IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Optimized Traffic Classification of High Entropy Encrypted Packets

**M.VAISHNAVI**
**Computer Science Engineering**
**Prist University,**
**Thanjavur,India.**

**D.THULASI RAMAN**
**Computer Science Engineering**
**Prist University**
**Thanjavur , India**

## ABSTRACT

Due to the complexity and volume, outsourcing ciphertexts to a cloud is deemed to be one of the most effective approaches for big data storage and access. Nevertheless, verifying the access legitimacy of a user and securely updating a ciphertext in the cloud based on a new access policy designated by the data owner are two critical challenges to make cloud-based big data storage practical and effective. Findings from the evaluation show that our approach outperforms current state of the art. We also make available our statistically sound dataset, based on known benchmarks, to the wider research community. this proposed work, a secure and verifiable access control scheme based on the HASHED KEY TECHNIQUE cryptosystem for big data storage in clouds. We first propose a new HASHED KEY TECHNIQUE decryption algorithm to overcome the decryption failures of the original HASHED KEY TECHNIQUE, and then detail our scheme and analyze its correctness, security strengths, and computational efficiency. Our scheme allows the cloud server to efficiently update the ciphertext when a new access policy is specified by the data owner, who is also able to validate the update to counter against cheating behaviors of the cloud. It also enables (i) the data owner and eligible users to effectively verify the legitimacy of a user for accessing the data, and (ii) a user to validate the information provided by other users for correct plaintext recovery. Rigorous analysis indicates that our scheme can prevent eligible users from cheating and resist various attacks such as the collusion attack.In such a system,combined with verifiable computation and encrypt-then-mac mechanism, the data confidentiality, the fine-grained access control and the correctness of the delegated computing results are well guaranteed at the same time. Besides, our scheme achieves security against

chosen-plaintext attacks under the *k*-multilinear Decisional hashed technique. Moreover, an extensive simulation campaign confirms the feasibility and efficiency of the proposed solution.

**INDEXTERMS:HASHKEY, CIPHERTEXT**

# 1.INTRODUCTION

Most security and privacy mechanisms rely on (strong) encryption algorithms to protect the communications. However,flawed implementations often use cleartext communication to transfer sensitive information. The latter is often in IoT devices due to their lack of processing resources. The problem is amplified by the fact that IoT firmware is closed source and cannot be easily extracted since access to the storage and processing units can be concealed or access-protected. The use of compression instead of encryption from some devices may perplex the problem even more as there are no efficient and accurate methods to distinguish high entropy sources, e.g. encryption from compression. Therefore, a security investigator cannot easily determine whether a device is using a custom compression algorithm or encryption when evaluating its security.

Although existing data auditing schemes already have various properties (see Section 2), potential risks and inefficiency such as security risks in unauthorized auditing requests and inefficiency in processing small updates still exist. In this paper, we will focus on better support for small dynamic updates, which benefits the scalability and efficiency of a cloud storage server. To achieve this, our scheme utilizes a flexible data segmentation strategy and a ranked Merkle hash tree (RMHT). Meanwhile, we will address a potential security problem in supporting public verifiability to make the scheme more secure and robust, which is achieved by adding an additional authorization process among the three participating parties of client, CSS and a third-party auditor (TPA). Research contributions of this paper can be summarized as follows:

1. For the first time, we formally analyze different types of fine-grained dynamic data update requests on variable-sized file blocks in a single dataset. To the best of our knowledge, we are the first to propose a public auditing scheme based on BLS signature and Merkle hash tree (MHT) that can support fine-grained update requests. Compared to existing

schemes, our scheme supports updates with a size that is not restricted by the size of file blocks, thereby offers extra flexibility and scalability comparedto existing schemes.

2. For better security, our scheme incorporates an additional authorization process with the aim of eliminating threats of unauthorized audit challenges from malicious or pretended third-party auditors, which we term as ‘authorized auditing’.

3. We investigate how to improve the efficiency in verifying frequent small updates which exist in many popular cloud and big data contexts such as social media. Accordingly, we propose a further enhancement for our scheme to make it more suitable for this situation than existing schemes. Compared to existing schemes, both theoretical analysis and experimental results demonstrate that our modified scheme can significantly lower communication overheads.

