CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

In this chapter, the aim was to explore and analyze the current landscape of secure cloud technologies which focused on malware analyzing and AES encryption. It also included information about the existence of cloud and technology considerations about encryption and malware. This review examines existing methods for malware analyzing and encryption methods, identifying gaps in current practices and laying the groundwork for a comprehensive system to enhance the security and usability of cloud storage solutions.

2.2 Cloud Computing

Cloud Computing has become crucial technology since it provides various services like distributed applications, Ecommerce, Social Media Streaming services, Email activities, Storing files and Banking services. Cloud storage is a way of storage mechanism of automated digital data storage to store that information remotely and save different types of data securely so that it can be accessed at anytime from anywhere. The aim is to prevent Data leakage, intellectual property theft, malware attacks, control Hackers and also control insecure access points. However, the selection of an appropriate cloud model whether private, public, hybrid, or community cloud depends on factors such as security, performance, and cost. Each deployment model has unique strengths and limitations, making it critical to understand their characteristics to maximize efficiency and safeguard data (Kavuri & Anithaashri, 2023).

2.2.1 Private Cloud

Private clouds are designed exclusively for a single organization, ensuring full control over data, infrastructure, and security. They are commonly used in industries like healthcare, government, and finance, where regulatory compliance and data confidentiality are priorities. Organizations using private clouds can customize security protocols, such as firewalls and encryption, to meet specific needs. While this model offers robust data protection and operational control, it is associated with higher costs due to the dedicated infrastructure and maintenance required (IBM, 2020; Kavuri & Anithaashri, 2023)

2.2.2 Public Cloud

Public clouds operate in a multi-tenant environment, where services are shared among numerous users. Managed by third-party providers like AWS and Google Cloud, they offer affordability, scalability, and rapid deployment. These attributes make public clouds an attractive option for non-sensitive data storage and applications. However, the shared nature of public clouds introduces security risks, such as data breaches and unauthorized access. To mitigate these concerns, providers implement encryption and authentication protocols. Despite these challenges, the public cloud model remains widely used for its cost-effectiveness and accessibility (IBM, 2020; Kavuri & Anithaashri, 2023).

2.2.3 Hybrid Cloud

Hybrid clouds integrate private and public cloud environments, providing the benefits of both. This model allows organizations to manage sensitive data securely in private clouds while leveraging the scalability of public clouds for less critical workloads. Hybrid clouds are ideal for handling dynamic workloads, enabling seamless transitions between environments through methods like cloud bursting. However, managing hybrid infrastructures can be complex and requires sophisticated tools for integration and orchestration (IBM, 2020; Kavuri & Anithaashri, 2023).

2.2.4 Community Cloud

Community clouds are tailored for a specific group of organizations with shared interests, such as academic institutions or industry consortia. These clouds facilitate collaboration while ensuring data security and compliance with sector-specific regulations. Community clouds strike a balance between privacy and cost-efficiency, as resources are shared among members. However, challenges include governance complexities and resource allocation among participants. Despite these limitations, community clouds offer significant benefits for groups with aligned objectives (Kavuri & Anithaashri,2023).

Table 2. 1 Comparison between cloud-type

Parameters\Type	Public Cloud	Private Cloud	Hybrid Cloud	Community Cloud
Description	In public cloud, services are available for public users.	Private cloud is build up with existing private infrastructure. This type of cloud has some authentic users who can dynamically provision the resources.	Hybrid cloud is a heterogeneous distributed system, resulting from a private cloud, which incorporates different types of services and resources from public clouds.	Different types of cloud are integrated together to meet a common or particular need for some organizations.
Scalability	Very High	Limited	Very High	Limited
Reliability	Moderate	Very High	Medium to High	Very High
Security	Totally Depends on service provider	High class security	Secure	Secure
Performance	Low to medium	Good	Good	Very Good
Cost	Cheaper	High Cost	Costly	Costly
Examples	Amazon EC2, Google AppEngine	VMWare, Microsoft, KVM, Xen	IBM, HP, VMWare vCloud, Eucalyptus	SolaS Community Cloud, VMWare

Table 2.1 compares four types of cloud models: Public, Private, Hybrid, and Community Cloud. Public Cloud is open to everyone, offering high scalability and low cost (e.g., Amazon EC2). Private Cloud is for specific users, with high security and reliability but higher costs (e.g., VMware). Hybrid Cloud combines public and private clouds, providing high scalability, good performance, and higher costs (e.g., IBM vCloud). Community Cloud is shared by organizations for common purposes, with high reliability and security but limited scalability and high cost (e.g., SolaS Community Cloud).

2.3 File Storage and Cloud

Cloud and file storage are integral to modern data management, each offering distinct advantages to meet diverse organizational and individual needs. By integrating the scalability of cloud environments with the hierarchical structure of file storage, users can effectively manage, store, and access their digital assets.

2.3.1 File Storage in Cloud Environments

Cloud storage allows users to store data remotely on central servers accessed via the internet, eliminating the need for local physical storage (Eswari et al., 2023). The efficiency of cloud storage is enhanced by mechanisms like dividing files into smaller blocks, which facilitates easier management, retrieval, and dynamic updates (Eswari et al., 2023). Advanced data structures, such as the Merkle Hash Tree (MHT) and B+ Tree, are commonly employed to ensure file integrity and optimize data operations (Eswari et al., 2023).

In the realm of file-sharing systems, cloud object storage (COS) provides a high degree of reliability and scalability, supporting features like version control, unique file mapping using MD5 hashes, and efficient space utilization by avoiding file duplication (Zhan et al., 2023). The combination of tree structures for metadata and hashing for file verification forms the backbone of efficient cloud storage solutions (Eswari et al., 2023; Zhan et al., 2023).

2.3.2 Integration of Cloud and File Storage

The integration of cloud services with file storage systems introduces functionalities like real-time data synchronization, collaborative access, and secure data sharing across devices and users (Zhan et al., 2023). Systems designed for such integration often adopt microservices architectures, allowing independent management of storage, file sharing, and user permissions (Zhan et al., 2023).

