



Comparative Study of Data Security Techniques in Distributed Cloud Environments

Isandro Hathurusinghe

Department of Electrical and Computer Engineering, Kelana Technical and Management College, Malaysia
isandro.hathurusinghe@ktmc-my.net

Peer Review Information	Abstract
<i>Submission: 09 Aug 2017</i> <i>Revision: 26 Sep 2017</i> <i>Acceptance: 20 Oct 2017</i> Keywords <i>Distributed cloud, Data security, Attribute-based encryption, Public auditing, Secure deduplication</i>	Distributed cloud environments improve availability, elasticity, and geographic reach by placing data and services across multiple cloud nodes, but distribution also enlarges the attack surface and complicates confidentiality, integrity, access control, key management, and accountability. This methodology paper develops a structured comparative framework for data-security techniques that were available in the literature up to 2017. The study compares conventional encryption with managed keys, attribute-based encryption (ABE), provable data possession and privacy-preserving public auditing, message-locked/convergent encryption for deduplication, and hybrid-cloud authorized deduplication. The comparison uses six criteria: confidentiality strength, integrity assurance, fine-grained access control, scalability across distributed nodes, computational overhead, and communication/storage overhead. A weighted analytical scoring method is used so that heterogeneous techniques can be assessed consistently without claiming unreported experimental measurements. The analysis shows that no single technique satisfies all security and efficiency objectives. Symmetric encryption remains highly efficient for bulk confidentiality, ABE offers superior policy-driven access control at higher computational cost, auditing mechanisms strengthen integrity without replacing encryption, and deduplication-aware cryptography reduces storage cost but introduces distinct leakage and key-management concerns. A layered architecture combining efficient encryption, distributed policy/key control, independent integrity verification, and authorization-aware deduplication provides the most balanced security posture for distributed cloud deployments.

Introduction

Cloud computing shifted computing resources from locally controlled infrastructure toward elastic services hosted by third parties. The cloud model is attractive because resources can be provisioned rapidly and shared at scale, but outsourcing also separates data owners from the physical systems that store and process their information [1]. In a distributed cloud environment, this separation is amplified because replicas, services, or storage blocks can

be placed across multiple data centers, administrative domains, or geographic regions. Distribution improves resilience and latency, yet it creates more trust boundaries, inter-node communication paths, privileged administrators, cryptographic keys, and policy enforcement points that must remain secure.

Early cloud-security studies identified data confidentiality, insecure interfaces, virtualization isolation, authentication, trust, and privacy as

major barriers to cloud adoption [2]-[5]. Multi-tenancy can create information leakage risks when workloads belonging to different customers share physical resources [6]. A distributed deployment additionally raises questions about where encryption and decryption occur, how keys are synchronized or separated across nodes, how access policies follow replicated data, and how a data owner can verify that remote copies remain intact. These requirements mean that a single perimeter-oriented mechanism is insufficient; protection has to continue with the data throughout storage, transfer, replication, retrieval, and deletion.

Cryptography therefore became a central design element for cloud data security. Conventional symmetric encryption can protect bulk data efficiently, but access is generally coarse-grained unless it is combined with a policy and key-management service. Attribute-based encryption (ABE) was introduced to bind cryptographic access directly to descriptive attributes and policies [7], [8]. Remote integrity techniques such as provable data possession (PDP) and privacy-preserving public auditing allow owners or third-party auditors to challenge a cloud server without downloading the complete data set [9], [10]. Cryptographic cloud-storage proposals further combine encryption, searchable or structured protection, and integrity ideas to reduce reliance on the storage provider [11]. Storage optimization creates another security tension: ordinary encryption can defeat deduplication, while message-locked encryption and hybrid authorized deduplication attempt to preserve space savings without completely abandoning confidentiality [12], [13]. By 2017, outsourced ABE schemes also demonstrated that heavy cryptographic work could be shifted to cloud services while preserving fine-grained control [14].

The objective of this paper is to provide a step-by-step methodology for comparing these techniques under a common set of criteria. Rather than treating each technique as a direct substitute for the others, the study identifies the security property each mechanism primarily contributes, measures its relative operational burden, and determines which combinations are suitable for distributed cloud environments. All references used in the study were published in or before 2017, as required.

