International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Systematic and Comprehensive Review of Literature on Security Threats, Mitigation Strategies and Optimization Techniques in Cloud Computing

Ajay N[1], Dr. Mohan H S[2], Dr. Shrihari M R[3], Dr. Vikas Reddy S[4], Santhosh Kumar M[5], Shwetha B V[6]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: This paper presents a survey evaluating known vulnerabilities, identifying threats, and discussing cloud security requirements. It delves into the significance of resource optimization and recent security advancements in the realm of cloud computing. The primary objective of this research is to explore various facets of cloud computing, with a specific focus on privacy and security concerns. It scrutinizes security vulnerabilities within cloud services while evaluating existing standard cyber security solutions. Furthermore, the study seeks to comprehend the security risks encountered by cloud users, data owners, and service providers. Additionally, it underscores the importance of optimizing cloud resources and review standard load balancing techniques. Various comparative analysis are drawn for thorough inspection of issues and solutions.

Keywords: *Cloud Computing, Load Balancing, Optimization, Cyber Security, Access Control*

1. Introduction

The Internet of Things (IoT) is rapidly growing, connecting more devices to the internet. By 2025, it's expected to reach 41.6 billion IoT devices, producing 79.4 ZettaBytes (ZB) of data. Cloud service providers play a crucial role in managing this massive amount of data. Moreover, cloud computing has garnered significant attention due to its affordability, sustainability, scalability, flexibility, and reliability [3], [4]. The fundamental concept of pay-per-use has attracted both businesses and individuals seeking to leverage this new revenue model [5–6]. A survey conducted in 2020 with 750 global cloud specialists revealed that organizations have increased their spending on cloud

services by 47% in 2021, driven by the impact of the COVID-19 pandemic [6]. Industries projected to experience the highest growth rates in cloud service adoption include IoT, machine learning/AI, data warehousing, and server-less computing, with an average growth rate of 47.2% [7].

At the heart of the accelerating modern economic system lies cloud computing, facilitating the seamless integration of the internet, big data, artificial intelligence, and the real economy. According to Gartner, Inc., the global market for public cloud services was grown by 17% in 2020, reaching around $266.4 billion, up from $227.8 billion in 2019 [8].

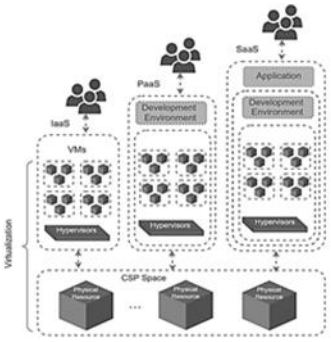

Fig 1. Service Oriented Cloud Computing

1.1. Overview of Cloud Computing

NSIT categorizes cloud security challenges into three main groups: characteristics, deployment models, and service-based paradigms.

1.1.1. Cloud Computing Enabling Technologies

The popularity of cloud computing has become a reality, thanks to the key technologies such as virtualization, multitenancy, and Service-Oriented Architecture (SOA) [9]. These techniques enable the sharing of user resources from a physical instance.

(a) Virtualization

Resource partitioning in cloud environments is made possible by virtualization, which establishes an abstract architecture for computers. By utilizing a virtual machine (VM), users can share resources using an image file, which they can create or obtain from external sources [10].

[1]*Assistant Professor,Dept. of CSE, SJCIT, Chickballapur*
Research Scholar, NHCE, Bangalore
Visvesvaraya Technological University, Belagavi, Karnataka, India
[2]*Professor & HoD, Dept. of CSE(DS), RNSIT, Bangalore*
Research Supervisor,NHCE, Bangalore
Visvesvaraya Technological University, Belagavi, Karnataka, India.
[3]*Associate Professor, Dept. of CSE, SJCIT, Chickballapur*
Visvesvaraya Technological University, Belagavi, Karnataka, India
[4]*Associate Professor & HoD, Dept. of AI&ML, SJCIT, Chickballapur*
Visvesvaraya Technological University, Belagavi, Karnataka, India
[5]*Assistant Professor, Dept. of CSD, SJCIT, Chickballapur*
Visvesvaraya Technological University, Belagavi, Karnataka, India,
[6]*Assistant Professor, Dept. of CSD, SJCIT, Chickballapur*
Visvesvaraya Technological University, Belagavi, Karnataka, India

Virtually any shareable IT resource can be virtualized in practice, enabling multiple users to access a single resource instance. Common forms of virtualization include desktop, network, storage, data, application, CPU, and cloud virtualization. Cloud virtualization encompasses models such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [11]. Figure 1 illustrates an abstract depiction of the service-based cloud computing environment. In this model, physical resources can be shared by multiple users across various layers, facilitated by virtualization through a hypervisor.

(b) Hypervisor

Virtual Machine Monitor (VMM) is also known as a hypervisor. The role of the hypervisor in a cloud environment is similar to that of an operating system on a PC. It organizes and ensures that the various virtual machines (VMs) receive the resources they have requested, acting as the intermediary layer between virtual machines and physical hardware [12]. With hypervisor (HV) technology, one machine can run multiple virtual machines simultaneously.

(c) Multitenancy

Multiple users can simultaneously access the same instance of a program thanks to the software architecture known as multitenancy. Through this approach, various virtual machines (VMs) on a server can serve end users by utilizing the same physical entities. Service-Oriented Architecture (SOA) employs various intermediary technologies, including HTTP and Simple Object Access Protocol (SOAP), to deliver the promised services to multiple clients.

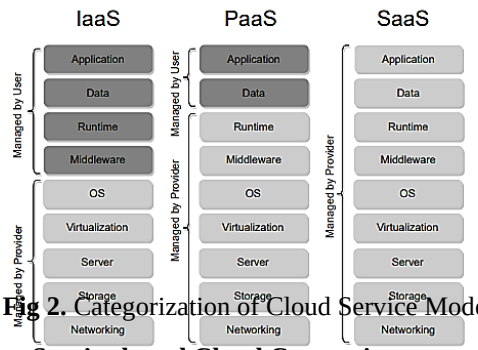

Fig 2. Categorization of Cloud Service Models

1.1.2. Service based Cloud Computing

The primary objective of the cloud computing paradigm is to offer high-quality, cost-effective services to clients worldwide. These services encompass any shared IT resource, including networks, software, and hardware. Because of their accessibility, availability, and scalability, three widely recognized service-based cloud platforms—IaaS, PaaS, and SaaS—are favoured by mid-size to large businesses [26].

(a) SaaS: The SaaS approach allows customers to access products and various programs on the cloud. This solution eliminates the need for data storage, application deployment assistance, and in-house applications. Organizations pay on a per-user basis to utilize the SaaS resources [13].

(b) PaaS: PaaS is a cloud computing solution that enables customers to develop cloud-related apps and services, supporting software throughout its entire lifecycle [14]. Instead of procuring their own hardware, programmers and developers utilize intermediary hardware and deliver the created programs to customers via the internet. Individuals or organizations can engage in application development using PaaS without the need to purchase the necessary hardware and software. Notable examples of the PaaS concept include Google App Engine, Amazon's Relational Database Services (RDS), and Microsoft's Azure services platform.

(c) IaaS: IaaS is a platform-based cloud computing solution provided in a virtual environment [15]. Users are not required to purchase any servers, data centers, network hardware, or space, such as Amazon EC2.

Table 1. Survey of Cloud Deployment Models

Cloud deployment model	Advantages	Disadvantages
Public	Scalable and reliable with on-demand resources	Might be unreliable
	Using is easy	Not very secure
Private cloud	Company specific	Expensive
	Can be customized	Needs IT expertise
Hybrid Cloud	Flexible Infrastructure	Lacks visibility
	Faster Speed	Application and data integration is difficult

1.1.3. Cloud Deployment Model

This deployment approach goes into more detail about the exclusive access to shared resources.

- **Private cloud:** It is a type of cloud deployment model where the sending environment is exclusively dedicated to private segments for the secure storage of corporate data [16]. Private clouds are essentially managed by external vendors but are located on-site.
- **Community cloud:** A community cloud is a cloud environment shared by multiple organizations for the same purpose. While community clouds and private clouds share common security and operational processes, they differ in that multiple organizations exclusively control computational resources and underlying infrastructure. Additionally, costs are higher compared to open clouds, and

access to information may not be precisely regulated due to the potential presence of untrusted parties. However, the community cloud offers the advantage of fair third-party access for security reviews.

- **Public cloud:** Large companies such as Google Apps, Amazon AWS, and Microsoft Office 365 dominate the public cloud market. In public clouds, resources are typically provided as pay-per-use services. The primary advantages include on-demand purchases, where higher utilization results in higher costs. Users of public clouds are typically individual users who connect from their homes to the providers' networks over the internet. Because the public cloud is highly accessible, concerns arise regarding data protection and privacy. There is no way to regulate the transmission or access to sensitive data [16]. Despite its significant security limitations, small firms have benefited from its services due to their minimal involvement with sensitive data.