Additionally, permission levels are essential for secure file operations in shared cloud environments, enabling administrators to control access at granular levels, from viewing to editing files (Zhan et al., 2023). Dynamic operations such as file modification and deletion are streamlined by leveraging structures like B+ trees, ensuring efficiency and scalability even for extensive datasets (Eswari et al., 2023)

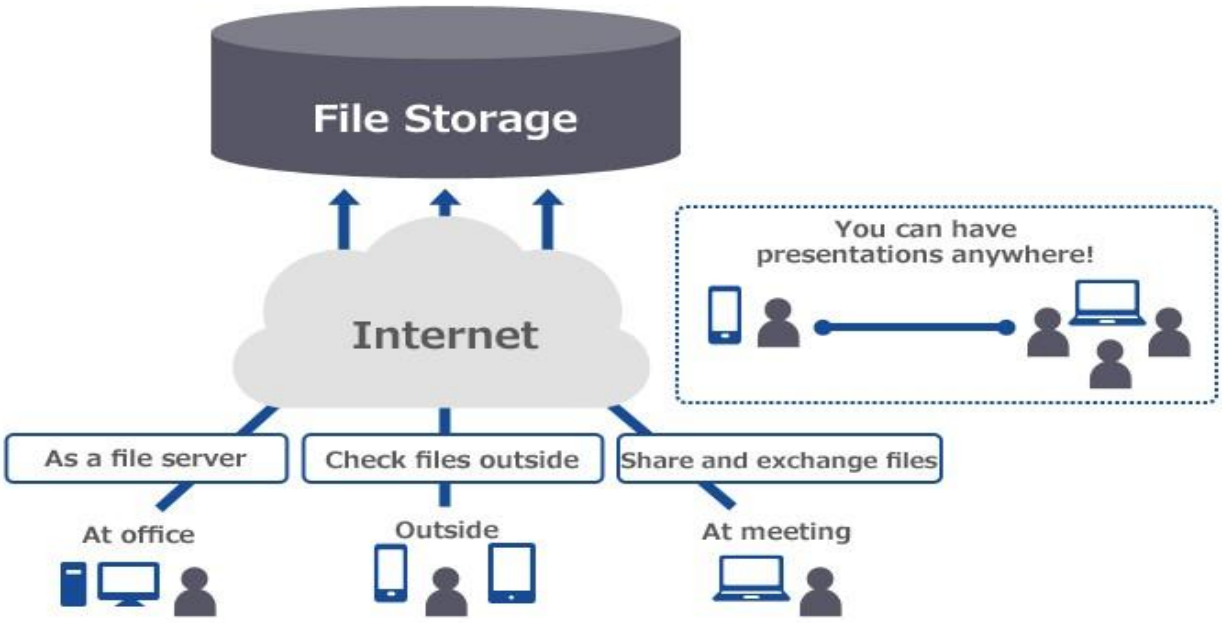

Figure 2. 1 File storage in cloud environment

2.4 Data Breach in Cloud Services

As organizations increasingly adopt cloud services, the incidence of data breaches has emerged as a critical concern (Choudhary et al., 2023). The unique architecture of cloud computing, characterized by shared resources and multi-tenancy, introduces vulnerabilities that can be exploited by malicious actors. Understanding the causes, impacts, and mitigation strategies related to data breaches in cloud services is essential for organizations aiming to safeguard sensitive information, particularly in the context of implementing secure cloud systems that utilize malware analysis and AES encryption (Ashok et al., 2023).

2.4.1 Causes of Data Breaches

Data breaches in cloud environments can be attributed to several factors. Human error is a predominant cause, with misconfigurations of cloud settings often leading to unintended exposure of sensitive data (Choudhary et al., 2023). For instance, a lack of proper access controls can allow unauthorized users to gain access to confidential information, which is a significant risk in cloud systems where multiple users share resources (Ashok et al., 2023). Insider threats, whether intentional or accidental, also contribute significantly to data breaches, as employees may inadvertently leak sensitive information or may be coerced into providing access (Choudhary et al., 2023).

Moreover, cybercriminals frequently exploit vulnerabilities in cloud security through tactics such as phishing and social engineering (Choudhary et al., 2023). The dynamic nature of technology means that new vulnerabilities can emerge rapidly, often outpacing the ability of organizations to implement effective defenses (Ashok et al., 2023). In this context, integrating a malware analyzer can help identify and mitigate these threats in real-time, enhancing the overall security posture of cloud systems (Choudhary et al., 2023). By analyzing potential malware threats, organizations can proactively address vulnerabilities before they are exploited.

2.4.2 Impacts of Data Breaches

The consequences of data breaches in cloud services are extensive and multifaceted. Financially, organizations can incur significant costs related to breach remediation, legal fees, and potential regulatory fines (Choudhary et al., 2023). For example, the 2017 Equifax breach, which exposed the personal information of approximately 147 million individuals, resulted in costs exceeding $4 billion for the company (Kaur&Kaimal,2023).

Beyond financial implications, data breaches can severely damage an organization's reputation, leading to a loss of customer trust and loyalty (Choudhary et al., 2023). Affected individuals may experience identity theft, emotional distress, and a long-term sense of vulnerability regarding their personal information (Ashok et al., 2023). Implementing AES encryption can significantly mitigate these impacts by ensuring that even if data is accessed, it remains unreadable without the appropriate decryption key (Choudhary et al., 2023). This encryption method serves as a robust defense mechanism, protecting sensitive data from unauthorized access and reducing the potential fallout from a breach.

2.4.3 Mitigation Strategies

To effectively mitigate the risks associated with data breaches in cloud services, organizations must adopt a comprehensive approach to security (Choudhary et al., 2023). Implementing robust encryption protocols, such as AES, for data both at rest and in transit is crucial to protecting sensitive information from unauthorized access (Ashok et al., 2023). Regular security audits and vulnerability assessments can help organizations identify and rectify weaknesses in their cloud configurations (Choudhary et al., 2023). These proactive measures are essential for maintaining a secure cloud environment.

Additionally, fostering a culture of security awareness among employees through ongoing training programs can significantly reduce the likelihood of human error (Choudhary et al., 2023). Organizations should also implement multi-factor authentication and strict access controls to enhance security measures (Ashok et al., 2023). Furthermore, integrating a malware analyzer can provide proactive threat detection and response capabilities, further strengthening the security framework of cloud systems (Choudhary et al., 2023). By continuously monitoring potential threats, organizations can respond swiftly to incidents, minimizing the risk of data breaches.

2.4.4 Research Gaps and Opportunities

Despite the growing body of literature on cloud security, several research gaps remain (Choudhary et al., 2023). There is a pressing need for empirical studies that quantify the effectiveness of various mitigation strategies, including the use of malware analysis and AES encryption, in real-world scenarios (Ashok et al., 2023). Additionally, research exploring the psychological impacts of data breaches on end-users is limited, presenting an opportunity for future investigations (Choudhary et al., 2023).