Literature Review

Armbrust et al. [1] described the architectural and economic foundations of cloud computing and highlighted obstacles created by dependence

on remote infrastructure. Jensen et al. [2] examined technical security issues in early cloud systems, while Takabi et al. [3] emphasized identity management, access control, trust, and privacy challenges caused by cloud service and deployment models. Subashini and Kavitha [4] organized security concerns across SaaS, PaaS, and IaaS, demonstrating that data and application risks vary with the service layer. Zissis and Lekkas [5] proposed a trusted-third-party security approach integrating public-key infrastructure, single sign-on, and directory services. Together, these studies establish that distributed cloud security requires mechanisms at several layers rather than a single cryptographic primitive.

Virtualization and physical co-residency add another security dimension. Ristenpart et al. [6] showed that cloud co-residency can enable information leakage and side-channel opportunities, which is important for distributed environments where data may migrate between physical hosts. Encryption limits the value of stolen storage content, but encryption alone does not determine who is allowed to decrypt. Goyal et al. [7] introduced key-policy ABE and Bethencourt et al. [8] introduced ciphertext-policy ABE, providing fine-grained cryptographic access control. Their approaches are particularly relevant when distributed users are described by roles, departments, projects, sensitivity levels, or geographic constraints.

Integrity verification addresses a different question: whether outsourced data still exists in its expected form. Ateniese et al. [9] proposed PDP, allowing probabilistic proof of remote data possession with low communication overhead. Wang et al. [10] extended the integrity theme with privacy-preserving public auditing and batch verification, enabling a third-party auditor to verify remote cloud data without learning the underlying content. Kamara and Lauter [11] presented cryptographic cloud-storage concepts showing how encryption and related cryptographic services can reduce the trust placed in cloud storage operators.

Deduplication conflicts with randomized encryption because two identical plaintext files normally become different ciphertexts. Bellare et al. [12] formalized message-locked encryption for secure deduplication, while Li et al. [13] proposed authorized deduplication in a hybrid-cloud architecture where privilege-sensitive duplicate checking is separated from public storage. These techniques improve storage and bandwidth efficiency but require careful control of equality leakage and dictionary attacks. Sun et

al. [15] reviewed cloud data security and privacy mechanisms and reinforced the need to combine software- and hardware-oriented protections. Zhang, Ma, and Lu [14] later demonstrated fully outsourced CP-ABE in 2017, reducing client-side cryptographic cost and showing a practical path toward fine-grained access control on resource-constrained devices.

A recurring observation across the reviewed studies is that security guarantees depend on where trust is placed. Centralized key services simplify revocation but become high-value targets; ABE reduces dependence on an online

authorization server but complicates key revocation and cryptographic computation; public auditing transfers verification work to an auditor but requires privacy-preserving proof design; and deduplication systems intentionally reveal some relation between equal files. These trust assumptions are especially important in distributed clouds because an organization may use several providers or regions with different administrators and failure domains. The comparative methodology therefore records not only the nominal security property but also the additional trusted component required to deliver it.

Table 1. Representative security techniques considered in the comparative study

Technique	Primary goal	Strength	Main limitation	Key refs.
Symmetric encryption + KMS	Confidentiality	Very low data-path cost	Key distribution and coarse policies	[2]-[5], [11]
CP/KP-ABE	Confidentiality + access control	Cryptographic fine-grained policy	Pairing/exponentiation overhead; revocation complexity	[7], [8], [14]
PDP / public auditing	Integrity / possession	Remote verification without full download	Does not itself encrypt content	[9], [10]
Message-locked encryption	Confidential deduplication	Retains deduplication benefit	Equality leakage; low-entropy data risk	[12]
Hybrid authorized deduplication	Privilege-aware deduplication	Separates sensitive authorization from storage	Added private-cloud component and protocol complexity	[13]

Methodology

1. Research Design

The study uses a comparative analytical methodology. It does not claim a new cryptographic construction and it does not fabricate cloud-provider benchmark data. Instead, it converts the security and operational characteristics reported in the pre-2018 literature into a common evaluation matrix. The unit of analysis is a security technique deployed for protecting data distributed across more than one cloud node. Five technique families are evaluated: (T1) symmetric encryption with centralized or replicated key management, (T2) ABE for fine-grained access control, (T3) PDP/privacy-preserving public auditing, (T4) message-locked encryption for deduplication, and (T5) hybrid-cloud authorized deduplication.