- **Hybrid cloud:** A cloud service provided by a private cloud owner in collaboration with a public cloud owner is known as a hybrid cloud. This arrangement adds complexity as multiple cloud providers are involved. However, it offers the advantage of flexibility and cost-effectiveness without exposing sensitive data to outside parties or purpose-specific software. The scalability features of a public cloud are enhanced by incorporating private cloud features into the hybrid system.

Table 2 Characteristics of Cloud Deployment Models

Attributes	Deployment Model		
	Public Cloud	Private cloud	Hybrid cloud
Ownership	Owned by customers	Owned by single organization	Partially owned by service provider and partially by consumer
Performance	Low to medium	Excellent	Good
Setup cost of building data centre	Low initial cost	Excellent	Good
Used by	Anyone can access	Limited people can access	Medium accessibility
Security	Less	Highest	Moderate
Space required	Very low	Very large	Medium
Workload	Normal workload with short spikes in demand	Not suitable for handling the large workload	Highly dynamic or changeable
Virtualization	Server virtualization optimizes resource utilization by efficiently allocating and managing server resources.	Efficiency gains in resource utilization are achieved through server virtualization.	Server virtualization optimizes resource utilization.
Reliability	Medium	Highest	Medium

1.2. Load Balancing

Load balancing is the technique of equally distributing workloads among all available nodes to maximize resource utilization and enhance customer satisfaction. This strategy, known as load balancing, aims to ensure equitable distribution of work, promote resource efficiency, increase throughput, and reduce response times by distributing tasks among virtual machines [17]. The primary objectives of load balancing include maximizing throughput, minimizing energy and costs, optimizing resource utilization, and enhancing overall system performance [5]. By distributing resources across multiple nodes, load balancing ensures system consistency, resilience, and failure prevention in the context of cloud computing, making it an essential method for effectively managing application demands.

1.2.1. Challenges and Goals of Load Balancing

In the backdrop of cloud computing, load balancing encounters several challenges stemming from the dynamic and distributed nature of the environment. Here are the main challenges:

- **Dynamic Workload Variation**: In cloud environments, workload demand undergoes unpredictable changes. Load balancers must dynamically adapt to varying traffic patterns and efficiently distribute workloads to handle fluctuations in demand [6].

- **Scalability:** Load balancers must seamlessly scale with the number of resources and adapt to the changing size of the infrastructure.

- **Security Concerns:** Ensuring the security of data and applications while load balancing is critical. Load balancers need to implement secure communication protocols and avoid becoming a potential point of vulnerability in the system.

• **Cost Management:** Efficient load balancing entails minimizing costs associated with data transfer, server usage, and overall infrastructure.

The main goals of load balancing schemes are as follows: Optimize Resource Utilization, improve in system performance, Handle Workload Fluctuations, Ensure High Availability, Minimize Response Time, Cost Optimization, adapt to Changing Conditions, Least reaction time

1.2.2. Scheduling

Pinedo, in his [18], defines scheduling as a decision-making procedure regularly employed in various manufacturing and service industries. It involves allocating resources to tasks over specified time periods with the goal of optimizing one or more objectives.

The scheduler's tasks include finding ways to evenly distribute the load across nodes to meet load balancing goals by efficiently utilizing resources. In the early days of cluster computing, the concept aimed to merge isolated clusters into a single unit. However, reliance on local resources proved to be a bottleneck in cluster systems, leading to the development of Grid computing. Grid computing integrated heterogeneous systems across geographically distributed locations. Now, the transition from Grid to Cloud computing leverages the strengths of both Cluster and Grid. There exists no algorithm for optimizing computing resources, as most scheduling algorithms are non-proprietary (NP-Complete) and NP-Hard.

1.2.3. Need of Load Balancing

In contexts based on cloud computing, the allocation of various tasks to virtual machines (VMs) constitutes what is known as the load. Within a cloud system, these loads can be classified as underloaded, overloaded, or balanced. Load balancing algorithms aim to evenly distribute the overall system loads by transferring workloads from heavily burdened nodes to lightly burdened ones, typically through cloud migration. The objective is to optimize the total system throughput. One of the most critical aspects of assignment planning for cloud systems is balancing the workload of jobs, whether with or without VMs.

1.2.4. Categorization of Load Balancing Algorithms

In the realm of cloud computing, the allocation of various tasks to virtual machines (VMs) constitutes what is known as the load. Within a cloud system, these loads are categorized as underloaded, overloaded, or balanced. Load balancing algorithms play a crucial role in evenly distributing system loads by transferring workloads from heavily burdened nodes to lightly burdened ones, typically through cloud migration. This process aims to optimize the total system throughput, a fundamental objective in cloud computing research.

One of the most critical aspects of assignment planning for cloud systems is the balancing of workload for jobs, regardless of whether they involve VMs. This aspect warrants considerable attention in cloud computing research, as it directly impacts system performance and resource utilization efficiency.

1.3. Cloud Security

In recent years, significant advancements in information technology have emerged, with cloud computing playing a pivotal role in providing consumers with various storage options. Through cloud computing, manufacturers can now offer consumers space on their physical systems and lease their services on an hourly basis. Despite its benefits, cloud computing presents several security risks for consumers. A paper from the Cloud Security Alliance focuses on vulnerabilities related to application program interfaces (APIs) and cloud computing platforms, specifically addressing misuse, insecure interfaces, and criminal usage [21]. The paper emphasizes the three primary goals of information security—availability, confidentiality, and integrity—as paramount. Persistent concerns regarding data confidentiality pose a threat to these goals, as both current and older encryption techniques may be deemed insecure. Additionally, there is a risk of information leakage when data is outsourced, and the possibility of data tampering further jeopardizes data confidentiality.

Table 3: Load Balancing Algorithms and Their Characteristics

Nature	Name	Request Serve Method	Advantage	Disadvantage
Static	WRR	Every server is used by its weight.	Sends most requests to better abled and loaded servers	All the approximations require the implementation of this algorithm, and this is a major drawback.
	Source Hash	The source IP address will undergo hashing and categorization based on the total number of operating servers to determine which server receives requests.	Allowing users to reconnect to a currently active session after disconnection can enhance performance.	Managing dynamic IP addresses provided by ISPs can be challenging.
	Least Bandwidth	Every server is used, in turn, by network bandwidth.	Requests can be distributed among more capable servers with higher network bandwidth utilization.	Estimating network bandwidth can be challenging in networks with varying data packet sizes, potentially leading to bandwidth exhaustion.
Dynamic	RR	Requests are processed sequentially in a rotating manner as they reach the server.	Easy configuration, deployment widely used algorithm	Servers with varying resource capacities may become overburdened and crash due to differences in processing capabilities.
	Least Connections	The request is directed to the server with the fewest active connections.	It prevents server overload by verifying the number of connected servers.	When calculating the number of current connections, the server's capacity cannot be taken into consideration.
	Least Response Time	The request is directed to the server with the lowest response time to ensure optimal performance and efficiency.	Taking into account the server's capacity, response time, and current number of connections is crucial to prevent overload and potential crashes.	If simple and basic-level virtual machines are utilized, the non-uniform traffic routing may experience slowdowns, making this algorithm unsuitable for cookie-based session applications.

1.3.1. Cloud Security Threats

The three fundamental pillars that encompass the majority of security risks in cloud computing are: Confidentiality,

integrity, and availability. Adverse events that may impact digital assets stored in cloud environments are referred to as cloud security issues. These assets include client trust, infrastructure, software, data, and company reputation. This study categorizes security concerns into four distinct groups: (1) security issues related to data; (2) security issues related to networks and services; (3) security issues related to applications; and (4) security issues related to people. This classification was developed based on the latest trends in attacks on cloud computing platforms.

A.Data Security Issues

Cloud computing represents a unique form of data sharing, where user data is distributed across multiple sites, processed, and made available to stakeholders as needed. Consequently, customers of Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) platforms are particularly concerned with ensuring data security in cloud environments. Data security aims to enable authorized users to access, transfer, or modify data in accordance with their rights, while preventing unauthorized requests from accessing resources. This is achieved by limiting data access to only those who have been authenticated.

- **Storage:** Cloud-based computing solutions usually give users very little authority over the information kept in cloud service provider-run data centres. While some control may be granted over virtual machines, users often lack control over data storage. This lack of control leaves data vulnerable to tampering by attackers after users upload it to the cloud. Additionally, cloud service providers may have the ability of replication, manipulation, or advanced modification of user data without the user knowing about it.

- **Location:** In cloud computing settings, data is scattered over various physical locations and forms, making it challenging to identify the site of every data item. Moreover, each geographical location may have its own laws and regulations governing data handling, which must be adhered to. Users could need to know how close their data is to them, and cloud service providers might have to tell them of this. Additionally, unnecessary apps may be stored on public clouds, making data management even more difficult.

- **Access:** To stop illegal access to services and data held in cloud settings, it is crucial to keep an eye on user identity and activity. Although access controls aid in maintaining the confidentiality of data, administering access and identity restrictions is difficult because data owners and data are spread across several platforms and places. In cloud environments, organisations cannot rely only on their authentication and authorization procedures because cloud resources are dynamic and IP addresses are subject to frequent changes.