Moreover, the evolving nature of cyber threats necessitates ongoing research into emerging technologies and their potential to enhance cloud security (Ashok et al., 2023). As organizations continue to adopt cloud services, understanding the implications of these technologies on data security will be crucial for developing effective regulatory frameworks and best practices (Choudhary et al., 2023).

Data breaches in cloud services present significant challenges that require a multifaceted approach to address (Choudhary et al., 2023). By understanding the causes, impacts, and mitigation strategies, organizations can better protect their sensitive data and build resilience against future breaches. The integration of malware analysis and AES encryption into secure cloud systems is essential for enhancing data protection and ensuring compliance with evolving security standards (Ashok et al., 2023). The ongoing evolution of technology and the regulatory landscape underscores the importance of continuous research and adaptation in the field of cloud security (Choudhary et al., 2023).

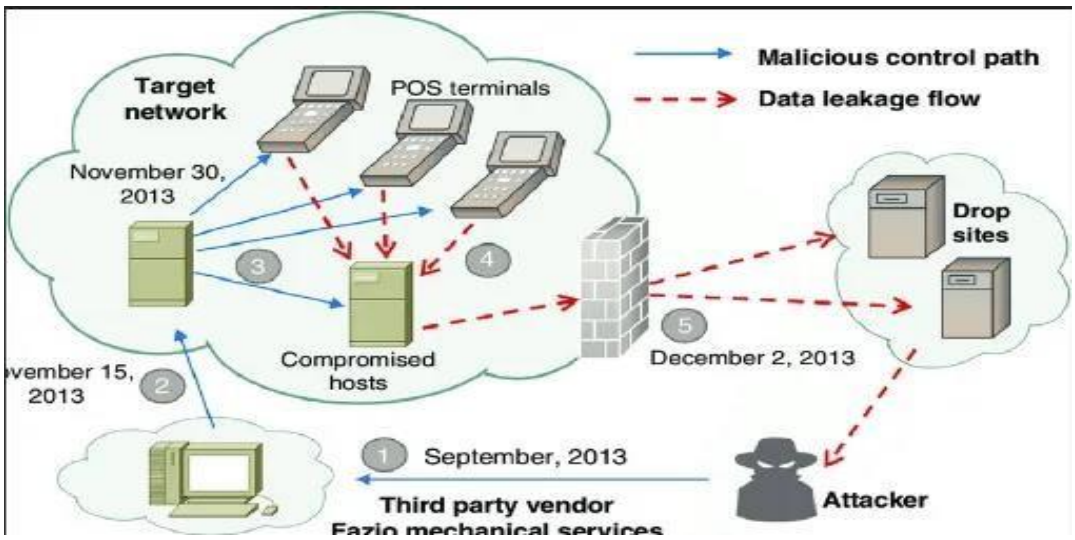

Figure 2. 2 Cloud breach scenarios and attack vector

2.5 Malware in Cloud Storage

Malware, or malicious software, represents one of the most significant challenges in modern computing environments, with cloud storage systems being particularly vulnerable. Cloud storage offers organizations and individuals a convenient, scalable, and cost-effective way to store, access, and share data. However, its interconnected and accessible nature makes it an attractive target for malicious actors. Malware in cloud storage often infiltrates these systems via disguised malicious files, posing a serious threat to data integrity, user privacy, and organizational security. The growing complexity of such attacks necessitates advanced mitigation strategies, including the integration of sophisticated technologies such as malware analyzer APIs. This review examines the operations of malware within cloud storage, the types of malware commonly targeting these platforms, their impact, and the role of malware analyzer APIs in combating these threats.

2.5.1 How Malware Exploit in Cloud Storage

Malware in cloud storage exploits the collaborative and open-access features of these platforms, taking advantage of their architecture to propagate and execute harmful activities. Attackers often introduce malware into cloud storage systems by uploading infected files, either directly or by tricking users into doing so through phishing attacks or compromised software. These files, when accessed or downloaded, infect devices and systems, leading to further compromises. In some cases, malware uses cloud APIs to automate its propagation across the entire storage infrastructure or to other connected systems, amplifying its impact (Kaipu et al., 2023).

One of the primary ways malwares evades detection is through obfuscation techniques. For instance, malicious payloads may be embedded within seemingly harmless files, making it difficult for conventional antivirus solutions to detect them. Additionally, some malware takes advantage of cloud's ability to synchronize files across multiple devices, spreading quickly to all connected systems. The dynamic nature of cloud environments further complicates detection, as malware can exploit the scalability and elasticity of these platforms to remain concealed and operate across different virtualized resources (Shan & Channu, 2023).

2.5.2 Types of Malware Files in Cloud Storage

The types of malware files targeting cloud storage are diverse and evolving. Among the most common are ransomware, Trojans, worms, and spyware. Ransomware is particularly devastating in cloud storage environments, as it encrypts files and demands payment for their release. This type of malware can paralyze organizations that rely on cloud systems to store critical data, causing severe financial and operational disruption. The collaborative nature of cloud storage amplifies the risks, as a single infected file in a shared folder can render all related files inaccessible to users (Samuel et al., 2023). Figure 2.3 below shows how ransomware works to the victim by having malicious code in file.

Trojans are another prevalent threat in cloud storage. Often disguised as legitimate files or software, Trojans infiltrate systems when users unknowingly download or execute them. Once active, they can steal sensitive information, modify data, or provide attackers with backdoor access to cloud systems. Similarly, worms leverage the interconnected structure of cloud storage to replicate and spread rapidly, infecting other files, folders, and even accounts connected to the same system.

Spyware, designed to covertly monitor user activity, often targets sensitive data stored in the cloud, such as financial records, login credentials, or proprietary business information. The stealthy nature of spyware allows it to remain undetected for extended periods, causing significant data breaches. These various types of malware exploit the features that make cloud storage efficient—such as file sharing, scalability, and remote access—to magnify their impact (Kaipu et al., 2023).

