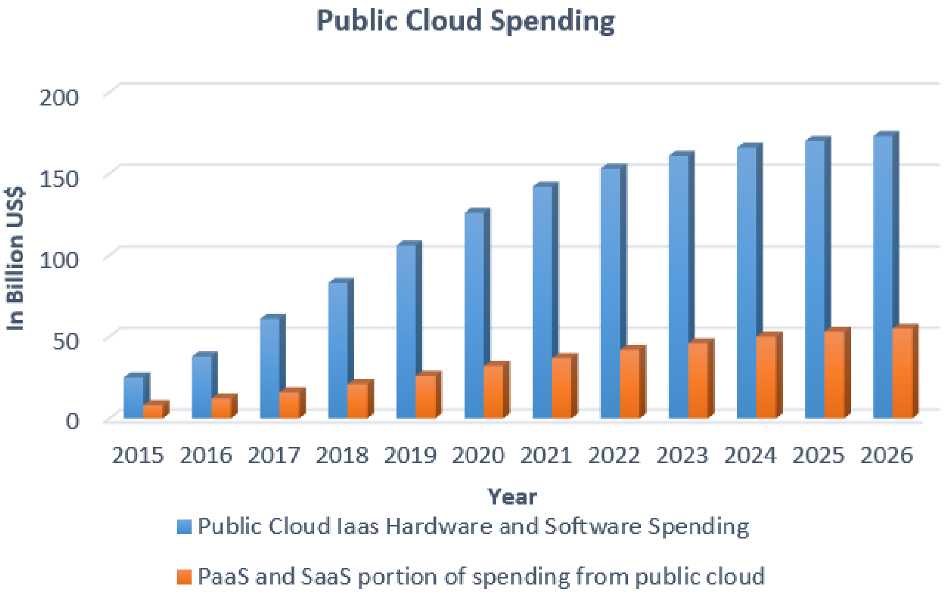

Fig. 3. Public cloud spending. Past, present and future.

VII. Conclusion

Cloud computing is an extremely rapidly developing technology in the domain of computing. There exist several benefits to applying cloud computing, such as anytime-anywhere accessibility, more effective geographical coverage, greater time efficiency, and reduced infrastructure costs. Nonetheless, there are also obstacles to applying cloud computing, such as lack of expertise and resources, cloud services management, privacy, and the need for data security. The majority of the services pertaining to the infrastructure of the hosting cloud, including storage and computing resources, exist in data centres. The hosting of applications in a single provider's cloud is seen to be simple and to offer various benefits. Nonetheless, there are a myriad of associated risks and challenges with cloud computing. Information privacy, security and data integrity are among the top of these. Trends and results from the literature shows that cloud computing is still emerging and new associated technologies are being developed to cope up with the existing challenges. The use of AI in cloud computing can mitigate some of the risks and provide solutions to previously unresolved issues.

Acknowledgment

The author highly appreciates the support provided by Saudi Electronic University for this research.

References

[1] A. Vafamehr and M. E. Khodayar, "Energy-aware cloud computing," Electr. J., vol. 31, no. 2, pp. 40–49, Mar. 2018, doi: 10.1016/j.tej.2018.01.009.

[2] "Final Version of NIST Cloud Computing Definition Published," NIST, Oct. 2011, doi: 10/final-version-nist-cloud-computing-definition-published.

[3] W. D. Tian and Y. D. Zhao, Optimized Cloud Resource Management and Scheduling: Theories and Practices, 1st edition. Waltham, MA: Morgan Kaufmann, 2014.

[4] K. D. Foote, "A Brief History of Cloud Computing," DATAVERSITY, Dec. 17, 2021. https://www.dataversity.net/brief-history-cloud-computing/ (accessed Jan. 04, 2023).

[5] L. Coppolino, S. D'Antonio, G. Mazzeo, and L. Romano, "Cloud security: Emerging threats and current solutions," Comput. Electr. Eng., vol. 59, pp. 126–140, Apr. 2017, doi: 10.1016/j.compeleceng.2016.03.004.

[6] L. Helali and M. N. Omri, "A survey of data center consolidation in cloud computing systems," Comput. Sci. Rev., vol. 39, p. 100366, Feb. 2021, doi: 10.1016/j.cosrev.2021.100366.

[7] V. Gehlot, D. S. P. Singh, and D. A. Saxena, "A Survey On Energy-Aware Load Balancing In Cloud Computing Environment," Int. J. Sci. Technol. Res., vol. 8, no. 12, pp. 4055–4060, Dec. 2019.

[8] P. Kumari and P. Kaur, "A survey of fault tolerance in cloud computing," J. King Saud Univ. - Comput. Inf. Sci., vol. 33, no. 10, pp. 1159–1176, Dec. 2021, doi: 10.1016/j.jksuci.2018.09.021.