- **Privacy Breaches:** In cloud computing, data privacy is challenging since unencrypted data is kept on servers run and owned by parties other than the data owner. Sensitive information belonging to an organization could be made available to other users who share the same storage space via a cloud data breach. The impact of a breach can be increased since users running various apps on the virtual machines can share the same database due to multi-tenancy. The date, mode, and degree of data exposure must all be taken into account in investigations into data privacy incidents. Laws, policies, and procedures that protect personally identifiable information must be implemented in order to preserve privacy. When sensitive data is accessed without authorization, cloud service providers should quickly discover it and take the necessary action. Depending on the usage situations and cloud models, privacy problems may vary.

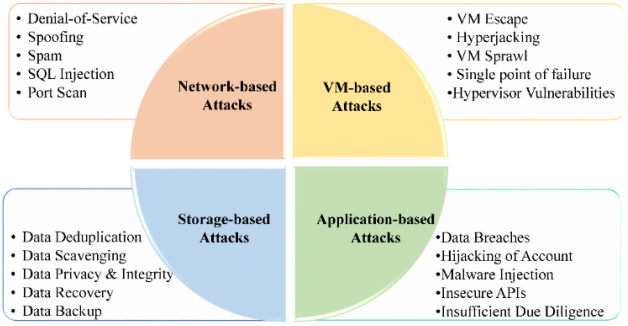

Fig 3. Types of Cyber Attacks in Cloud Computing

1.3.2. Network and Services Related Security Issues

This section addresses various security issues related to networking and cloud computing services, encompassing virtualization, availability issues, account or session hijacking, and multi-tenancy.

- **Account or Session Hijacking:** Consumers accessing information and cloud computing services via cloud-based systems are susceptible to account or session hijacking, where unauthorized individuals exploit passwords to gain access to cloud resources. This unauthorized access could lead to data alteration, theft, or other malicious activities.

- **Multi-Tenancy:** Multi-tenancy in cloud computing involves various users sharing the same computational resources including any kind of hardware and software provided by a cloud vendor while maintaining principle of data segregation. This architecture allows users to share resources like software, hardware, services, and network resources while ensuring data isolation and security.

- **Virtualization:** Virtualization technology is essential to cloud computing in order to maximize resource utilization. Resources are available to users on a pay-per-use basis, selecting processors, RAM, bandwidth, or operating systems based on their requirements and paying only for the

resources used. However, virtualization introduces security risks due to increased entry points and interconnection density, making virtualized environments susceptible to various attacks.

• **Availability:** Continuous service delivery depends on cloud systems being simple to use. To satisfy the needs of businesses that depend on vital services, cloud service providers must guarantee on-demand service delivery. By consuming available resources, attacks like denial of service (DoS) can cause availability to be disrupted, leading to delayed or unavailable services. Non-availability can also be caused by hardware malfunctions, under-provisioned bandwidth, cloud outages, and resource mismanagement.

• **Backup:** Data backup is essential for maintaining data security and facilitating recovery in case of disasters. Regular backups help ensure data availability by adhering to security guidelines to prevent unauthorized access or tampering. Consistent backup practices are crucial to enable quick recovery and mitigate the impact of potential data loss incidents, protecting against unauthorized access and tampering.

Table 4. Classification of Cyber Attacks in Cloud Computing

Classification of Attack	Description	Attack Name
Denial of Service Distributed Denial of Service	Huge quantity of data traffic is caused by the attacker to prevent the availability of all services A DDoS is the distributed form of DoS where the system is flooded in a distributed manner.	• SMURF: ICMP: generation of echo request to a targeted IP address. • LAND: transfer of spoofed SYN packets with the same source and target IP address. • SYN Flood: reduction in efficiency of storage via IP spoofed packets. • Teardrop: exploitation of flaw TCP/IP stacks. • HTTP Flooding: exploiting legitimate • HTTP POST or GET requests. • Zero Day Attacks: exploiting security loopholes • unknown to CSPs.
Remote to Local User to Root	Attacker compromises the computer system as they have executed commands that grant access Attacker gains root access to destroy The system.	• SPY: software that runs a • Machine for phishing purposes. • Password guess. • IMAP: finding a vulnerable IMAP Mail server. • Root kits: Offering privileged access while masking its existence. • Buffer Overflowing
Probing	Breaching the PII of a victim	• Ports Sweeping. • NMAP: port scanning.

1.3.3. Application Related Security Issues

The following are the issues that cloud apps face:

• Malware Injections: Due to the significant security risk posed by virus injections, internet-connected devices require meticulous configuration for multiple user support. Inadequate setups may result in malware infestations and data leaks, thereby jeopardizing the entire cloud computing environment of both the organization and the cloud service provider. Malware injections occur when embedded code in cloud services, often operating as Software as a Service (SaaS) on cloud servers, is executed.

• **User Interfaces:** While cloud applications allow users to customise their cloud experience, they can seriously compromise the security of the cloud infrastructure as a whole. Many container-based solutions, however, lack built-in security shields. Programmers can create programmes and integrate them with the cloud using Application Programming Interfaces (APIs), also called user interfaces. Even though the purpose of this interface is to give users access to cloud services, certain APIs allow users to access cloud customers' potentially vulnerable systems, so there is potential for abuse. It is essential to keep software services' updates up to date since users could unintentionally become targets of attacks that compromise their data.

• **Development Life Cycle:** Weak software may allow vulnerabilities to remain even when protective measures like firewalls, antivirus programmers, and encryption are put in place. Compared to traditional approaches, cloud software development is more complex and introduces security flaws at every stage of the software development life cycle. Throughout the development life cycle, frequent modifications could jeopardize security while advancing development. It is essential to take preventative measures to deal with security flaws and malicious attacks when developing, testing, implementing, and designing cloud applications. Applications that use Platforms as a Service (PaaS) in particular need special consideration. Incorrect Software Development Life Cycle (SDLC), an over reliance on programmers, risky reverse engineering techniques, and post-deployment issue detection are some of the security vulnerabilities in the development life cycle.

Table 5. Types of Cloud Computing Attacks

Attack Name	Description	Affected Layer	Additional Parameters
Service Injection	This attack compromises the integrity of services, targeting both VMs and application layers. Viruses are injected into legitimate files, providing malicious services.	PaaS	Impact on data confidentiality, resource utilization, and cost
Zombie	This attack disrupts service availability by flooding legal VMs directly or indirectly through host machines.	PaaS, IaaS and Saas	Impact on network bandwidth, latency, and scalability
Hypervisor and VM Attack	Malicious actors compromise the hypervisor, gaining access to users' VMs by bypassing the virtualization layer.	IaaS	Impact on VM isolation, hypervisor security, and workload migration
Man in the Middle	Interception of data transfer or communication to users, compromising message integrity and confidentiality.	PaaS, IaaS and Saas	Impact on user trust, data integrity, and compliance
Back Door Channel	Attackers exploit data privacy and service availability by creating a backdoor in a valid VM, allowing unauthorized access to resources.	IaaS	Impact on auditability, incident response, and governance
Phishing	Users are directed to fake or malicious websites, posing a threat to the privacy of sensitive data.	PaaS, IaaS and Saas	Impact on user awareness, security awareness, and education
Spoofing Meta Data	Privacy and security of services are compromised through anomalous behavior, such as modifying web service descriptions.	PaaS and SaaS	Impact on service discovery, API security, and data validation
Side Channel Attack	Data integrity is compromised as hackers gain access to plaintext or ciphertext from encrypted data through side channels.	SaaS and PaaS	Impact on encryption strength, data leakage prevention, and regulatory compliance
Authentication Attack	Exploitation of authentication protocol vulnerabilities to gain unauthorized access.	PaaS, IaaS and SaaS	Impact on identity management, access control, and incident response

2. RELATED WORK

Although cloud computing safety issues have been studied from a variety of angles, the most popular subjects in the literature right now include virtualization, multitenancy, data security, and general vulnerabilities. The entirety of the literature review is made up of reviews of studies that are special to clouds, discussing different security risks and defenses against such attacks. The studies pertaining to cloud-generic vulnerabilities are covered in the next subsection.