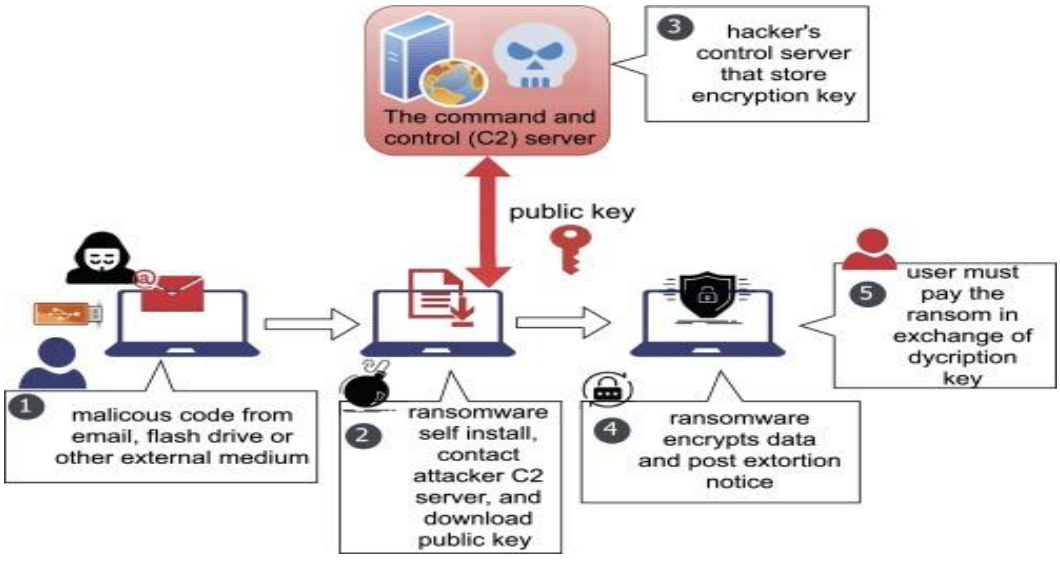

Figure 2. 3 Ransomware attack scenario

2.5.3 The Impact of Malware on Cloud Storage

The consequences of malware in cloud storage can be far-reaching and severe, affecting both individuals and organizations. One of the most immediate impacts is the compromise of sensitive data. Malware can steal confidential information, including personal details, financial records, intellectual property, and business-critical files. For organizations, this often results in financial losses, regulatory fines, and damage to reputation. The operational impact can also be significant, particularly in ransomware attacks where encrypted files disrupt workflows and require costly recovery efforts (Kaipu et al., 2023).

Another key concern is the rapid spread of malware within cloud environments. A single malicious file uploaded to a shared directory can infect all users with access, creating a cascade of compromises that extend beyond the initial target. This amplifies the scale of attacks, making containment and recovery more challenging. Moreover, cloud malware often serves as a staging ground for more sophisticated attacks, such as Distributed Denial of Service (DDoS) attacks or credential theft, further increasing the damage potential (Shan & Channu, 2023).

The dynamic and distributed nature of cloud storage complicates traditional security measures. Malware can exploit the elasticity of cloud systems to migrate across virtual machines or containers, bypassing detection tools and remain operational for extended periods. Additionally, attackers may leverage cloud APIs to manipulate files or automate malicious processes, further hindering security efforts. These challenges highlight the need for more advanced and integrated solutions to combat malware effectively in cloud storage environments.

2.5.4 Mitigation Strategies for Malware in Cloud Storage

Mitigating malware in cloud storage requires a multi-layered approach that combines prevention, detection, and response strategies. One of the most promising developments in this area is the integration of malware analyzer APIs into cloud storage systems. These APIs provide advanced detection capabilities, enabling real-time analysis and mitigation of malicious files before they can cause significant damage.

Malware analyzer APIs use sophisticated algorithms, including behavior-based analysis, to evaluate the actions of files uploaded to the cloud. Unlike signature-based detection methods, which rely on predefined patterns of known threats, behavior-based analysis examines the actual behavior of files to identify anomalies. For example, malware analyzer APIs can detect suspicious activities such as unauthorized file encryption, unusual access patterns, or attempts to execute hidden commands within cloud systems (Samuel et al., 2023).

The integration of malware analyzer APIs into cloud platforms enhances the security infrastructure by providing continuous monitoring and automated responses to detected threats. When a potentially malicious file is uploaded, the API scans it dynamically, flagging or quarantining it if suspicious behavior is detected. This not only prevents the file from spreading but also reduces the risk of data breaches or system compromises (Kaipu et al., 2023).

Other critical mitigation strategies include the use of encryption to protect sensitive data, regular updates and patch management to address vulnerabilities, and stringent access controls such as multi-factor authentication (MFA). Encryption ensures that even if attackers gain access to cloud-stored files, the data remains unreadable without the appropriate decryption keys. Similarly, access controls limit the ability of malware to move laterally within the system, reducing the potential damage of an infection.

2.6 Encryption Algorithm

Encryption algorithms are essential for safeguarding data in digital systems, providing security for data at rest, in transit, and during processing. They form the foundation of modern cryptography, ensuring confidentiality, integrity, and authentication. Encryption techniques can be broadly categorized into symmetric and asymmetric algorithms, each suited to specific use cases based on their operational mechanisms and security features (Shilin et al., 2023; Mallouli et al., 2020).

2.6.1 Advanced Encryption Standard (AES)

The Advanced Encryption Standard (AES) is a symmetric encryption algorithm widely regarded as a cornerstone of modern data security. AES operates on fixed block sizes of 128 bits, with key lengths of 128, 192, or 256 bits, enabling a high level of customization and security. Its substitution-permutation structure ensures robustness against differential and linear cryptanalysis, making it highly secure for a range of applications, from secure file storage to communication systems (Shilin et al., 2023). The efficiency of AES is another notable feature, with its relatively fast encryption and decryption processes making it suitable for environments requiring both performance and security. Recent studies have explored enhancements in AES, such as dynamic key generation, to further increase its resistance to attacks while maintaining its computational efficiency (Awasthi & Kohli, 2023).