Most PDP and POR schemes can support public data verification. In such schemes, there are three participating parties: client, CSS and TPA. In brief, both CSS and TPA are only semi-trusted to the client. In the old model, the challenge message is very simple so that everyone can send a challenge to CSS for the proof of a certain set of file blocks, which can enable malicious exploits in practice. First, a malicious party can launch distributed denial-of-service (DDOS) attacks by sending multiple challenges from multiple clients at a time to cause additional overhead on CSS and congestion to its network connections, thereby causing degeneration of service qualities. Second, an adversary may get privacy-sensitive information from the integrity proofs returned by CSS. By challenging the CSS multiple times, an adversary can either get considerable information about user data (due to the fact that returned integrity proofs are computed with client-selected data blocks), or gather statistical information about cloud service status. To this end, traditional PDP models cannot quite meet the security requirements of 'auditing-as-a-service', even though they support public verifiability.

## 2.LITERATURE SURVEY

### FINE-GRAINED CONTROL OF SECURITY CAPABILITIES:

We present a new approach for fine-grained control over users' security privileges (fast revocation of credentials) centered on the concept of an on-line semi-trusted mediator (SEM). The use of a SEM in conjunction with a simple threshold variant of the RSA cryptosystem (mediated RSA) offers a number of practical advantages over current revocation techniques. The benefits include simplified validation of digital signatures, efficient certificate revocation for legacy systems and fast revocation of signature and decryption capabilities. This paper discusses both the architecture and the implementation of our approach as well as its performance and compatibility with the existing infrastructure. Experimental results demonstrate its practical aspects. We begin this paper with an example to illustrate the premise for this work. Consider an organization – industrial, government, or military – where all employees (referred to as users) have certain authorizations. We assume that a Public Key Infrastructure (PKI) is available and all users have digital signature, as well as en/de-cryption, capabilities. In the course of performing routine everyday tasks, users take advantage of secure applications, such as email, file transfer, remote log-in and web browsing.

### OBLIVIOUS TRANSFER WITH ACCESS CONTROL :

We present a protocol for anonymous access to a database where the different records have different access control permissions. These permissions could be attributes, roles, or rights that the user needs to have in order to access the record. Our protocol offers maximal security guarantees for both the database and the user, namely (1) only

authorized users can access the record; (2) the database provider does not learn which record the user accesses; and (3) the database provider does not learn which attributes or roles the user has when she accesses the database. We prove our protocol secure in the standard model (i.e., without random oracles) under the bilinear Diffie-Hellman exponent and the strong Diffie-Hellman assumptions. Also to protect sensitive information such as medical or financial data we need to provide strong access control to be sure that only those people who have the necessary permissions can access it. But statistics about what sort of data people query also reveals a lot of information about them. It is possible to build a complete picture of someone's movements, transactions, locations and relationships from the trail left from interaction with websites and various databases. So personal security has become a serious issue.

## SIGNATURE SCHEMES AND ANONYMOUS CREDENTIALS FROM BILINEAR MAPS

We propose a new and efficient signature scheme that is provably secure in the plain model. The security of our scheme is based on a discrete-logarithm-based assumption put forth by Lysyanskaya, Rivest, Sahai, and Wolf (LRSW) who also showed that it holds for generic groups and is independent of the decisional Diffie-Hellman assumption. We prove security of our scheme under the LRSW assumption for groups with bilinear maps. We then show how our scheme can be used to construct efficient anonymous credential systems as well as group signature and identity escrow schemes. To this end, we provide efficient protocols that allow one to prove in zero-knowledge the knowledge of a signature on a committed (or encrypted) message and to obtain a signature on a committed message. Signature schemes exists if and only if one-way functions exist [NY89, Rom90]. However, the effi- ciency of these general constructions, and also the fact that these signature schemes require the signer's secret key to change between invocations of the signing algorithm, makes these solutions undesirable in practice.

## A VERIFIABLE RANDOM FUNCTION WITH SHORT PROOFS AND KEYS

We give a simple and efficient construction of a verifiable random function (VRF) on groups equipped with a bilinear mapping. Our construction is direct; it bypasses an expensive Goldreich-Levin transformation from a unique signature to a VRF in contrast to prior works of Micali-Rabin-Vadhan [MRV99] and Lysyanskaya [Lys02]. Our proofs of security are based on a decisional bilinear Diffie-Hellman inversion assumption (DBDHI), previously used in [BB04a] to construct an identity based encryption scheme. Our VRF's proofs and keys have constant size in contrast to proofs and keys of VRFs in [Lys02] and [Dod03], which are

linear in the size of the message. We operate over an elliptic group, which is significantly shorter than the multiplicative group Z ∗ n used in [MRV99], yet we achieve the same security. Furthermore, our scheme can be made distributed and proactive.