2.1.Load Balancing and QoS Management Techniques

As shown in Fig. 4, we then divide balancing techniques into four major categories according to the kinds of algorithms that are used for this, which range from traditional strategies to hybrid heuristics. All of these algorithms are briefly discussed in the ensuing subsections. We have analysed numerous algorithms used and developed by researchers in each sector by taking into account multiple factors. As a result, the load balancing algorithms fall into the following categories:

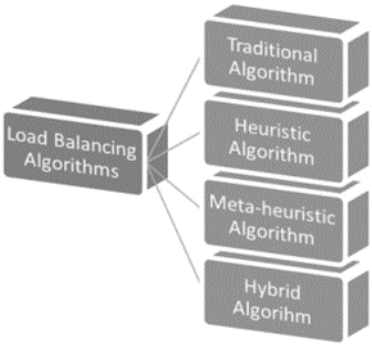

Fig 4. Different Types of Load Balancing Algorithms

2.1.1. Traditional Algorithm

This method integrates well-established CPU scheduling techniques with precision. CPU scheduling is a technique that enables the full utilization of the CPU by allowing one process to run while others wait in the queue. We can see operating system (OS) selecting a process from the queue that has all the ready processes and assigns it to a CPU for execution. In distributed computing systems, various load balancing scheduling strategies exist. Classical algorithms fall into two main categories: preemptive and non-preemptive. Preemptive algorithms involve pausing an ongoing execution to service a higher priority task, resuming it after the high priority task completes. Additionally, a priority—either internal or external—determines this [33]. Each task receives a priority based on its scheduled completion time. Priority-based scheduling and Round Robin scheduling represent examples of such methods.

Fathalla et al. [22] provided a methodology that helped categorize the most advanced preemptive resource allocation techniques into two groups: heuristic and brute force. While brute force-based strategies could maintain system performance, heuristic-based solutions lacked speed. In their study, they presented a novel multi-objective preemptive resource allocation policy that capitalized on these two categories. The Best K-First-Fit (Best-KFF) heuristic was introduced. Each physical machine (PM) was assigned k preemption choices determined by the Best-KFF algorithm. These choices were then ranked by the PMs based on various objectives, such as resource utilization. The Best-KFF algorithm selected the optimal option for preserving the performance of the cloud computing system. Thus, the Best-KFF algorithm represented a compromise between the heuristic and brute force categories. The search space expanded with an increasing value of k.

Kumar et al. [23] presented an understanding of the emergence of workload prediction schemes as highly intelligent solutions in contemporary times, owing to their scalability in automatic resource management, cost-effectiveness, and enhancement of resource utility in the cloud. They highlighted that while workload prediction, based on the single-model prediction approach, offers several schemes, arriving at a reasonable conclusion using traditional methods is challenging due to the vast scale of heterogeneous data delivered over the internet to the cloud. Their research conducted analyzes a significant amount of heterogeneous data in the cloud and proposed a proactive strategy for resource allocation. Their study demonstrates superior accuracy in resource prediction compared to current approaches, all while requiring less time and financial investment.

Sohani et al. [24] addressed the ongoing challenge faced by cloud providers in managing resources due to fluctuating cloud workloads in heterogeneous system environments. They propose the Predictive Priority-based Modified Heterogeneous Earliest Finish Time (PMHEFT) approach, which effectively predicts future resource demands of applications, thus resolving this issue. Their research aims to develop a prediction-based model for dynamic and efficient resource provisioning in a heterogeneous system context, catering to the end user's needs. They highlighted the shortcomings of current methods in integrating cloud computing characteristics, such as resource elasticity and heterogeneity, and in meeting user Quality of Service (QoS) requirements, such as minimizing makespan and satisfying budget constraints. The paper proposed the PMHEFT technique to enhance load balancing across virtual machines and minimize the makespan of specific workflow applications.

2.1.2. Heuristic, Metaheuristic and hybrid Algorithm

Kaur et al. [25] introduced a novel hybrid approach to optimize load balancing and resource provisioning during workflow execution, aiming to maximize virtual machine utilization while ensuring balanced load distribution. Their framework seeks to achieve optimal behaviour in terms of both makespan and cost by integrating heuristic techniques with metaheuristic algorithms. Specifically, they planned two hybrid techniques within the HDD-PLB framework: the Hybrid Predict Earliest Finish Time (PEFT) Heuristic with Ant Colony Optimization (ACO) metaheuristic (HPA) and the Hybrid Heterogeneous Earliest Finish Time (HEFT) Heuristic with ACO (HHA).

Table 6. Optimization Techniques in Resource Allocation and Load Balancing: A Comparative Analysis

Article	Work done	Optimization	Pros/Cons
Fathalla et al. [22]	Multi-objective pre-emptive optimization for resource allocation	Best K-First-Fit (Best-KFF).	It is limited to the search space which relies on value of k
Kumar et al. [23]	Pre-emptive resource allocation for heterogeneous data in cloud	Customized optimization based on master slave servers	
Sohani et al. [24]	Modified Heterogeneous Earliest Finish Time (PMHEFT)	Heterogeneous Earliest Finish Time	Better resource allocation, low cost and time
Kaur et al. [25]	resource provisioning and load balancing	Combination of Hybrid PEFT Heuristic with ACO	
Alghamdi et al. [26]	BPSO for work load scheduling	Meta-heuristic	Slow convergence
Muhammad Junaid et al. [28]	SVM based model to classify the input request	Machine Learning model	Data set labelling and training time consumption are challenging
Yadav et al. [29]	Genetic algorithm and PSO	Hybrid meta-heuristic	It provides refined optimal solution
Princess et al. [30]	Load balancing using Harries Hawks Optimization and Pigeon inspired Optimization	Hybrid meta-heuristic	Pre-mature convergence and Exploration vs. Exploitation Balance are challenging

Alghamdi et al. [26] presented an emphasized view on the potential for significant cost reduction and resource optimization through an efficient resource allocation system. To address the limitations of inefficient heuristic methods, they developed Binary Particle Swarm Optimization (BPSO). However, to achieve the optimal solution, it was essential to integrate these algorithms with other heuristic or meta-heuristic algorithms. Despite their effectiveness, these algorithms may be less practical due to their considerable temporal complexity in real-world applications. The binary variant of PSO is proposed specifically for addressing cloud computing workload scheduling and balancing for the NP issue. Their objective function evaluates whether there is a significant difference in completion time among heterogeneous Virtual Machines (VMs), considering the optimization and updating constraints outlined in their research. Additionally, they devised a mechanism to update particle placements in conjunction with load balancing.

Mishra et al. [27] introduced an alternative approach by employing the BSO-LB algorithm, inspired by the characteristics of a flock of birds, to devise the LB technique. In this analogy, jobs are equated to birds, while Virtual Machines (VMs) are likened to food particles. They utilized the cloudlet-based GoCJ to access the datasets used for their measurements. Their contribution led to a significant reduction in reaction time, facilitating equitable effort distribution.

Muhammad Junaid et al. [28] presented an innovative method by utilizing a support vector machine for classifying input requests. Subsequently, they provided an assignment to a hybrid metaheuristic approach, which combined file type formatting with Ant Colony Optimization, based on its classification. They presented their assertion that the stability of cloud systems can be maintained with the assistance of the hybrid metaheuristic algorithm they developed. The effectiveness of the proposed method was evaluated using metrics such as throughput, quality of service, overhead times, migration times, and violations of service level agreements.

Yadav et al. [29] introduced a method aimed to select the best solution from multiple options in optimization. They noted the expense of finding optimal algorithms for NP-hard problems, leading to a focus on approximations for virtual machine load balancing. They observed the increasing popularity of heuristic, meta-heuristic, and hybrid optimization strategies for solving difficult problems efficiently. They developed a hybrid meta-heuristic approach, drawing from Particle Swarm Optimization and Genetic Algorithm techniques inspired by nature.

Annie Poornima Princess et al. [30] presented a study on Load Balancing (LB), which refers to the distribution of dynamic workloads among cloud systems to ensure equitable resource sharing and prevent server overload or underload. They combined the benefits of the Pigeon-inspired Optimization Algorithm and the Harries Hawks Optimization to develop an efficient load balancing technique, aiming to optimize resource utilization and task response times in the cloud. Their proposed method was implemented using the Java Net Beans IDE integrated with the CloudSim framework. Performance evaluation was conducted by analyzing various tasks. Simulation results indicated that the load balancing scheme based on the proposed Hawks Optimization and Pigeon-inspired Optimization method achieved optimal load distribution among virtual machines in a shorter timeframe compared to existing techniques.

Kakkottakath et al. [31] offered an improved version of PSO combined with Ant-Lion Optimization (ALO) technique for better workflow scheduling in the cloud. They secured cloud data using Data Encryption Standard (DES) during scheduling to enhance security. Their goal was to enhance workflow scheduling in a safer manner compared to current methods, considering factors like makespan, load, and cost.

Khan et al. [32] presented an innovative dynamic load-balancing that computes load values for each Virtual Machine (VM) using a deep learning model that combines Recurrent and Convolutional neural networks. The objective of this method is to get better cloud performance by optimising job scheduling and workload distribution. Their suggested paradigm uses a dynamic clustering technique based on computed loads to divide virtual machines (VMs) into overloaded and underloaded clusters. To increase the efficiency of clustering, they integrated the Hybrid Lyrebird Falcon Optimisation (HLFO) method with Reinforcement Learning (RL). To increase load balancing efficiency, HLFO combines the Lyrebird Optimisation Algorithm (LOA) and the Falcon Optimisation Algorithm (FOA). In addition, they introduced a Multi-Objective Hybrid Optimisation model that maximises task scheduling while taking into account Quality of Service (QoS)

requirements such efficient makespan, reduced energy usage, balanced CPU utilisation, and RAM use, as well as task hierarchy.