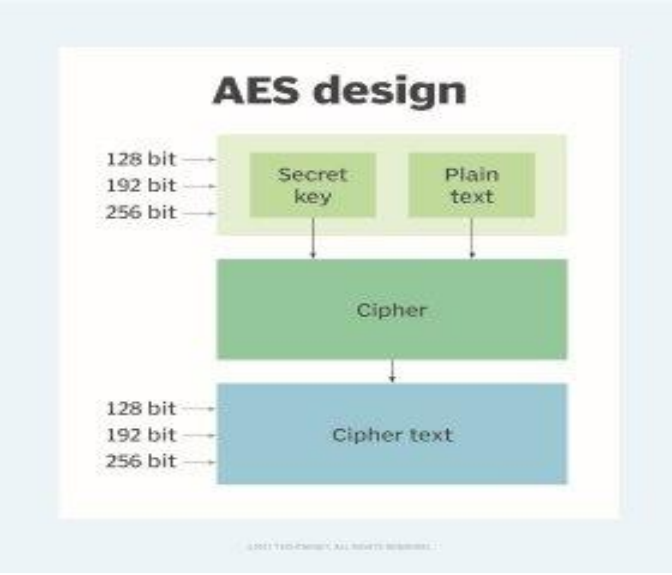

Figure 2. 4 AES encryption architecture design

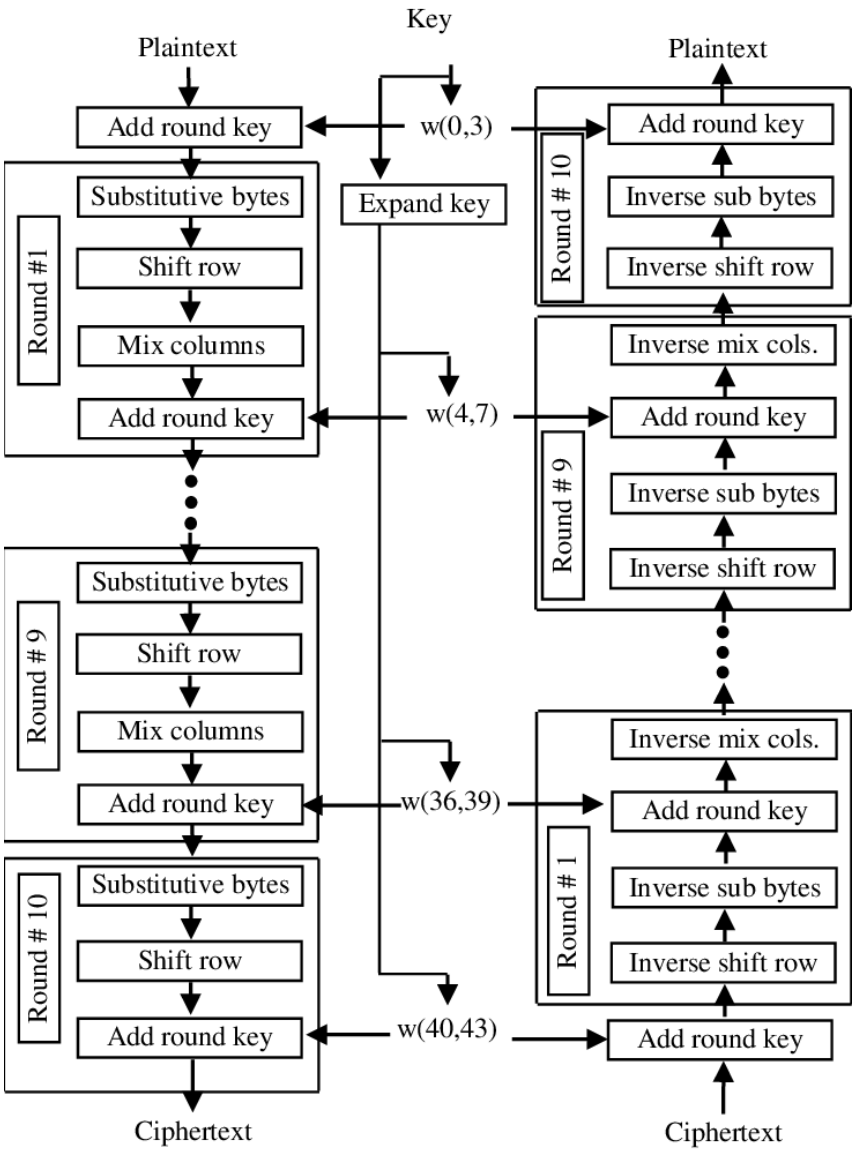

Figure 2. 5 AES Encryption and Decryption process diagram

Figure 2.4 and 2.5 Illustrate overview on how AES encryption works in storing data and files, including encryption and decryption process to maintain a secure encrypted data in system environment .

2.6.2 Data Encryption Standard

The Data Encryption Standard (DES) was once the gold standard in cryptography, operating on 64-bit blocks with a 56-bit key. Despite its initial success, DES is now considered obsolete due to its vulnerability to brute-force attacks, especially given the advancements in computational power. Nevertheless, DES played a crucial role in the evolution of encryption technologies by introducing a structured approach to secure data communication (Awasthi & Kohli, 2023). The development of DES laid the groundwork for more secure algorithms such as Triple DES (3DES) and AES, which address its key length limitations and enhance resistance to cryptanalysis.

2.6.3 RSA (Rivest-Shamir-Adleman)

The RSA algorithm, named after its creators Rivest, Shamir, and Adleman, is a cornerstone of asymmetric cryptography. RSA utilizes a pair of keys: a public key for encryption and a private key for decryption. Its security lies in the computational difficulty of factoring large prime numbers, making it a preferred method for secure data transmission and digital signatures (Mallouli et al., 2020). Despite its robust security, RSA is computationally intensive, requiring significant resources, particularly when using large key sizes for enhanced security. This limitation has led to its partial replacement by more efficient algorithms like Elliptic Curve Cryptography (ECC) in certain applications (Awasthi & Kohli, 2023; Mallouli et al., 2020).

2.6.4 Elliptic Curve Cryptography (ECC)

Elliptic Curve Cryptography (ECC) represents a significant advancement in asymmetric encryption. Unlike RSA, ECC achieves comparable security with much smaller key sizes; for example, a 256-bit ECC key offers the same security level as a 3072-bit RSA key (Mallouli et al., 2020). This efficiency makes ECC ideal for applications with resource constraints, such as mobile devices, IoT systems, and blockchain technology. ECC's mathematical foundation in the algebraic structure of elliptic curves over finite fields enables its use in secure key exchange, digital signatures, and encryption. Its computational efficiency and scalability have contributed to its growing adoption in modern security protocols (Shilin et al., 2023; Mallouli et al., 2020)

2.6.5 Comparison of Encryption Algorithms and Focus on AES

Encryption algorithms play a pivotal role in securing cloud storage by protecting data during transit and at rest. Symmetric encryption methods like AES and DES offer speed and efficiency, while asymmetric algorithms such as RSA and ECC provide robust mechanisms for key management and authentication. The choice of algorithm depends on the specific security requirements and operational constraints of the cloud storage environment (Shilin et al., 2023; Mallouli et al., 2019).