The assumed system contains users, an identity/policy service, a key-management function, and at least three storage or compute nodes. Data may be replicated between nodes. Threats considered include an honest-but-curious cloud operator, an external attacker

obtaining stored blocks, unauthorized users, accidental or malicious data modification, compromised tenant isolation, and deduplication-based information leakage. Full compromise of all cryptographic keys or all non-colluding security services is outside the base model because no cryptographic system can maintain confidentiality after complete key disclosure.

The data life cycle is divided into creation, encryption, upload, replication, access, audit, and deletion. At creation, the methodology checks whether sensitive data is encrypted before leaving the owner domain. During upload and replication, it checks whether nodes receive plaintext or ciphertext and whether inter-node traffic is protected. At access time, it examines the granularity and revocability of authorization. During audit, it considers whether missing or modified replicas can be detected without retrieving complete objects. During deletion and deduplication, it considers residual copies, duplicate identifiers, and whether content equality can be inferred. This life-cycle view prevents a technique from receiving a high

overall judgment merely because it performs strongly at one stage.

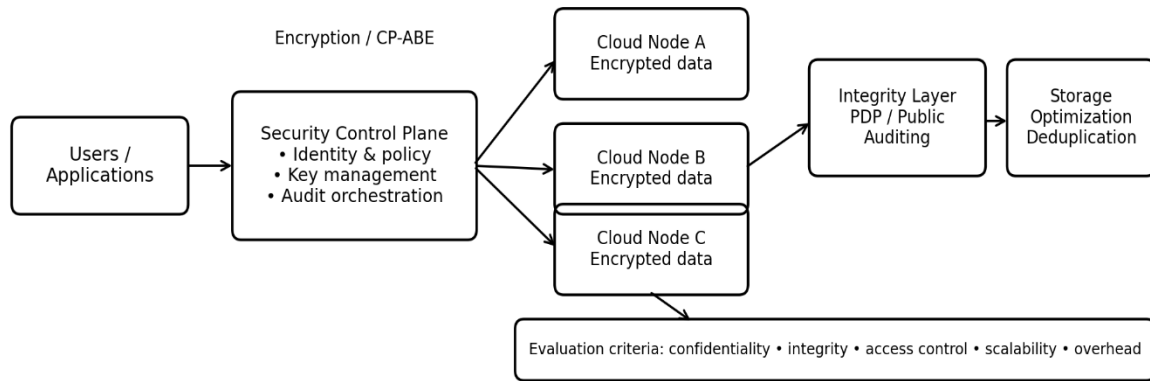


Figure 1. Security evaluation framework for a distributed cloud environment.

2. Evaluation Criteria and Scoring Procedure

Each technique is scored from 1 (weak/unfavorable) to 5 (strong/favorable) on six criteria. For security criteria, a higher score means stronger protection. For overhead criteria, the score is reversed so that 5 represents low overhead and 1 represents high overhead. The criteria and weights are: confidentiality (0.25), integrity assurance (0.20), fine-grained access control (0.20), distributed scalability (0.15), computational efficiency (0.10), and

communication/storage efficiency (0.10). The weights emphasize the security properties most directly associated with outsourced distributed data. A weighted security-efficiency index (WSEI) is calculated as $WSEI = \sum(w_i \times s_i)$, where w_i is the criterion weight and s_i is the 1-5 score. The maximum possible WSEI is 5.00. Scores are assigned from the documented functional properties of each technique and are used only for comparison within this methodology; they are not presented as measured throughput or latency values.

Table 2. Comparative evaluation criteria and weights

Criterion	Weight	Score 5 means	Score 1 means
Confidentiality	0.25	Strong protection against content disclosure	Little/no confidentiality
Integrity assurance	0.20	Strong independent verification	No integrity verification
Fine-grained access	0.20	Cryptographic/policy-level fine granularity	Coarse access only
Distributed scalability	0.15	Fits multi-node deployment with manageable control	Central bottleneck or poor scaling
Computational efficiency	0.10	Low client/server cryptographic burden	High expensive operations
Communication/storage efficiency	0.10	Low network/storage overhead	High metadata, proof, or storage overhead

Step-by-Step Comparative Procedure

Step 1 - Define the distributed data scenario. A data owner stores encrypted objects across three or more nodes and authorizes multiple user groups. Replication and remote verification are permitted.

Step 2 - Map each technique to its primary security property. Encryption protects

confidentiality; ABE adds policy-driven decryption; PDP/public auditing tests integrity and possession; message-locked encryption preserves deduplication; hybrid authorized deduplication adds privilege-aware duplicate checking.