## ATTRIBUTE-BASED ENCRYPTION FOR FINE-GRAINED ACCESS CONTROL OF ENCRYPTED DATA

As more sensitive data is shared and stored by third-party sites on the Internet, there will be a need to encrypt data stored at these sites. One drawback of encrypting data, is that it can be selectively shared only at a coarse-grained level (i.e., giving another party your private key). We develop a new cryptosystem for fine-grained sharing of encrypted data that we call Key-Policy Attribute-Based Encryption (KP-ABE). In our cryptosystem, cipher texts are labeled with sets of attributes and private keys are associated with access structures that control which ciphertexts a user is able to decrypt. We demonstrate the applicability of our construction to sharing of audit-log information and broadcast encryption. Our construction supports delegation of private keys which subsumes Hierarchical Identity-Based Encryption (HIBE).

## 3. PROPOSED SYSTEM

In the challenge/verification process of our scheme, we try to secure the scheme against a malicious CSS who tries tocheat the verifier TPA about the integrity status of the client's data, which is the same as previous work on both PDP and POR. In this step, aside from the new authorization process, the only difference compared to is the RMHT and variable-sectored blocks. Therefore, the security of this phase can be proven through a process highly similar with, using the same framework, adversarial model and interactive games defined in. A detailed security proof for this phase is therefore omitted here.

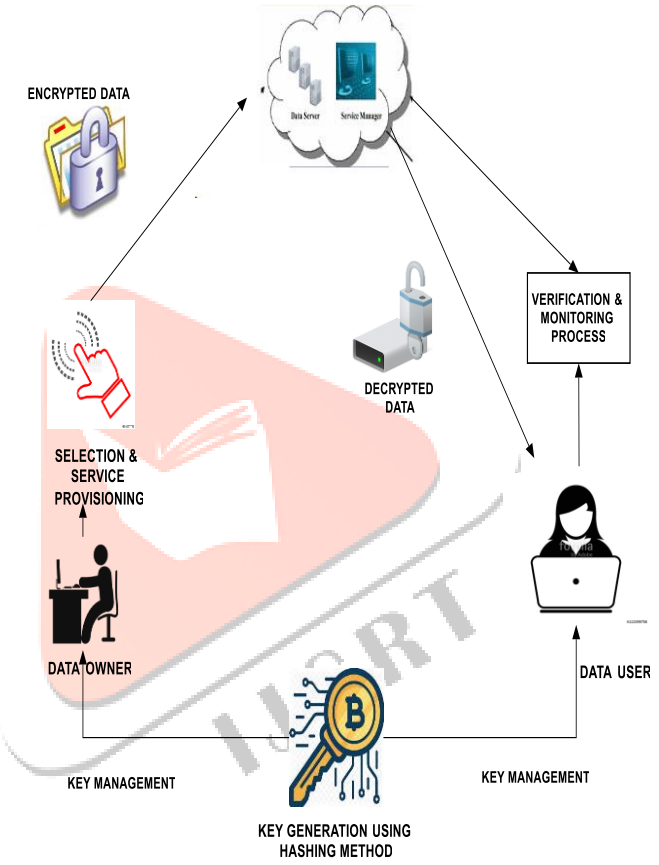


**Fig.1.shown the Work Flow of Model**

## 1.USER DATA METRIC

A data owner designates the access policy for its data, encrypts the data based on the access policy before outsourcing the data to the cloud server, and requests the cloud server to update the encrypted data when a new access policy is adopted. It can also check whether the cipher text at the cloud server is correctly updated.

## 2. SECURE SERVICE PROVISIONING

Services for data security and access control when users outsource sensitive data for sharing on cloud servers. This challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and on the other hand allowing the data owner to delegate most of the computation tasks involved in fine grained data access control to u Hashed Key Techniquested cloud servers without disclosing the underlying data contents. The data owner to delegate tasks of data file re-encryption and user secret key update to cloud servers without disclosing data contents or user access privilege information. This goal can be achieved by exploiting and uniquely combining techniques of attribute-based encryption (ABE), proxy re-encryption, and lazy re-encryption

## 3.KEY GENERATION

Key mechanisms to protect data confidentiality in traditional data networks. It is designed to block unauthorized users and malicious hackers from accessing data. Although the objective of access control in cloud storage does not differ from that in traditional data network, the requirement does change. Traditional access control enforced by the service provider could not stop a curious cloud service provider accessing users' sensitive data, which was stored in the service provider's infrastructure and managed by the service provider. A curious cloud storage server trying to derive sensitive information from its stored data or from data operations performed by data owner and authorized users, is a new threat model against data confidentiality in cloud storage service. Moreover, a malicious service provider could intentionally leak the data to unauthorized parties for profit, or a malicious attacker could compromise the service provider and get unauthorized access to the data.