In the context of cloud storage, AES has become the preferred choice due to its combination of high security and computational efficiency. Compared to DES, AES offers significantly stronger encryption with key lengths of 128, 192, or 256 bits, while DES's 56-bit key is inadequate against modern brute-force attacks (Awasthi & Kohli, 2023). While RSA and ECC are vital for secure key exchange and digital signatures, they are computationally intensive and less suitable for encrypting large datasets typically found in cloud storage systems. For instance, RSA requires key sizes of over 3072 bits to match the security of AES's 256-bit key, which increases computational overhead (Mallouli et al., 2019).

AES's efficiency in handling large volumes of data makes it particularly well-suited for cloud storage applications. Its ability to encrypt and decrypt files with minimal latency ensures seamless integration into cloud workflows. Furthermore, advanced implementations of AES, such as leveraging hardware acceleration and parallel processing, have further enhanced its performance in cloud environments. These advancements allow AES to secure data in transit through encrypted communication channels and protect data at rest stored in cloud servers (Shilin et al., 2023; Awasthi & Kohli, 2023).

Given its adaptability, AES also supports secure multi-tenant environments by encrypting each user's data separately, ensuring isolation and confidentiality. This feature aligns well with the shared nature of public and hybrid cloud infrastructures, where data from multiple users or organizations coexists. As cloud storage continues to evolve, AES remains central to addressing both current and emerging data security challenges. Table 2.2 shows comparison between cryptographic methods to shows how AES is relevant choice for this project.

Table 2. 2 Comparison of Cryptographic Method

	RSA	AES	ECC	DES
ALGORTIHM TYPE	ASYMMETRIC	SYMMETRIC	ASYMMETRIC	SYMMETRIC
KEY LENGTH	2048 TO 4096 BITS	128,192 OR 256 BITS	VARIES	56 BITS
SECURITY	HIGH	HIGH	MODERATE	MODERATE
SPEED	MODERATE TO SLOW	FAST	FAST	FAST
USE CASES	SECURE DATA TRANSMISSION, DIGITAL SIGNATURES	ENCRYPTING ELECTRONIC DATA	ENCRYPTION, SIGNATURES, KEY EXCHANGE	ENCRYTPTION ELECTRONIC DATA

2.7 Data Integrity Techniques and Hashing Algorithms

Data integrity refers to the preservation of data accuracy and consistency, ensuring it remains unaltered during storage, processing, or transmission. Cryptographic hash functions play a crucial role in maintaining integrity by transforming data into fixed-length hash values that are unique to the input. These functions are characterized by their one-way nature, collision resistance, and efficiency, which make them indispensable for secure systems (Shwetha & Premananda, 2021).

2.7.1 MD5 (Message Digest 5)

MD5, developed by Ron Rivest, produces a 128-bit hash value and was once widely used due to its simplicity and speed (Shwetha & Premananda, 2021). However, vulnerabilities to collision attacks, where two different inputs produce the same hash value, have significantly reduced its application in security-critical contexts (Shwetha & Premananda, 2021). MD5 is now primarily used for less secure purposes, such as verifying data integrity in file transfers, as it is unsuitable for applications requiring strong cryptographic guarantees (Sy et al., 2023).

2.7.2 SHA-256

SHA-256, part of the SHA-2 family, is a cryptographic hash function developed by the National Security Agency (NSA). It generates a 256-bit fixed-length hash output and is renowned for its robustness against cryptographic attacks, such as preimage and collision attacks (Shwetha & Premananda, 2021). The algorithm is widely used in applications requiring high levels of security, including blockchain systems, database management, and digital signatures (Sy et al., 2023). Its deterministic nature ensures the same input always produces the same output, while its avalanche effect guarantees that even minor input changes result in significantly different hash values (Sy et al., 2023). Furthermore, SHA-256 complies with established security standards, such as FIPS 180-4, ensuring its acceptance across various industries (Sy et al., 2023).

2.7.3 SHA-1

SHA-1, a predecessor to SHA-256, generates 160-bit hash output and was initially designed for secure applications. However, it has been deprecated due to its susceptibility to collision attacks, where attackers can generate identical hash values for different inputs (Fajar et al., 2023). Despite these vulnerabilities, SHA-1 is still used in legacy systems and applications where security requirements are lower (Fajar et al., 2023).

2.7.4 Comparison of Hash Functions with Focus on SHA-256

When comparing the cryptographic hash functions MD5, SHA-1, and SHA-256, several key differences highlight their varying levels of security, efficiency, and suitability for different applications. MD5, developed in the early 1990s, produces the smallest hash size of 128 bits. Its design prioritizes speed and simplicity, which made it a popular choice for data integrity checks and authentication processes in its early adoption. However, over time, significant vulnerabilities have been discovered in MD5, particularly its susceptibility to collision attacks, where different inputs can produce identical hash outputs. This flaw undermines its utility in secure applications, relegating MD5 to low-security tasks such as non-critical file integrity verification or checksums for data transmission (Shwetha & Premananda, 2021).

SHA-1, introduced as an improvement over MD5, generates a 160-bit hash, offering a higher level of security. While SHA-1 was initially widely adopted in digital signatures, certificates, and encryption protocols, advancements in cryptanalysis have revealed vulnerabilities, including its susceptibility to collision attacks. These weaknesses render SHA-1 unsuitable for modern systems requiring robust cryptographic integrity, though it may still be found in legacy systems with lower security demands (Fajar et al., 2023).

In contrast, SHA-256, part of the SHA-2 family, represents a significant advancement in cryptographic hash function design. By producing a 256-bit fixed-length hash, SHA-256 offers exceptional resistance to both preimage and collision attacks. This robustness is achieved through its larger hash size, which increases the computational difficulty for attackers attempting to reverse-engineer or generate identical hashes. Additionally, SHA-256 exhibits properties such as determinism, where identical inputs consistently produce the same hash, and the avalanche effect, where even minor changes to input data result in drastically different hash outputs. These features make SHA-256 highly reliable for securing sensitive information and detecting unauthorized data modifications (Sy et al., 2023).

Although SHA-256 requires more computational resources than MD5 and SHA-1, its enhanced security justifies its use in critical applications. For instance, SHA-256 is widely employed in securing blockchain transactions, protecting digital signatures, and managing database integrity. Its compliance with modern security standards, such as the Federal Information Processing Standards (FIPS 180-4), further solidifies its status as the preferred choice for secure environments (Shwetha & Premananda, 2021; Sy et al., 2023).