[9] D. Widyastuti and I. Irwansyah, "Benefits And Challenges Of Cloud Computing Technology Adoption In Small And Medium Enterprises (SMEs)," Jan. 2018. doi: 10.2991/bcm-17.2018.46.

[10] S. S. Gill et al., "Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges," Internet Things, vol. 8, p. 100118, Dec. 2019, doi: 10.1016/j.iot.2019.100118.

[11] K. N. Qureshi, G. Jeon, and F. Piccialli, "Anomaly detection and trust authority in artificial intelligence and cloud computing," Comput. Netw., vol. 184, p. 107647, Jan. 2021, doi: 10.1016/j.comnet.2020.107647.

[12] M. R. Dorsala, V. N. Sastry, and S. Chapram, "Blockchain-based solutions for cloud computing: A survey," J. Netw. Comput. Appl., vol. 196, p. 103246, Dec. 2021, doi: 10.1016/j.jnca.2021.103246.

[13] M. Taghipour, E. Mowloodi, M. Mahboobi, and J. Abdi, "Application of Cloud Computing in System Management in Order to Control the Process," vol. 3, pp. 34–55, May 2020, doi: 10.31058/j.mana.2020.33003.

[14] "(PDF) Cloud Computing: A review of the Concepts and Deployment Models." https://www.researchgate.net/publication/317413701_Cloud_Computing_A_review_of_the_Concepts_and_Deployment_Models (accessed Jan. 03, 2023).

[15] S. Slimani, T. Hamrouni, and F. Ben Charrada, "Service-oriented replication strategies for improving quality-of-service in cloud computing: a survey," Clust. Comput., vol. 24, no. 1, pp. 361–392, Mar. 2021, doi: 10.1007/s10586-020-03108-z.

[16] P. T. Endo, M. Rodrigues, G. E. Gonçalves, J. Kelner, D. H. Sadok, and C. Curescu, "High availability in clouds: systematic review and research challenges," J. Cloud Comput., vol. 5, no. 1, p. 16, Oct. 2016, doi: 10.1186/s13677-016-0066-8.

[17] "Defining Cloud Computing in Business Perspective: A Review of Research - C. Madhavaiah, Irfan Bashir, Syed Irfan Shafi, 2012." https://journals.sagepub.com/doi/abs/10.1177/0972262912460153 (accessed Jan. 03, 2023).

[18] P. K. Senyo, E. Addae, and R. Boateng, "Cloud computing research: A review of research themes, frameworks, methods and future research directions," Int. J. Inf. Manag., vol. 38, no. 1, pp. 128–139, Feb. 2018, doi: 10.1016/j.ijinfomgt.2017.07.007.

[19] ["Making Internet Business Personal | Intercom." https://www.intercom.com (accessed Jan. 13, 2023).

[20] "Manage Your Team's Projects From Anywhere | Trello." https://trello.com/ (accessed Jan. 13, 2023).

[21] "Rackspace Technology | Multicloud Solutions Provider." https://www.rackspace.com/node/22215 (accessed Jan. 13, 2023).

[22] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi, "Cloud computing — The business perspective," Decis. Support Syst., vol. 51, no. 1, pp. 176–189, Apr. 2011, doi: 10.1016/j.dss.2010.12.006.

[23] "Cloud Computing Services | Microsoft Azure." https://azure.microsoft.com/en-us (accessed Jan. 13, 2023).

[24] "Cloud Computing Services - Amazon Web Services (AWS)," Amazon Web Services, Inc. https://aws.amazon.com/ (accessed Jan. 13, 2023).

[25] H. Mouratidis, S. Islam, C. Kalloniatis, and S. Gritzalis, "A framework to support selection of cloud providers based on security and privacy requirements," J. Syst. Softw., vol. 86, no. 9, pp. 2276–2293, Sep. 2013, doi: 10.1016/j.jss.2013.03.011.

[26] P.-F. Hsu, S. Ray, and Y.-Y. Li-Hsieh, "Examining cloud computing adoption intention, pricing mechanism, and deployment model," Int. J. Inf. Manag., vol. 34, no. 4, pp. 474–488, Aug. 2014, doi: 10.1016/j.ijinfomgt.2014.04.006.

[27] "The Role of Cloud Computing in the Development of Information Systems for SMEs," IBIMA Publishing. https://ibimapublishing.com/articles/JCC/2017/736545/ (accessed Jan. 03, 2023).

[28] R. O. Aburukba, M. AliKarrar, T. Landolsi, and K. El-Fakih, "Scheduling Internet of Things requests to minimize latency in hybrid Fog–Cloud computing," Future Gener. Comput. Syst., vol. 111, pp. 539–551, Oct. 2020, doi: 10.1016/j.future.2019.09.039.

[29] "An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing | EURASIP Journal on Wireless Communications and Networking | Full Text." https://jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-019-1557-3 (accessed Jan. 03, 2023).