## 4.VERIFICATION AND HASHING PROCESS

The HASHED KEY TECHNIQUE cryptosystem is based on the shortest vector problem(SVP) in a lattice that makes it lightning fast and resistant toquantum computing attacks.The outsourced data must be protected from eavesdropping attacks during upload, download, update, and retrieval of the data. The legitimacy of any user to access the data must be verifiable; the information provided by any user for plaintext recovery must be validated. The proposed scheme must be able to counter potential attacks (cheating, collusion attacks, forging attacks, etc.) launched by misbehaving users and adversaries.

## 5. REPORTS

Reporting and analysis of Web use is a critical element of legal compliance and maintaining a safe working environment for organizations. Content Keeper ARM provides a powerful, yet easy to use and highly flexible reporting platform that enables you to gain meaningful insight into your Web use and make informed decisions. It enables you to consolidate reporting data across multiple offices, devices or domains. Reporting can be provided to appropriate personnel for their

particular areas of concern. Each table we have report format. It should be viewed by the crystal report viewer button that should be retrieve by that object name we declared for that.

## 4. RESULTS AND DISCUSSION

In this section the result and discussion covers thoroughly analysing the results, we discovered that the randomness tests applied to compressed files exhibited different behaviours according to the input file-type. More precisely, there is a strong relationship between the randomness of the input file-type and the randomness of the compressed file generated. To document such behaviour, we classify the outcomes of all data streams according to input file-types for all the compressed (see Fig 2) algorithms tested

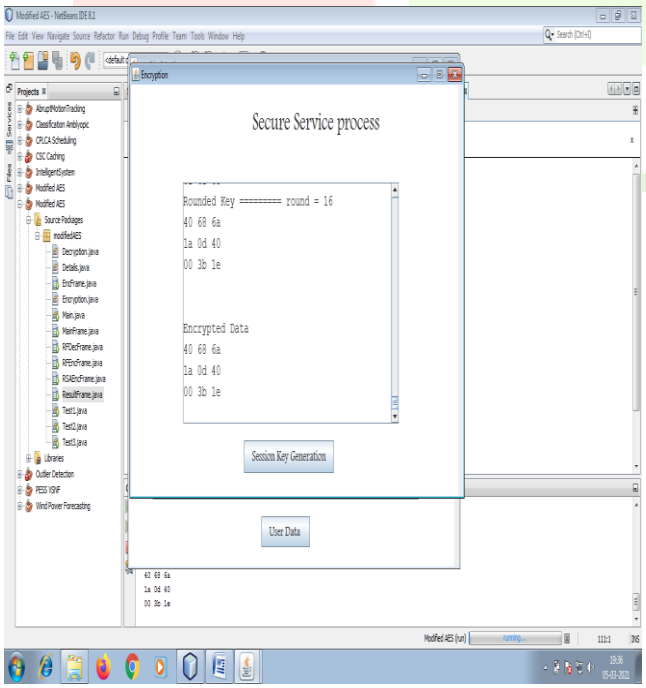


Figure 2. shown the performance of the system

Next, after thoroughly analysing the results, we discovered that the randomness tests applied to compressed files exhibited different behaviours according to the input file-type. More precisely, there is a strong relationship between the randomness of the input file-type and the randomness of the compressed file generated. To document such behaviour, we classify the outcomes of all data streams according to input file-types for all the compressed (see Fig 5) algorithms tested. First, Fig. 2. shows the accuracy detection of 64KB encrypted data streams, whose behaviour is similar to other file sizes and therefore we omit the results. It is apparent that when we analyse encrypted data streams we cannot distinguish the input file-type which is a result that complies with the theory. On the contrary, one can observe that compressed binary and compressed text files, as well as compressed video files are more easily classified by our method (achieving 100% accuracy with binary and text in 64KB file streams) than the rest.

This implies that their randomness is much lower than other files, such as PDF or MP3 (i.e. MP3 is already a compressed file, as well as JPG files). This behaviour applies to all file sizes except for files lower than 4KB, where compressed text files become more indistinguishable, but we still find notable differences between each file-type. Such findings are relevant, especially from a security perspective, because discovering the content of compressed data streams enables the proper management when detecting exchange of unexpected/suspicious files (e.g. executables). Again, the result complies with the theory since the output of compression algorithms like Huffman or members of the Lempel-Ziv family highly depends on the statistical properties of the

input stream. In summary, the results show that we can distinguish compressed from encrypted bit streams accurately in an efficient way, and in the former case we may even determine the content of compressed files with certainty.