Overall, the comparison of these hash functions underscores the importance of selecting the appropriate algorithm based on security needs. While MD5 and SHA-1 may suffice for tasks requiring minimal security, SHA-256 stands out as the most reliable option for contemporary applications demanding robust data integrity and protection against cryptographic attacks. The algorithm's superior collision resistance, compliance with stringent security protocols, and wide adoption across industries demonstrate its critical role in modern cryptography (Sy et al., 2023; Shwetha & Premananda, 2021). Table 2.3 shows the total comparison between SHA-256, SHA-1 and MD5 hash.

Table 2. 3 Comparison between hash

Feature	SHA-256	SHA-1	MD5
Hash Length	256 bits (32 bytes)	160 bits (20 bytes)	128 bits (16 bytes)
Security Level	Very High	Moderate (vulnerable to collision attacks)	Low (severely broken, collision-prone)
Speed	Slower than MD5 and SHA-1 (more secure)	Faster than SHA-256, slower than MD5	Fastest of the three
Collision Resistance	Excellent (no practical collisions found)	Weak (collisions have been demonstrated)	Very weak (collisions are trivial)
Use Cases	Cryptographic applications, blockchain, SSL/TLS	Legacy systems, outdated standards	Checksums, non-security-critical tasks
Year of Development	2001 (as part of SHA-2 family)	1995	1991
Algorithm Type	Cryptographic hash function	Cryptographic hash function	Cryptographic hash function
Primary Weaknesses	Computationally intensive	Collision vulnerability since 2005	Easily exploited with modern computing
NIST Approval	Approved and recommended	Deprecated by NIST	Deprecated by NIST

2.8 Malware Analyzer

Malware analyzers are critical tools in cybersecurity, designed to detect, analyze, and mitigate malicious software threats (Pandey et al., 2020). They employ various techniques, including static and dynamic analysis, to understand malware behavior and develop effective countermeasures (Sharma & Kaushik, 2021).

2.8.1 Malware Analyzer in Cloud

Cloud-based malware analyzers offer scalable solutions for handling large datasets efficiently (Singh et al., 2019). They utilize deep learning models to adapt to evolving malware patterns and provide high accuracy in threat detection (Wang et al., 2022). Cloud-native frameworks also reduce the computational load on local systems by leveraging remote resources for real-time analysis (Alzahrani et al., 2021). A review highlighted that these systems significantly improve the detection of advanced persistent threats in dynamic environments (Kumar et al., 2020).

2.8.2 API in Malware Analyzer

APIs play a pivotal role in integrating malware analysis capabilities into security workflows (Lee & Kim, 2021). They enable automated submission of suspicious files, retrieval of analysis reports, and real-time integration with other security tools (Chen et al., 2020). For example, API call sequence analysis has been shown to effectively detect malicious software behaviors using machine learning techniques (Patel et al., 2021). Furthermore, APIs enhance the scalability of detection frameworks, allowing organizations to adapt quickly to emerging threats (Zhou & Zhang, 2022).

2.8.3 VirusTotal

VirusTotal aggregates results from multiple antivirus engines to provide comprehensive assessments of files and URLs (Almeida et al., 2020). Its API allows users to automate submissions and retrieve detailed analysis reports, streamlining its integration into organizational workflows (Smith et al., 2021). However, research indicates that VirusTotal's detection rate may vary depending on the configuration and update frequency of the antivirus engines it uses (Jones et al., 2022). Despite these limitations, it remains one of the most popular tools for malware analysis due to its ease of use and extensive database (Gupta et al., 2021).

2.8.4 Hybrid Analysis

Hybrid analysis combines static and dynamic techniques to improve malware detection rates (Rahman et al., 2021). By addressing the limitations of individual methods, it provides deeper insights into malware behavior (Hassan & Ahmed, 2022). Recent studies emphasize the role of hybrid techniques in analyzing polymorphic malware, showcasing their ability to adapt to evolving threats (Sharma et al., 2021). Additionally, hybrid systems integrate memory analysis to capture behavioral nuances, enhancing overall detection accuracy (Kumar et al., 2022).

2.8.5 Cuckoo Sandbox

Cuckoo Sandbox is an open-source tool for dynamic malware analysis, offering detailed insights into system changes, network activity, and API calls (Lopez et al., 2021). It executes suspicious files in isolated virtual environments, monitoring their behavior to identify malicious activities (Ali & Khan, 2020). Research has demonstrated its effectiveness in detecting sophisticated malware samples, particularly those targeting critical infrastructure (Singh et al., 2021). Furthermore, its scalability and customization options make it a preferred choice for advanced threat analysis (Chowdhury et al., 2022).

2.8.6 Comparison of API and Focus on VirusTotal API

When comparing malware analysis APIs, factors such as detection capabilities, integration flexibility, and reporting features are critical (Rahman et al., 2020). VirusTotal's API stands out for its ability to aggregate results from numerous antivirus engines, providing a broad perspective on potential threats (Smith et al., 2021). However, studies indicate that it primarily relies on static analysis, limiting its effectiveness against advanced malware (Jones et al., 2022). In contrast, tools like Cuckoo Sandbox offer dynamic analysis capabilities through their APIs, enabling deeper behavioral analysis (Ali & Khan, 2020). Hybrid analysis systems further enhance detection accuracy by combining static and dynamic methods, making them more robust than VirusTotal for specific use cases (Sharma et al., 2021). Ultimately, the choice of API depends on organizational needs, such as the required depth of analysis and integration scope (Kumar et al., 2020).