Step 3 - Evaluate confidentiality and access control. The analysis considers whether a cloud operator can read plaintext, whether users need

a shared decryption key, and whether the cryptographic policy can distinguish users or attributes [7], [8], [14].

Step 4 - Evaluate remote integrity. The analysis records whether the technique can detect missing or altered outsourced data without a complete download and whether an external auditor can verify data while preserving privacy [9], [10].

Step 5 - Evaluate distributed overhead. Key distribution, proof traffic, pairing/exponentiation work, additional private-cloud services, and deduplication metadata are considered qualitatively using the published designs [11]-[14].

Step 6 - Assign 1-5 scores and compute WSEI. The same rubric is applied to every technique, making the comparison transparent and reproducible. A sensitivity check is performed conceptually: when access-control weight is increased, ABE rises; when storage efficiency is prioritized, deduplication approaches rise; when

integrity weight is increased, auditing mechanisms rise.

Step 7 - Derive a layered recommendation. Because mechanisms protect different properties, the final architecture is selected by combining complementary techniques rather than choosing the highest single score as a universal winner.

Results

Table 3 presents the analytical scores obtained from the defined rubric. Symmetric encryption with a managed key service achieves the best computational efficiency and strong confidentiality, but its fine-grained access-control score is limited because policy enforcement is external to the ciphertext. ABE produces the strongest access-control result and strong confidentiality, consistent with the purpose of KP-ABE and CP-ABE [7], [8]. Its main trade-off is computational cost, although fully outsourced ABE reported in 2017 reduces the client burden [14].

Table 3. Analytical comparative results (1 = weak/unfavorable, 5 = strong/favorable)

Technique	Conf.	Integrity	Access	Scale	Comp. eff.	Comm./storage eff.	WSEI
T1 Symmetric encryption + KMS	5	2	2	4	5	4	3.55
T2 CP/KP-ABE	5	2	5	4	2	3	3.75
T3 PDP / public auditing	1	5	1	4	4	4	2.85
T4 Message-locked encryption	3	2	2	4	4	5	3.05
T5 Hybrid authorized deduplication	4	2	4	3	3	5	3.45

The WSEI ranking places ABE and symmetric encryption near the top, but their similar aggregate values hide different security roles. This is why a single composite score should not be used as the only design decision. Symmetric encryption is the preferred bulk-data primitive when throughput is dominant, whereas ABE is preferable for data objects whose access must be enforced cryptographically across organizational boundaries. An effective distributed design can use symmetric data-encryption keys for each object and protect or distribute those keys using an attribute-based or policy-controlled mechanism, thereby combining high data-path efficiency with fine-grained authorization.

PDP/public auditing scores very highly for integrity because it directly answers the remote-possession problem [9], [10], yet it receives a low

confidentiality score because the mechanism is not intended to replace encryption. This result illustrates complementarity: an encrypted object can still be audited using appropriate integrity metadata, providing both secrecy and verifiable possession. In distributed clouds, auditing is especially useful for confirming that replicas placed at independent sites remain available and unmodified.

Message-locked encryption improves communication/storage efficiency by making identical plaintext map to deduplicable protected representations, but its deterministic relationship to content exposes equality information and can be risky for predictable, low-entropy files [12]. The hybrid authorized approach improves control by moving privilege-sensitive duplicate checking to a private

component [13], at the cost of an extra trusted service and protocol complexity. Therefore, deduplication-aware cryptography should be restricted to data classes where storage savings justify the additional leakage model.

A sensitivity interpretation further clarifies the ranking. If an enterprise gives the highest weight to transaction throughput and very frequent reads, T1 becomes the clear operational choice for the data plane. If the environment contains many organizations and rapidly changing sharing relationships, T2 becomes more attractive because policy can travel with encrypted data. If regulatory evidence and replica correctness dominate, T3 becomes mandatory even though its standalone confidentiality score is low. For backup and archival workloads with very high redundancy, T4 or T5 can reduce storage cost, but T5 is preferable when duplicate checks must reflect user privileges. Thus, the best technique depends on workload and threat priorities rather than on a universal maximum score.