Table 7. Methods for Optimization in Task Scheduling and Resource Allocation: A Comparative Study

Article	Work done	Methodology	Pros/Cons
Kakkottakath et al. [31]	Combined workflow scheduling and security	ALO and PSO for scheduling and DES for security	Improved search process
Khan et al. [32]	Combination of CNN and RNN	Deep Learning with HLFO, FOA	It suffers from vanishing gradient problem
Ghafir et al. [33]	task scheduling and resource allocation using optimisation	PSO with bilateral transposed filtering. Later, Double Deep Q proximal model with a feedback controller is used	Training complexity for limited data
Gupta et al. [34]	Multi-objective optimization problem	Whale Optimization Algorithm	performance sensitive to parameter settings
Thilak et al. [35]	Task scheduling and load distribution	Hybrid Genetic Algorithm for optimization	Handles nonlinearity, and multimodality
Apat et al. [36]	Task scheduling	Population based meta-heuristic approach	Local search abilities of SA improves the optimization process
Khan et al. [37]	Task scheduling	parallel enhanced whale optimization algorithm	not always guarantee the global optimum due to its stochastic nature.
Elsakaan et al. [38]	Clustering to identify the similar group of datacentre , round robin for task scheduling and GA for task allocation to serer	K-means ,round robin and genetic algorithm	Queuing delay

Ghafir et al. [33] introduced a technique that utilizes a unique PSO Approach based on Intelligent Weighted Filtering to decrease calculation time during resource allocation and task scheduling. Their method achieves good quality of service, throughput, scalability, short reaction time, and optimal bilateral transposed convolution filtering by employing a multi-objective PSO algorithm with Pareto dominance. Furthermore, current VM migration procedures result in violations of service level agreements due to ineffective VM assignment among PMs. To address these concerns, a Double Deep Q Proximal model associated with a feedback controller has been planned. The double weight set of the choice model in the offline and online updating procedure ensures the smooth operation of cloud service level agreements. Additionally, in complex scenarios involving mixed process instructions, both centralized and decentralized controller algorithms fail due to a single point of failure and coordination issues. Finally, the usage of the conditional GAN feedback controller eliminates a solo point of breakdown while maintaining elevated fault tolerance, little energy consumption, and short relocation time

Gupta et al. [34] introduced their Multi-Objective Whale Optimization-Based Scheduler (WOA-Scheduler), designed for efficient job scheduling in cloud computing environments. Their paper presents the scheduler, which utilizes the Whale Optimization Algorithm (WOA) to simultaneously optimize cost, time, and load balancing. One of its key features is its flexibility to accommodate user-defined weights for various purposes, enabling organizations to prioritize optimization targets according to their unique needs.

Thilak et al. [35] proposed an approach to address the need for appropriate job scheduling methods in data centers, ensuring equal load allocation to systems with higher scalability and performance. Their method aims to optimize output, reduce response time, minimize resource utilization, and conserve energy by matching resources to the workload effectively. Their recommended approach employs a two-stage task scheduling methodology. Firstly, virtual machines are generated using historical task data through classification and clustering techniques. Subsequently, a hybrid ant-genetic algorithm is utilized to schedule the optimal Virtual Machine (VM) for each task, leveraging the advantages of genetic algorithms and incorporating pheromone values from ant colony algorithms.

Apat et al. [36] provided an overview of explored effective resource allocation approaches as one of the optimal options for meeting the Quality of Service (QoS) criteria and enhancing system behaviour. They addressed the challenge of determining the optimal allocation approach for Internet of Things (IoT) applications, which involved multiple QoS parameters and is a Non-Deterministic Polynomial Time (NP)-Complete problem. Their study focused on optimizing the three parameters of makespan, cost, and energy using a traditional weighted multi-objective IoT service placement method. Given the non-convex nature of the resolution space, they opted to concentrate on population-based meta-heuristic algorithms such as Genetic Algorithm (GA), Simulated Annealing (SA), and Particle Swarm Optimization (PSO), along with their combinations GA-SA and GA-PSO.

Khan et al. [37] proposed a similar improved whale optimization algorithm for scheduling separate jobs with heterogeneous resources in the cloud. By employing an adaptive bubble net offensive mechanism and a modified encircling manoeuvre, their approach enhances solution diversity while avoiding local optima. Despite its inherent complexity, the parallelization strategy ensures low execution time. Their suggested technique improves throughput and resource efficiency while minimizing makespan. Their study also showcases the effectiveness of the proposed PEWOA compared to Multi-core Random Matrix Particle Swarm Optimization (MRMPSO) and the top-performing upgraded whale optimization algorithm (WOAmM).

Elsakaan et al. [38] introduced a hybrid strategy that surpasses existing methodologies in performance parameters such as makespan, response time, number of cloudlet migrations, and SLA violations. Their strategy operates in two stages. First, servers in each data center are grouped based on comparable utilization rates using the k-means clustering algorithm. Subsequently, task groups are sequentially assigned by a round-robin technique to clusters that are not overloaded. Within each cluster, genetic algorithm determines the optimal job assignment to servers, resulting in a multi-layered design that promotes robust

interoperability and the decoupling of essential mechanisms, facilitating easier hot-deployment and scaling in operational cloud environments.

2.2. Privacy Preservation techniques in cloud computing

Ge et al. [39] provided an in-depth analysis of the security measures implemented to ensure data security, emphasizing the encryption of data before transmission to the cloud. However, they noted that locating and exchanging encrypted data posed greater challenges compared to plain data. Despite this, the expectation for fast searches without compromising data confidentiality remained crucial for cloud service providers. To address these challenges, they introduced the Cypher text-Policy Attribute-Based method with Keyword Search and Data Sharing (CPAB-KSDS), which allowed attribute-based data sharing and keyword search simultaneously, marking a significant advancement. Additionally, they discussed the flexibility of CPAB-KSDS, where the scheme's keyword could be changed during the sharing stage without contacting the PKG. Furthermore, they provided a concrete solution in the random oracle model, demonstrating its security against chosen ciphertext and chosen keyword attacks.

Agarwal et al. [40] offered valuable insights into existing security protocols, emphasizing the use of various cryptographic algorithms to ensure the safety of transmitted information to the cloud. However, they highlighted potential inefficiencies in scenarios where a user's access was revoked, prompting the need for a secure data sharing framework employing enhanced cryptography approaches. Their proposed approach, leveraging cipher text-policy attribute-based encryption and dynamic unidirectional Proxy Re-Encryption (PRE), provided a secure, privacy-preserving mechanism. Furthermore, it guaranteed the security, privacy, and integrity of data during retrieval from the cloud.

Liu et al. [41] introduced a novel short-term fingerprint attack technique for the e-Finga program, demonstrating how adversaries could analyze specific secret settings and fingerprint characteristics by eavesdropping on a user's temporary fingerprint ciphertext. To mitigate this threat, they proposed the Secure e-fingerprint system, which encrypted users' temporary fingerprints using samples from learning with errors, thereby possessing the homomorphic addition property.

Tolba et al. [42] underscored the serious security risks posed by vulnerabilities in transmitting sensitive data from Internet of Medical Things (IoMT) devices to the cloud via open wireless connections. Despite existing cryptanalysis works on theoretical models of wireless encryption protocols, there remained a notable performance gap between theory and practical implementations of these attacks. To bridge this gap, they proposed a method based on a deep learning model for simulating attacks on specific IoMT protocols, utilizing a conceptual scenario involving smart eavesdropping on multiple layers of the IoMT protocol stack to simulate an attack and discover the private key of IoMT wireless communications.

Malik et al. [43] emphasized the threat of Cross-Site Scripting (XSS) attacks targeting web browsers, highlighting the critical aspect of cloud security in organizational health. Their approach aimed to unify the protection of private data across online platforms, applications, and organizations. In their proposed system, subsequent client requests containing encrypted cookies are routed to the web proxy for decryption before forwarding them to the web server, enhancing overall security.

Park et al. [44] addressed privacy concerns in integrating reinforcement learning (RL) methodologies into data-centric services operating within cloud computing infrastructures. They proposed leveraging homomorphic encryption (HE) schemes to enable cloud platforms to execute arithmetic operations on ciphertexts without decryption. This approach mitigated the risk of exposing sensitive data, with their solution involving the development of a privacy-preserving reinforcement learning (PPRL) framework tailored for cloud computing environments.

Mei et al. [45] proposed a secure blockchain-enabled privacy-preserving authentication scheme for transportation cyber-physical systems (CPS) in cloud-edge computing environments. Their scheme supported unconditional anonymity and data batch integrity verification, simplifying key management issues. They utilized elliptic curve cryptography to construct a pairing-free ring signature scheme, reducing resource overhead in transportation CPS with cloud-edge computing. Additionally, they demonstrated the security of their scheme based on the elliptic curve discrete logarithm problem under the random oracle model.