Table 2. 4 Comparison between malware analyzer APIs

Feature	VirusTotal API	Hybrid Analysis API	Cuckoo Sandbox API
Type of Analysis	Primarily static; aggregates results from multiple antivirus engines (Smith et al., 2021).	Combines static and dynamic analysis for deeper behavioral insights (Hassan & Ahmed, 2022).	Fully dynamic, with a focus on executing files in a controlled environment (Ali & Khan, 2020).
Detection Capabilities	Detection varies due to engine configurations; strong at identifying known threats (Jones et al., 2022).	Effective at detecting polymorphic and metamorphic malware (Sharma et al., 2021).	Best suited for detecting complex malware behavior in real-time (Chowdhury et al., 2022).
Integration	Easy to integrate; well-documented API for automated submissions and results retrieval (Smith et al., 2021).	Moderately easy to integrate; requires customization for hybrid setups (Rahman et al., 2021).	Requires technical expertise for integration and management (Lopez et al., 2021).
Data Output	Provides detailed antivirus reports with aggregated scores (Gupta et al., 2021).	Offers detailed behavioral insights and risk scoring (Hassan & Ahmed, 2022).	Generates in-depth logs, including file system changes, API calls, and network activity (Ali & Khan, 2020).
Scalability	High; suitable for handling large numbers of files due to cloud support (Almeida et al., 2020).	High; designed for scalable hybrid analysis workflows (Rahman et al., 2021).	Medium; dependent on system resources and sandbox environments (Chowdhury et al., 2022).
Strengths	Aggregates multiple antivirus results; strong reputation database (Gupta et al., 2021).	Comprehensive analysis with both static and dynamic methods (Sharma et al., 2021).	Excellent for advanced behavioral analysis and detection of zero-day threats (Ali & Khan, 2020).
Weaknesses	Limited dynamic analysis capabilities; reliance on external antivirus engines (Jones et al., 2022).	Higher computational overhead; moderately complex integration (Rahman et al., 2021).	Requires significant setup and maintenance for optimal performance (Chowdhury et al., 2022).

2.9 Related works

The first related work, Secure File Storage on Cloud Using Hybrid Cryptography (Sharma et al., 2020), integrates AES and ElGamal encryption with SHA-256 hashing and two-factor authentication (2FA). This combination ensures robust security and user authentication. However, the solution does not address the risk of malware uploads, leaving the storage system vulnerable. The proposed project overcomes this limitation by adding a malware analyzer that proactively detects and prevents malicious files from being uploaded, thus enhancing the overall security framework.

Improved Cloud Storage Security Using Three Layers Cryptography Algorithms (Kumar et al., 2020) introduces a multi-layered encryption approach combining AES, ECC, and RSA to secure data storage. This method offers robust protection and resolves key distribution challenges. However, the computational overhead associated with the multi-layered approach makes it less efficient, particularly for systems with limited resources. Furthermore, the absence of malware detection is a notable gap. The proposed project simplifies encryption by using AES alone to ensure efficiency while incorporating malware detection to protect against malicious files.

Another significant contribution is Securing Cloud Storage Data Using Hybrid AES-ECC Cryptographic Approach (Chhabra & Gupta, 2021), which leverages the efficiency of AES alongside the advanced security of ECC. While this hybrid method is advantageous for resource-constrained environments, it does not include malware detection mechanisms, leaving malicious files encrypted as a potential threat. By integrating a malware analyzer, the proposed project ensures that only secure files are stored, thus complementing the strengths of this hybrid approach while addressing its shortcomings.

Development of a Secure Web-Based Cloud Storage System Using 3D-AES Block Cipher Cryptography Algorithm (Ali & Hassan, 2019) focuses on enhancing the AES encryption process by adding a third dimension to its block cipher. While this improves security, the added complexity increases the implementation challenges and computational overhead, especially for larger datasets. Additionally, the lack of malware detection capabilities exposes the system to threats. The proposed project simplifies encryption by adopting standard AES and compensates for the lack of threat detection by integrating a malware analyzer, thereby ensuring both security and performance**.**

Feistel and S-Box Extensions (2022) enhanced AES with a Feistel structure and S-Box algorithm to improve encryption robustness. Although this approach strengthens security, it slightly increases computational overhead, making it less suitable for large-scale or real-time cloud storage applications. The proposed solution mitigates this limitation by focusing on optimizing AES for performance in time-sensitive cloud environments, ensuring efficient encryption without compromising speed or scalability.

Table 2. 5 Related Works

Title	Key Features	Strengths	Limitations	Proposed Solution
Secure File Storage on Cloud Using Hybrid Cryptography (Sharma et al., 2020)	AES, ElGamal encryption, SHA-256, and 2FA	Strong encryption with 2FA authentication	Lacks malware detection capabilities	Adds malware analyzer to prevent malicious file uploads
Improved Cloud Storage Security Using Three Layers Cryptography Algorithms (Kumar et al., 2020)	AES, ECC, and RSA-based multi-layer encryption	Robust key management and encryption	High computational cost, no malware detection	Simplifies encryption using AES and integrates malware analyzer
Securing Cloud Storage Data Using Hybrid AES-ECC Cryptographic Approach (Chhabra & Gupta, 2021)	Combination of AES and ECC	Optimized for resource-constrained environments	Lacks malware detection mechanisms	Incorporates malware detection before encryption

Title	Key Features	Strengths	Limitations	Proposed Solution
Development of a Secure Web-Based Cloud Storage System Using 3D-AES Block Cipher Cryptography Algorithm (Ali & Hassan, 2019)	Enhanced AES block cipher	Advanced encryption techniques	Complex implementation, no malware detection	Balances security and efficiency with AES and malware analyzer
Enhanced AES Implementation (Feistel & S-Box Extensions 2022)	AES enhanced with Feistel structure and SBox algorithm to increase encryption robustness.	Provides high security for data transmission and storage.	Slightly increases encryption and decryption time.	Focus on algorithmic optimizations to reduce overhead while maintaining encryption strength.

2.10 Summary

The literature review for the project Secure Cloud Storage With Malware Analyzer and AES Encryption synthesizes key advancements in cloud storage security, encryption algorithms, and authentication mechanisms. It highlights the strengths and limitations of existing approaches, paving the way for the proposed system to address identified gaps.

Research on encryption algorithms underscores the dominance of AES due to its high security, efficiency, and adaptability for large-scale data encryption. Enhanced implementations, such as those incorporating Feistel structures and advanced S-Box algorithms, demonstrate improved functionality but face challenges related to increased computational overhead. Comparisons with DES, RSA, and ECC further establish AES as the most suitable choice for cloud storage environments, particularly when optimized for speed and scalability in handling time sensitive data.

Research on malware analysis highlights the importance of static and dynamic techniques for detecting threats in cloud environments. VirusTotal stands out for its use of multiple antivirus engines but relies mainly on static analysis. Hybrid Analysis combines static and dynamic methods for more comprehensive detection, while Cuckoo Sandbox excels in real-time dynamic analysis, offering insights into malware behavior. In terms of APIs, VirusTotal is popular for its ease of integration, while Hybrid Analysis and Cuckoo Sandbox provide deeper behavioral analysis but require more setup. The proposed system seeks to combine these strengths for enhanced features to provide double-layered security and efficient cloud storage security.