[30] S. A. Bello et al., "Cloud computing in construction industry: Use cases, benefits and challenges," Autom. Constr., vol. 122, p. 103441, Feb. 2021, doi: 10.1016/j.autcon.2020.103441.

[31] T. Vasiljeva, S. Shaikhulina, and K. Kreslins, "Cloud Computing: Business Perspectives, Benefits and Challenges for Small and Medium

Enterprises (Case of Latvia),” Procedia Eng., vol. 178, pp. 443–451, Jan. 2017, doi: 10.1016/j.proeng.2017.01.087.

[32] A. Aljumah and T. A. Ahanger, “Cyber security threats, challenges and defence mechanisms in cloud computing,” IET Commun., vol. 14, no. 7, pp. 1185–1191, 2020, doi: 10.1049/iet-com.2019.0040.

[33] M. Humayun, “Role of Emerging IoT Big Data and Cloud Computing for Real Time Application,” Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 11, no. 4, Art. no. 4, Jun. 2020, doi: 10.14569/IJACSA.2020.0110466.

[34] “THE 17 GOALS | Sustainable Development.” https://sdgs.un.org/goals (accessed Jan. 20, 2023).

[35] M. Ghobakhloo, “Industry 4.0, digitization, and opportunities for sustainability,” J. Clean. Prod., vol. 252, p. 119869, Apr. 2020, doi: 10.1016/j.jclepro.2019.119869.

[36] T. Masood and P. Sonntag, “Industry 4.0: Adoption challenges and benefits for SMEs,” Comput. Ind., vol. 121, p. 103261, Oct. 2020, doi: 10.1016/j.compind.2020.103261.

[37] N. Subramanian and A. Jeyaraj, “Recent security challenges in cloud computing,” Comput. Electr. Eng., vol. 71, pp. 28–42, Oct. 2018, doi: 10.1016/j.compeleceng.2018.06.006.

[38] S. Kamara and K. Lauter, “Cryptographic Cloud Storage,” in Financial Cryptography and Data Security, Berlin, Heidelberg, 2010, pp. 136–149. doi: 10.1007/978-3-642-14992-4_13.

[39] R. Bhadauria and S. Sanyal, “Survey on Security Issues in Cloud Computing and Associated Mitigation Techniques,” Int. J. Comput. Appl., vol. 47, no. 18, pp. 47–66, Jun. 2012, doi: 10.5120/7292-0578.

[40] “Data Security and Privacy Protection Issues in Cloud Computing | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/abstract/document/6187862 (accessed Jan. 03, 2023).

[41] H. A. Kholidy, “Detecting impersonation attacks in cloud computing environments using a centric user profiling approach,” Future Gener. Comput. Syst., vol. 117, pp. 299–320, Apr. 2021, doi: 10.1016/j.future.2020.12.009.

[42] H. Abusaimeh, “Security Attacks in Cloud Computing and Corresponding Defending Mechanisms,” Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 3, pp. 4141–4148, Jun. 2020, doi: 10.30534/ijatcse/2020/243932020.

[43] B. T. Devi, S. Shitharth, and M. A. Jabbar, “An Appraisal over Intrusion Detection Systems in Cloud Computing Security Attacks,” in 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Mar. 2020, pp. 722–727. doi: 10.1109/ICIMIA48430.2020.9074924.

[44] A. Saeed, P. Garraghan, and S. A. Hussain, “Cross-VM Network Channel Attacks and Countermeasures Within Cloud Computing Environments,” IEEE Trans. Dependable Secure Comput., vol. 19, no. 3, pp. 1783–1794, May 2022, doi: 10.1109/TDSC.2020.3037022.

[45] “Current Cloud Computing Statistics Send Strong Signal of What’s Ahead,” Insight. https://www.insight.com/en_US/content-and-resources/2016/11032016-current-cloud-computing-statistics.html (accessed Jan. 27, 2023).

[46] S. Q. Ren et al., “Secure searching on cloud storage enhanced by homomorphic indexing,” Future Gener. Comput. Syst., vol. 65, pp. 102–110, Dec. 2016, doi: 10.1016/j.future.2016.03.013.

[47] “Docker: Accelerated, Containerized Application Development,” May 10, 2022. https://www.docker.com/ (accessed Jan. 24, 2023).

[48] D. Pop, “Machine Learning and Cloud Computing: Survey of Distributed and SaaS Solutions.” arXiv, Mar. 29, 2016. doi: 10.48550/arXiv.1603.08767.

[49] U. A. Butt et al., “A Review of Machine Learning Algorithms for Cloud Computing Security,” Electronics, vol. 9, no. 9, Art. no. 9, Sep. 2020, doi: 10.3390/electronics9091379.

[50] H. Kim, J. Kim, Y. Kim, I. Kim, and K. J. Kim, “Design of network threat detection and classification based on machine learning on cloud computing,” Clust. Comput., vol. 22, no. 1, pp. 2341–2350, Jan. 2019, doi: 10.1007/s10586-018-1841-8.