Abirami et al. [46] enhanced distributed secure outsourcing using crypto-deep neural networks to mitigate impersonation attacks and enhance trust among cloud users. Their proposed framework, incorporating cloud server, web server, data center, and cloud agent, aimed to handle impersonation attacks using crypto-deep neural network cloud security (CDNNCS), offering enhanced security compared to secure linear algebraic equation schemes.

Salim et al. [47] proposed a privacy-preserving scheme using homomorphic encryption to secure medical plaintext data from unauthorized access. Their approach involved distributing computations to several virtual nodes on the edge and masking all arithmetic operations, preventing untrusted cloud servers from learning the tasks performed on encrypted patient data.

Ghayvat et al. [48] introduced a scheme integrating blockchain (BC)-based confidentiality-privacy (CP) preserving scheme, CP-BDHCA, operating in two phases. They proposed an elliptic curve cryptographic (ECC)-based digital signature framework, HCA-ECC, to establish a session key for secure communication among healthcare entities. Additionally, they proposed a two-step authentication framework, HCA-RSAE, integrating RSA and AES to safeguard against possible attack vectors.

Shukla et al. [49] proposed a novel ECC based provably secure and privacy-preserving multi-factor authentication protocol for the cloud environment. Their protocol delivered user anonymity, unlinkability, perfect forward secrecy, and session key security as security and privacy authentication features. The security of their protocol was theoretically proven under the Real-Or-Random (ROR) model.

Huang et al. [50] proposed a privacy-preserving multi-dimensional media sharing scheme named SMACD in mobile cloud computing. Their scheme encrypted each media layer with an access policy based on attribute-based encryption, guaranteeing media confidentiality and fine-grained access control. Moreover, they introduced decentralized key servers to achieve both intra-server and inter-server deduplication by associating different access policies into the same encrypted media.

2.3. Access control based mechanism in cloud computing

Xiong et al. [52] propose a novel CP-ABE-based storage model tailored for securely storing and accessing data in the cloud, particularly for IoT applications. Their framework introduces an attribute authority management (AAM) module within the cloud storage system, functioning as an agent to provide user-friendly access control while significantly reducing the storage overhead of public keys. Additionally, they present a secure and efficient multiauthority access control scheme for the cloud storage system, named SEM-ACSIT, which ensures both backward and forward security in case of attribute revocation.

Prince et al. [53] address the increasing concerns regarding data security and privacy in cloud-based pervasive healthcare systems. They propose an innovative access control model utilizing privacy ratings (PR) to ensure high privacy, data confidentiality, and availability of health data. This model employs a PR-based approach to provide access control to various system users, calculating PR for both users and data to grant access based on predefined thresholds.

Saini et al. [54] aim to establish an access control framework based on smart contracts atop a distributed ledger (blockchain) to secure the sharing of Electronic Medical Records (EMRs) among entities in smart healthcare systems. They propose four forms of smart contracts for user verification, access authorization, misbehavior detection, and access revocation, respectively. In their framework, EMRs are encrypted using cryptographic functions like elliptic curve cryptography (ECC) and Edwards-curve digital signature algorithm (EdDSA) before being stored in the cloud, with corresponding hashes packed into the blockchain.

BenMarak et al. [55] introduce a cryptographic solution, Chaos-ABAC, to enhance the security and reliability of the Attribute-Based Access Control (ABAC) model. Their proposal integrates chaotic algorithms for both data encryption and decryption within the existing structure of attribute-based encryption, offering a robust encryption model.

Qin et al. [57] propose an access control scheme combining lightweight decryption based on Attribute-Based Encryption (ABE) and blockchain technologies to address computing overhead challenges in the big data environment, especially for resource-constrained IoT devices. Their scheme ensures the accuracy of proxy re-encryption calculations based on the blockchain and introduces a user credibility incentive mechanism to dynamically adjust the endorsement protocol based on users' access behavior.

Hwang et al. [58] focus on enhancing the efficiency and security of CP-ABE access control in dynamic cloud environments. They propose a scheme to block access of withdrawn users, eliminate user access post-removal, and output ciphertexts of constant size. Their scheme aims to improve efficiency by revoking attributes of withdrawn users and minimizing the computational burden on users during decryption operations.

Bera et al. [59] propose a secure lightweight Attribute-Based Verifiable Data Storage and Retrieval Scheme (ABDSRS) for cloud environments, offering features such as lightweight design, provable security, fine-grained data access control, and data owner anonymity. ABDSRS utilizes an attribute-based online-offline mechanism, allowing only authorized data owners to upload data anonymously to the cloud, and enables data users to search over encrypted data using keyword policies while verifying the correctness of search results.

Dhal et al. [60] propose a scheme named Centralized Multi-Authority Cloud Storage with Revocation (CEMAR) to address secure data sharing challenges in cloud computing, particularly in scenarios with high computational demands at the user end. Their model partially outsources the decryption process to the cloud server, reducing computational burden while storing the required keys in the cloud server to enhance decryption process communication cost.

3. ISSUES AND CHALLENGES

Through the utilization of cloud computing, enterprises can now leverage state-of-the-art cloud infrastructures that offer increased productivity, reduced costs, and enhanced efficiency. The management of traditional cloud infrastructures needs to be revaluated given the advancements in 5G technology, reliable internet connectivity, intelligent smartphones, portable gadgets infrastructure, and sophisticated AI-driven data analysis systems. The provisioning of IT resources through cloud-based platforms has become more feasible. As a result, an organization needs to invest little time and expertise in cloud configuration. However, a user may also find themselves with an infrastructure that is susceptible to various cyber security problems due to their lack of expertise about a particular cloud and the diverse characteristics of the cloud. This could lead to data intrusions, denial of service attacks, hijacking of a session, and other such attacks.

3.1. Confidentiality, Integrity and Availability (CIA)

Ease of use, reliability, and confidentiality maintenance prove to be major issues associated to cloud computing. Unauthorised access must be prevented to data gathered by IoT devices. Data may be added, changed, copied, or removed as a result of this. Furthermore, while communicating data via any insecure channel, maintaining discretion is key before transferring the information to cloud servers.

3.2. Aspect of Application Security

One major obstacle and point of vulnerability in data security is the security of software applications. Numerous frameworks and application platforms may be linked to different vulnerabilities. Cloud computing application security vulnerabilities provide a significant challenge. This is especially important because millions of lines of code are written while developing programmes in different dialects through various programmers, which increases the variety of vulnerabilities that go in addition to them. In cloud computing, developers might be the only ones in charge of cloud apps, nevertheless there is no unanswered questions about application networking or programming safety. Furthermore, operating systems might play an essential component in ensuring the information safety of the internet.

3.3. Limited Computation Resources

Until recently, most organisations had no idea where, how, or how much data and burden was held in cloud-based platforms. It is now imperative to be dependent on cloud service providers to manage the issues. With uneven workloads, service ability must be modified in response to demanding that helps avoid oversizing in times of low demand or service performance degradation in times of high demand. More information has been made available by forensic investigations, and network recording and surveillance on platforms, the Internet of Things, information, and networks that are physically connected have been made easier. Yet, cloud service providers charge for mirroring because it takes extra bandwidth, which raises the costs. It is hard for cloud service providers to suit the expectations of every cloud user, especially without charging more. The compromised node is used to create traffic during a resource exhaustion attack, which uses up the nodes' energy. In an attempt to bring down the network, these nodes expend all of their resources. The layer of routing protocols is thus made the attack target. These kinds of attacks, which deliberately drain memory and network bandwidth, can be directed towards any computing resource, including cloud-based computing resources. The cloud is vulnerable to attacks of this kind, where the attack's commencement results in resource depletion, because of its ability to manage workloads on a scale basis. Examples of such attacks include taking advantage of weaknesses in application communication and flooding protocols that are dependent on volume.

3.4. Security Issue Classification

Cloud computing has faced an amount of security issues since its launch. However, given the present condition of cloud computing and technology, researchers need to be aware of several new security issues, such as virtualization, multiple occupancy, and other types of cyberthreats. Data resources might exist in several areas and take on different forms in a cloud computing scenario. Thus, it's critical to categorise data possessions and address safety concerns in accordance with the appropriate categorization level. As a result, less cash and work would be required to maintain security. It can be challenging to categorise information shared by several users and organisations since various organisations may value particular bits of information over others.

Due to the complexity of modern cloud infrastructures, security organisations have to cope with issues including data duplication, early attack detection, a decline in regulatory compliance, and the requirement for control over information access. Moreover, the cloud infrastructure—the buildings and data—must be secured against known and unknown intrusions in order to attain complete cloud security. This is a challenging task to complete across all cloud components.