The accuracy of the proposed methodology is highly dependent on the size of the investigated packets, decreasing as the packet size decreases. Moreover, our threshold-based method achieves higher accuracy (see Figure 1) with more efficiency than the state-of-the-art [4] (k-NN and convolutional Neural networks), since they have a complexity of at least O(n2) whilst our method has linear cost O(n). Moreover, in most cases we need only to compare one feature (and not n) to classify a bit stream (e.g. if the chi-square test fails, the file is discarded and there is no need to compute the rest of tests). In addition, results are reproducible since the variability of threshold values is almost negligible, so that a universal threshold can be considered regardless of data, avoiding costly training procedures (such in the case of neural networks) and other data-dependent methods.

## CONCLUSION

This proposed model, an improved PROPOSED TECHNIQUE cryptosystem to overcome the decryption failures of the original PROPOSED TECHNIQUE and then present a secure and verifiable access control scheme based on the improved PROPOSED TECHNIQUE to protect the outsourced big data stored in a cloud. This model approach was evaluated using a statistically sound benchmark was developed, achieving an accuracy between 68.68% (worst case scenario) and 94.72%. The accuracy rate depends on the packet size. It also determined that the randomness of compressed files has a strong dependence on the input file-type and we analysed their behaviour. We were able to conclude that (compressed) binary, text and video files can be easily detected by our system. Moreover, our method is more efficient than other competing state-of-the-art works and enables real-time traffic classification.

## FUTURE WORKS

In the future, this model will be refine this method to further increase its accuracy; especially for low packet sizes, and enable more accurate content classification of compressed files, to improve pro-active security and specific file-type detection in public clouds.

## REFERENCES

[1] M. A. Beyer and D. Laney, “The importance of big data: a definition,” Stamford, CT: Gartner, 2012.

[2] V. Marx, “Biology: The big challenges of big data,” Nature, vol. 498, no. 7453, pp. 255–260, 2013.

[3] G. P. Consortium et al., “A map of human genome variation from population-scale sequencing,” Nature, vol. 467, no. 7319, pp. 1061–1073, 2010.

[4] A. Sahai and B. Waters, “Fuzzy identity-based encryption,” Advances inCryptology–EUROCRYPT 2005, pp. 457–473, 2005.

[5] C. Hu, F. Zhang, X. Cheng, X. Liao, and D. Chen, “Securing communications between external users and wireless body area networks,” in Proceedings of the 2nd ACM workshop on Hot topics on wireless network security and privacy. ACM, 2013, pp. 31–36.

[6] C. Hu, H. Li, Y. Huo, T. Xiang, and X. Liao, “Secure and efficient data communication protocol for wireless body area networks,” IEEE Transactions on Multi-Scale Computing Systems, vol. 2, no. 2, pp. 94–107, 2016.

[7] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-based encryption for fine-grained access

control of encrypted data," in Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006, pp. 89–98.

[8] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization," Public Key Cryptography– PKC 2011, pp. 53–70, 2011.

[9] C. Hu, N. Zhang, H. Li, X. Cheng, and X. Liao, "Body area network security: a fuzzy attribute-based signcryption scheme," IEEE journal on selected areas in communications, vol. 31, no. 9, pp. 37–46, 2013.

[10] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," Advances in Cryptology–EUROCRYPT 2011, pp. 568–588, 2011.

[11] C. Hu, X. Cheng, Z. Tian, J. Yu, K. Akkaya, and L. Sun, "An attributebasedsigncryption scheme to secure attribute-defined multicast communications," in SecureComm 2015. Springer, 2015, pp. 418–435.

[12] A. Shamir, "Identity-based cryptosystems and signature schemes," in Advances in cryptology. Springer, 1985, pp. 47–53.

[13] M. Dehkordi and S. Mashhadi, "An efficient threshold verifiable multisecret sharing," Computer Standards & Interfaces, vol. 30, no. 3, pp.187–190, 2008.

[14] Z. Eslami and J. Z. Ahmadabadi, "A verifiable multi-secret sharing scheme based on cellular automata," Information Sciences, vol. 180, no. 15, pp. 2889–2894, 2010.

[15] M. H. Dehkordi and S. Mashhadi, "New efficient and practical verifiable multi-secret sharing schemes," Information Sciences, vol. 178, no. 9, pp. 2262–2274, 2008.