Overall, the comparative study supports a layered distributed-cloud security architecture: (1) efficient symmetric encryption for all sensitive bulk data, (2) strong key management and ABE or equivalent policy-bound protection for cross-domain authorization, (3) PDP/privacy-preserving auditing for integrity and accountability, (4) isolation and hardening to reduce co-residency risks identified in [6], and (5) authorized deduplication only for suitable data categories. This layered outcome is consistent with survey findings that cloud security is a multi-layer problem involving data, applications, infrastructure, identity, and trust [3]-[5], [15].

Conclusion and Discussion

This paper presented a methodology for comparing data-security techniques in distributed cloud environments using literature available up to 2017. The principal finding is that confidentiality, access control, integrity verification, and storage efficiency are distinct goals and should not be collapsed into a single mechanism. Symmetric encryption with disciplined key management remains the most efficient baseline for bulk confidentiality. ABE extends confidentiality into fine-grained, cryptographically enforced authorization, but it introduces heavier computation and revocation/key-lifecycle challenges. PDP and privacy-preserving public auditing provide an independent way to verify remote data possession and integrity and therefore

complement, rather than replace, encryption. Message-locked encryption and authorized deduplication address storage efficiency, but they require a threat model that explicitly considers equality leakage, predictable files, and privilege management. For distributed deployments, the preferred design is consequently layered: encrypt data at rest and in transit, separate keys and security policy from public storage nodes, use attribute- or role-aware key release for sensitive shared objects, independently audit replicas, and enable deduplication only where its risk is acceptable. The weighted scoring model makes the trade-offs visible, but its numerical values are analytical ratings rather than measured provider performance. Future experimental work can implement the same framework on multiple cloud regions and measure encryption throughput, key-operation latency, audit proof time, bandwidth consumption, policy-update cost, node-failure recovery, and deduplication ratio. Such measurements would convert the methodology into an empirical performance study while preserving the same security criteria.

References

- M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50-58, 2010. DOI: 10.1145/1721654.1721672.
- M. Jensen, J. Schwenk, N. Gruschka, and L. L. Iacono, "On technical security issues in cloud computing," in 2009 IEEE International Conference on Cloud Computing, pp. 109-116, 2009. DOI: 10.1109/CLOUD.2009.60.
- H. Takabi, J. B. D. Joshi, and G.-J. Ahn, "Security and privacy challenges in cloud computing environments," *IEEE Security & Privacy*, vol. 8, no. 6, pp. 24-31, 2010. DOI: 10.1109/MSP.2010.186.
- S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," *Journal of Network and Computer Applications*, vol. 34, no. 1, pp. 1-11, 2011. DOI: 10.1016/j.jnca.2010.07.006.
- D. Zissis and D. Lekkas, "Addressing cloud computing security issues," *Future Generation Computer Systems*, vol. 28, no. 3, pp. 583-592, 2012. DOI: 10.1016/j.future.2010.12.006.

- T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds," in Proc. 16th ACM Conference on Computer and Communications Security, pp. 199-212, 2009. DOI: 10.1145/1653662.1653687.
- V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proc. 13th ACM Conference on Computer and Communications Security, pp. 89-98, 2006. DOI: 10.1145/1180405.1180418.
- J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in 2007 IEEE Symposium on Security and Privacy, pp. 321-334, 2007. DOI: 10.1109/SP.2007.11.
- G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proc. 14th ACM Conference on Computer and Communications Security, pp. 598-609, 2007. DOI: 10.1145/1315245.1315318.
- C. Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage," IEEE Transactions on Computers, vol. 62, no. 2, pp. 362-375, 2013. DOI: 10.1109/TC.2011.245.
- S. Kamara and K. E. Lauter, "Cryptographic cloud storage," in Financial Cryptography and Data Security Workshops, pp. 136-149, 2010. DOI: 10.1007/978-3-642-14992-4_13.
- M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Advances in Cryptology - EUROCRYPT 2013, pp. 296-312, 2013. DOI: 10.1007/978-3-642-38348-9_18.
- J. Li, Y. K. Li, X. Chen, P. P. C. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1206-1216, 2015. DOI: 10.1109/TPDS.2014.2318320.
- R. Zhang, H. Ma, and Y. Lu, "Fine-grained access control system based on fully outsourced attribute-based encryption," Journal of Systems and Software, vol. 125, pp. 344-353, 2017. DOI: 10.1016/j.jss.2016.12.018.
- Y. Sun, J. Zhang, Y. Xiong, and G. Zhu, "Data security and privacy in cloud computing," International Journal of Distributed Sensor Networks, vol. 10, no. 7, 2014. DOI: 10.1155/2014/190903.