Cloud service providers may find it challenging to ensure that measures are in place to stop data loss or alteration. The safe storage of applications and data, the protection of interfaces, and the restriction of data access to authorised personnel allow for the control of data breaches and hacking incidents. Cloud service providers need to keep controls in place to deal with these issues. Furthermore, it is imperative

to promptly identify malware that uses botnets to eavesdrop on systems. These dangers can do great harm and hide in a mist for longer than usual techniques. Inadequate security measures for spotting illegal access through network data monitoring, which needs attention, can also result in data breaches. Another challenge for cloud computing is managing insider attacks. There is an issue with open research here. Because of the inherent risks and uncertainties, cloud-based options require greater creativity. The current designs and tactics for cloud computing require service providers. Client-provider contracts ought to expressly address these security issues.

3.5. Limitations concerning Deep Learning/AI

Decentralised and online, cloud computing services are available to anybody with the right credentials. Due to the fact that firm data is available online via the cloud, many hackers find it appealing to explore systems, find vulnerabilities, and take advantage of them. The combination of artificial intelligence, cloud-based data and resources, cyber security, and security vulnerabilities makes it imperative to sense cyber threats and safety flaws in the cloud earlier than they have a significant detrimental impact. Machines can be taught to do tasks based on prior experiences thanks to artificial intelligence (AI) and deep learning. This gives machines a greater level of intelligence that allows them to recognise and detect cyberattacks. However, a lot of companies still don't know how dangerous cloud environments may be and how important it is to spend money on defence against emerging cyberthreats.

3.6. Obsolete Laws

Regulations that are occasionally out-of-date and irrelevant are relied upon by businesses and cloud service providers. Instead of relying simply on out-of-date legislation, new regulations must be developed to account for the rapidly changing nature of cloud technology and its extensive usage on the internet. Everyone working with cloud-based systems needs to understand the dangers that come with cloud computing and the steps users take to reduce those risks. Software development companies frequently undervalue the importance of providing cloud development teams with adequate, requirement-driven security training when developing applications. Inadequately controlled access, incorrectly configured cloud storage, and exposed APIs are further security issues. The task of finding workable and affordable answers to these problems falls on researchers. Subsequent cloud safety guidelines are vital for businesses to avoid financial and reputational harm.

3.7. Security Policy Issues

Security policies are guidelines that specify the safety measures implemented to prevent intrusions. It is expected that security policies or standards will safeguard the cloud's operating environment without sacrificing reliability or effectiveness. Along with several service-level agreements (SLAs), preceding confidence, and consumer oversight concerns, these security standards are also subject to various regulatory organizations.

4. CONCLUSION

Our research comprehensively examines the range of security issues, vulnerabilities, attack, furthermore risks that impede the acceptance of cloud computing. We delve into various aspects of cloud security challenges, analyzing the unique characteristics of cloud environments contributing to these concerns. By providing a holistic view of these issues, we stress how critical it is to comprehend the weaknesses in security that cloud computing platforms inherently have and to develop suitable workarounds. Our analysis leads us to recommend a revaluation of present safety protocols with the goal of reducing weaknesses and foiling any attacks. Throughout our investigation, we identify and outline a range of policies, protocols, and methodologies that establish secure management practices within cloud environments. Implementing these measures enables organizations to mitigate risks, enhance resilience, and instill confidence in an increasingly interconnected world. Our exploration also encompasses the safety for cloud-hosted services and information framework, categorizing various security challenges and conducting a comparative examination of recommended countermeasures. Furthermore, we discuss recent advancements in load distribution and task management within cloud computing environments, with a particular focus on Quality of Service (QoS) management. In summary, our study aims to offer a comprehensive understanding of the security landscape in cloud computing, providing practical insights and solutions to address the evolving threats and vulnerabilities. We emphasize the importance of ensuring connectivity and continuity in cloud environments to maintain operational efficiency and security.

References

[1] According to a New IDC Forecast. The Growth in Connected IoT Devices Is Expected to Generate 79.4 ZB of Data in 2025. Accessed: Mar. 1, 2020. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS45213219

[2] Y.-Y. Teing, A. Dehghantanha, K.-K.-R. Choo, and L. T. Yang, ‘‘Forensic investigation of P2P cloud storage services and backbone for IoT networks: BitTorrent sync as a case study,’’ Comput. Electr. Eng., vol. 58, pp. 350–363, Feb. 2017.

[3] B. Varghese and R. Buyya, “Next generation cloud computing: New trends and research directions,” CoRR, vol. abs/1707.07452, 2017.

[4] T. Vasiljeva, S. Shaikhulina, and K. Kreslins, “Cloud computing: business perspectives, benefits and

challenges for small and medium enterprises (case of latvia)," Procedia Engineering, vol. 178, pp. 443–451, 2017.

[5] S. Becker, G. Brataas, M. Cecowski, D. Huljenic, S. Lehrig, and I. Stupar, "Introduction," in Engineering Scalable, Elastic, and Cost-Efficient Cloud Computing Applications - The CloudScale Method, S. Becker, G. Brataas, and S. Lehrig, Eds. Springer, 2017, pp. 3–21.

[6] J. Weinman, "The economics of pay-per-use pricing," IEEE Cloud Comput., vol. 5, no. 5, p. 101, 2018.

[7] M. Bahrami and M. Singhal, "DCCSOA: A dynamic cloud computing service-oriented architecture," in 2015 IEEE International Conference on Information Reuse and Integration, IRI 2015, San Francisco, CA, USA, August 13-15, 2015. IEEE Computer Society, 2015, pp. 158–165

[8] Gartner: Gartner Forecasts Worldwide Public Cloud Revenue to Grow 17% in 2020. Accessed: Feb. 2020. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2019-11-13-gartner-forecastsworldwide-public-cloud-revenue-to-grow-17-percent-in-2020

[9] Parast, F. K., Sindhav, C., Nikam, S., Yekta, H. I., Kent, K. B., & Hakak, S. (2022). Cloud computing security: A survey of service-based models. Computers & Security, 114, 102580.

[10] Barrowclough, J. P., & Asif, R. (2018). Securing cloud hypervisors: a survey of the threats, vulnerabilities, and countermeasures. Security and Communication Networks, 2018, 1-20.

[11] Mansouri, Y., & Babar, M. A. (2021). A review of edge computing: Features and resource virtualization. Journal of Parallel and Distributed Computing, 150, 155-183.

[12] Nemati, H., Azhari, S. V., Shakeri, M., & Dagenais, M. (2021). Host-based virtual machine workload characterization using hypervisor trace mining. ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS), 6(1), 1-25.

[13] Rahman, A., & Subriadi, A. P. (2022, January). Software as a service (SaaS) adoption factors: individual and organizational perspective. In 2022 2nd International Conference on Information Technology and Education (ICIT&E) (pp. 31-36). IEEE.

[14] Isharufe, W., Jaafar, F., & Butakov, S. (2020, June). Study of security issues in platform-as-a-service (PaaS) cloud model. In 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE.

[15] [15] Nazarov, A. N., Koupaei, A. N. A., Dhoot, A., Azlan, A., & Siadat, S. M. R. (2020, March). Mathematical modelling of infrastructure as a service. In 2020 Systems of Signals Generating and Processing in the Field of on Board Communications (pp. 1-6). IEEE.

[16] Saxena, S., Yagyasen, D., Saranya, C. N., Boddu, R. S. K., Sharma, A. K., & Gupta, S. K. (2021, October). Hybrid Cloud Computing for Data Security System. In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-8). IEEE.

[17] Shafiq, D. A., Jhanjhi, N. Z., Abdullah, A., & Alzain, M. A. (2021). A load balancing algorithm for the data centres to optimize cloud computing applications. IEEE Access, 9, 41731-41744.

[18] Pinedo, M. L. (2012). Scheduling, Chapter 7.

[19] Murad, S. A., Muzahid, A. J. M., Azmi, Z. R. M., Hoque, M. I., & Kowsher, M. (2022). A review on job scheduling technique in cloud computing and priority rule based intelligent framework. Journal of King Saud University-Computer and Information Sciences, 34(6), 2309-2331.

[20] Aktan, M. N., & Bulut, H. (2022). Metaheuristic task scheduling algorithms for cloud computing environments. Concurrency and Computation: Practice and Experience, 34(9), e6513.

[21] Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W., Alyami, H., & Ayaz, M. (2021). A systematic literature review on cloud computing security: threats and mitigation strategies. IEEE Access, 9, 57792-57807.

[22] Fathalla, A., Li, K., & Salah, A. (2022). Best-KFF: a multi-objective preemptive resource allocation policy for cloud computing systems. Cluster Computing, 25(1), 321-336.

[23] Kumar, S. M., Senthil, P., Sabitha, E., Kumar, P. S., Balasundaram, A., & Ashokkumar, S. (2020, September). Pre-emptive Approach for Resource Scheduling in Cloud Computing. In 2020 International Conference on Smart Electronics and Communication (ICOSEC) (pp. 276-281). IEEE.

[24] Sohani, M., & Jain, S. C. (2021). A predictive priority-based dynamic resource provisioning scheme with load balancing in heterogeneous cloud computing. IEEE access, 9, 62653-62664.

[25] Kaur, A., & Kaur, B. (2022). Load balancing optimization based on hybrid Heuristic-Metaheuristic techniques in cloud environment. Journal of King Saud University-Computer and Information Sciences, 34(3), 813-824

[26] Alghamdi, M. I. (2022). Optimization of load balancing and task scheduling in cloud computing environments using artificial neural networks-based binary particle swarm optimization (BPSO). Sustainability, 14(19), 11982

[27] Mishra, K.; Majhi, S.K. A binary Bird Swarm Optimization based load balancing algorithm for cloud

computing environment. Open Comput. Sci. 2021, 11, 146–160.

[28] Junaid, M.; Sohail, A.; Ahmed, A.; Baz, A.; Khan, I.A.; Alhakami, H. A Hybrid Model for Load Balancing in Cloud Using File Type Formatting. IEEE Access 2020, 8, 118135–118155.

[29] Yadav, M., & Gupta, S. (2020). Hybrid meta-heuristic VM load balancing optimization approach. Journal of Information and Optimization Sciences, 41(2), 577-586

[30] Annie Poornima Princess, G., & Radhamani, A. S. (2021). A hybrid meta-heuristic for optimal load balancing in cloud computing. Journal of grid computing, 19(2), 21.

[31] Kakkottakath Valappil Thekkepuryil, J., Suseelan, D. P., & Keerikkattil, P. M. (2021). An effective meta-heuristic based multi-objective hybrid optimization method for workflow scheduling in cloud computing environment. Cluster Computing, 24(3), 2367-2384

[32] Khan, A. R. (2024). Dynamic Load Balancing in Cloud Computing: Optimized RL-Based Clustering with Multi-Objective Optimized Task Scheduling. Processes, 12(3), 519.

[33] Ghafir, S., Alam, M. A., Siddiqui, F., & Naaz, S. (2024). Load balancing in cloud computing via intelligent PSO-based feedback controller. Sustainable Computing: Informatics and Systems, 41, 100948.

[34] Gupta, S., & Singh, R. S. (2024). User-defined weight based multi objective task scheduling in cloud using whale optimisation algorithm. Simulation Modelling Practice and Theory, 102915.

[35] Thilak, K. D., Devi, K. L., Shanmuganathan, C., & Kalaiselvi, K. (2024). Meta-heuristic Algorithms to Optimize Two-Stage Task Scheduling in the Cloud. SN Computer Science, 5(1), 1-16.

[36] Apat, H. K., Sahoo, B., Goswami, V., & Barik, R. K. (2024). A hybrid meta-heuristic algorithm for multi-objective IoT service placement in fog computing environments. Decision Analytics Journal, 10, 100379.

[37] Khan, Z. A., Aziz, I. A., Osman, N. A. B., & Nabi, S. (2024). Parallel Enhanced Whale Optimization Algorithm for Independent Tasks Scheduling on Cloud Computing. IEEE Access.

[38] Elsakaan, N., & Amroun, K. (2024). A novel multi-level hybrid load balancing and tasks scheduling algorithm for cloud computing environment. The Journal of Supercomputing, 1-41.

[39] Ge, C., Susilo, W., Liu, Z., Xia, J., Szalachowski, P., & Fang, L. (2020). Secure keyword search and data sharing mechanism for cloud computing. IEEE Transactions on Dependable and Secure Computing, 18(6), 2787-2800.

[40] Agarwal, N., Rana, A., Pandey, J. P., & Agarwal, A. (2020). Secured sharing of data in cloud via dual authentication, dynamic unidirectional PRE, and CPABE. International Journal of Information Security and Privacy (IJISP), 14(1), 44-66.

[41] Liu, Y., Zhou, T., Yue, Z., Liu, W., Han, Y., Li, Q., & Yang, X. (2021). Secure and efficient online fingerprint authentication scheme based on cloud computing. IEEE Transactions on Cloud Computing, 11(1), 564-578.

[42] Tolba, Z., & Derdour, M. (2021, October). Deep learning for cryptanalysis attack on IoMT wireless communications via smart eavesdropping. In 2021 International Conference on Networking and Advanced Systems (ICNAS) (pp. 1-6). IEEE.

[43] Malik, M., Kumar, M., Kumar, V., Gautam, A. K., Verma, S., Kumar, S., & Goyal, D. (2022). High level browser security in cloud computing services from cross site scripting attacks. Journal of Discrete Mathematical Sciences and Cryptography, 25(4), 1073-1081.

[44] Park, J., Kim, D. S., & Lim, H. (2020). Privacy-preserving reinforcement learning using homomorphic encryption in cloud computing infrastructures. IEEE Access, 8, 203564-203579.

[45] Mei, Q., Xiong, H., Chen, Y. C., & Chen, C. M. (2022). Blockchain-enabled privacy-preserving authentication mechanism for transportation CPS with cloud-edge computing. IEEE Transactions on Engineering Management.

[46] Abirami, P., & Bhanu, S. V. (2020). Enhancing cloud security using crypto-deep neural network for privacy preservation in trusted environment. Soft Computing, 24(24), 18927-18936.

[47] Salim, M. M., Kim, I., Doniyor, U., Lee, C., & Park, J. H. (2021). Homomorphic encryption based privacy-preservation for iomt. Applied Sciences, 11(18), 8757.

[48] Ghayvat, H., Pandya, S., Bhattacharya, P., Zuhair, M., Rashid, M., Hakak, S., & Dev, K. (2021). CP-BDHCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications. IEEE Journal of Biomedical and Health Informatics, 26(5), 1937-1948.

[49] Shukla, S., & Patel, S. J. (2022). A novel ECC-based provably secure and privacy-preserving multi-factor authentication protocol for cloud computing. Computing, 104(5), 1173-1202.

[50] Huang, Q., Zhang, Z., & Yang, Y. (2020). Privacy-preserving media sharing with scalable access control and secure deduplication in mobile cloud computing. IEEE Transactions on Mobile Computing, 20(5), 1951-1964.

[51] Liu, Y., Zhou, T., Yue, Z., Liu, W., Han, Y., Li, Q., & Yang, X. (2021). Secure and efficient online fingerprint authentication scheme based on cloud computing. IEEE Transactions on Cloud Computing, 11(1), 564-578.

[52] Xiong, S., Ni, Q., Wang, L., & Wang, Q. (2020). SEM-ACSIT: secure and efficient multiauthority access control for IoT cloud storage. IEEE Internet of Things Journal, 7(4), 2914-2927.

[53] Prince, P. B., & Lovesum, S. J. (2020). Privacy enforced access control model for secured data handling in cloud-based pervasive health care system. SN Computer Science, 1(5), 239.

[54] Saini, A., Zhu, Q., Singh, N., Xiang, Y., Gao, L., & Zhang, Y. (2020). A smart-contract-based access control framework for cloud smart healthcare system. IEEE Internet of Things Journal, 8(7), 5914-5925.

[55] BenMarak, O., Naanaa, A., & Elasmi, S. (2024, April). A Security Evaluation of Chaos Attribute-Based Access Control (ABAC) for Cloud Computing. In International Conference on Advanced Information Networking and Applications (pp. 415-425). Cham: Springer Nature Switzerland.

[56] Han, D., Zhu, Y., Li, D., Liang, W., Souri, A., & Li, K. C. (2021). A blockchain-based auditable access control system for private data in service-centric IoT environments. IEEE Transactions on Industrial Informatics, 18(5), 3530-3540.

[57] Qin, X., Huang, Y., Yang, Z., & Li, X. (2021). LBAC: A lightweight blockchain-based access control scheme for the internet of things. Information sciences, 554, 222-235.

[58] Hwang, Y. W., & Lee, I. Y. (2020). CP-ABE access control that block access of withdrawn users in dynamic cloud. KSII Transactions on Internet and Information Systems (TIIS), 14(10), 4136-4156.

[59] Bera, S., Prasad, S., Rao, Y. S., Das, A. K., & Park, Y. (2023). Designing attribute-based verifiable data storage and retrieval scheme in cloud computing environment. Journal of Information Security and Applications, 75, 103482.

[60] Dhal, K., Rai, S. C., Pattnaik, P. K., & Tripathy, S. (2022). CEMAR: a fine grained access control with revocation mechanism for centralized multi-authority cloud storage. The Journal of Supercomputing, 78(1), 987-1009.

[61] Javed, I. T., Alharbi, F., Bellaj, B., Margaria, T., Crespi, N., & Qureshi, K. N. (2021, June). Health-ID: A blockchain-based decentralized identity management for remote healthcare. In Healthcare (Vol. 9, No. 6, p. 712). MDPI.

[62] Shrihari, M.R., Ajay, N., Shwetha, B.V., Mohan, H.S., Muniraju, M.(2024). Development of Fragmentation to Create a Big Data Security Framework using Blockchain Innovation, IJISAE, 12(1), pp. 492–499

[63] Ajay, N., Mohan, H.S., Shwetha, B.V., Manjunath, P.V., Anitha, T.N.(2022). Access Control Framework in the Cloud based on Multi-Blockchain with Light Privacy Protection, IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics.