Secure Encrypted Data with Authorized Deduplication

Dr. A. Noble Marry Juliet,Dr. N. Senthil Madasamy,N. Hari Priya and Dr. N. Suba Rani
Department of Computer Science and Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, India
Email: haripriyanarayanan2@gmail.com

***Abstract*—Data deduplication is a vital technique to compress a data for eradicating duplicate data. It is utilized broadly to handle the storage space effectively and helps in saving bandwidth and to safeguard the truthfulness of delicate data. Traditional encryption will end in multiple different ciphertexts produced from the same plaintext by different uses secret keys, which hiders data deduplication. In Secure hash algorithm which generates a hash code for each document and by comparing the hash code with the previously stored documents, once the code matched, instead of storing the document in the memory space the file is mapped to a specific field using mapping technique. This method gives more secure way to store data effectively.**

***Keywords*— Content based page sharing, kernel same page Merging, Utility based cache partitioning, Virtual Machine monitor, Secure hash algorithm.**

I. INTRODUCTION

Information security can be summed up to info, a group of steps, procedures, and strategies that are used to stop and observe illegal access, trouble-shooting, revelation, perturbation and adjustment of computer network sources. Enhancing the privacy, eligibility and reliability of the work requires a lot work to strengthen the current methods from constant trials to break them and to improve new ways that are resistant to most kinds of attacks. Accordingly, it was proven that encoding is one of the most reliable strategies used to secure information since the ancient days of the Romans who used similar methods to enable security on their valued information and documents. The rapid development of technology changes user's method and efficiency in processing data, where each system must provide the scalable computing and efficient storage to users. Along with the explosive growth of data, massive duplicate data occupied the storage space and the huge expenditure bring a severe challenge to the limited storage space. Therefore, how to reduce the management expenditure and improve the storage efficiency is an urgent issue to be solved. Recent research shows that data deduplication can reduce up to 83% for backup systems and 68% for memory systems. While storing a large amount of duplicate data, it may affect performance, bandwidth, storage inconsistencies, etc., for example by taking email storage system, many copies of the same messages and file attachments have been shared by many people causing data deduplication.

II.RELATED STUDY

In [1] Limited main memory size is considered as one of the major bottlenecks in virtualization environments. Content-Based Page Sharing is also on of the efficient memory deduplication technique to reduce server memory

Grenze ID: 02. ICIEICE.2021.1.6

requirements, where pages with same content are detected and shared into a single copy. As the widely used implementation of CBPS, Kernel Same page Merging (KSM) maintains the whole memory pages into two global comparison trees (a stable tree and an unstable tree). To detect page sharing opportunities, each candidate page needs to be compared with pages already in these two large global trees.

In [4] the problem of partitioning a shared cache between multiple concurrently executing applications. The commonly used policy implicitly partitions a shared cache on a demand basis, giving more cache resources to the application that has a high demand and fewer cache resources to the application that has a low demand. However, a higher demand for cache resources does not always correlate with a higher performance from additional cache resources. Utility- based cache partitioning (UCP), a low-overhead, runtime mechanism that partitions a shared cache between multiple applications depending on the reduction in cache misses that each application is likely to obtain for a given amount of cache resources.

Modern processors contain multiple cores which enables them to concurrently execute multiple applications (or threads) on a single chip[5]. As the number of cores on a chip increases, the pressure on the memory system to sustain the memory requirements of all the concurrently executing applications (or threads) increases. One of the keys to obtaining high performance from multicore architectures is to manage the largest level on- chip cache efficiently so that off-chip accesses are reduced. This paper investigates the problem of partitioning the shared largest-level on-chip cache among multiple competing applications. Traditional design for on-chip cache uses the (policy for replacement decisions.

In [11] Multicore processors contain new hardware characteristics that are different from previous generation single-core systems or traditional SMP (symmetric multiprocessing) multiprocessor systems. These new characteristics provide new performance opportunities and challenges. In this paper, show how hardware performance monitors can be used to provide a fine-grained, closely-coupled feedback loop to dynamic optimizations done by a multicore- aware operating system. Demonstrate three case studies on how a multicore-aware operating system can use these online capabilities for determining cache partition sizes, which helps reduce contention in the shared cache among applications, detecting memory regions with bad cache usage, which helps in isolating these regions to reduce cache pollution, and detecting sharing among threads, which helps in clustering threads to improve locality.

In Many modern applications result in a significant operating system (OS) component. The OS component has several implications including affecting the control flow transfer in the execution environment. This paper focuses on understanding the operating system effects on control flow transfer and prediction, and designing architectural support to alleviate the bottlenecks. characterize the control flow transfer of several emerging applications on a commercial operating system. propose two simple OS- aware control flow prediction techniques to alleviate the destructive impact of user/OS branch interference.

III. PROPOSEDWORK

A. Padding Process

In the proposed work, improvement of framework in terms of data deduplication and Mapping Technique is proposed. The main goal is to secure data deduplication system and there by verifying the authorization. Unauthorized users cannot unscramble the encrypted text even conspire with service provider. We firstly introduce the coordinate memory deduplication and partition. Here secure hash algorithm is proposed. SHA-1 produces a message digest based on principles MD4 and MD5. SHA-1 differs from SHA -0 only by a single bitwise rotation in the message schedule of its compression function. SHA – 1 produces a 160 bits hash value known as message digest. This hash value is rendered as hexadecimal number. It is 40 digits long. where initially the input given(file) is divided into 448 bits and it is carried to the second stage padding process. Padding Process In this process 64 bit is added to the 448 and by doing this the size of the input file is determined. The output from the second stage is passed into the compression function.

B. Compression function

This stage consists of round 0 to round 79 so totally 80 rounds. The output from the second stage (448+64) a total of 512 bits is taken as the input and divides it int 32 bits, since it runs for five times and as a result 160 bits hash code is generated, this 160 bits will occur only after the processing of all 80 round in the compression function.

C . Applications of Hash Functions

- Message authentication which is used to check if a message has been modified.
- Digital signatures encrypt digest with privatekey.
- Password storage where digest of password is compared with the storage so that hackers cannot get password from storage.
- Key generation can be generated from digest of pass- phrase and it can be made computationally expensive to prevent brute-force attacks.
- Pseudorandom number generation is used for iterated hashing of a seed value.
- Intrusion detection and virus detection which keep and check hash of files on system.

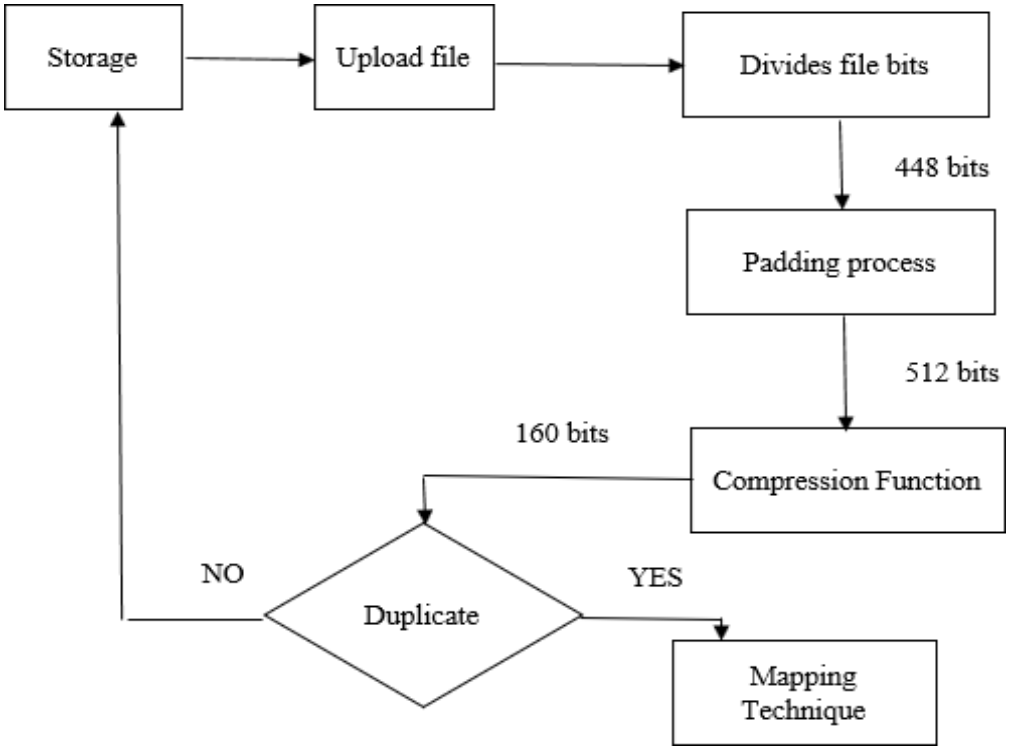

Fig.1 Flow diagram of Proposed System

D. Deduplication

Deduplication, is a technique to save storage cost by enabling us to store only one copy of identical data, which becomes unprecedentedly significant with the dramatic increase in data stored for the purpose of ensuring data confidentiality, they are usually encrypted before outsourced

E. Record Matching

Record matching is the main functionality in this application function. The data matching initially performs the pre- processing with data upload to the client. The has been uploaded by the SHA1 hashing technique. Every data has generated by the hash key value. It checks both the file and attitude in the server using the genetic fitness by all the has been encrypted using AES algorithm. If record attribute is matching with the records attribute is matching with existing document. Based on that duplication the file matching is performed.

F. The Mapping technique

The Mapping Technique is used to deduplication as well as performed in single copy of same data for multiple data owners in storage. If any of the data is stored in same data means the data cannot be stored it will be mapped

and linked to the document/data. And a virtual machine based memory partition called VMMP is added into our technique is to reduce interference among virtual machines.

IV. Implementation And Evaluation

The connection between the multiple storage system is connected with the host. Then upload a specific file or folder from your PC to the data location.

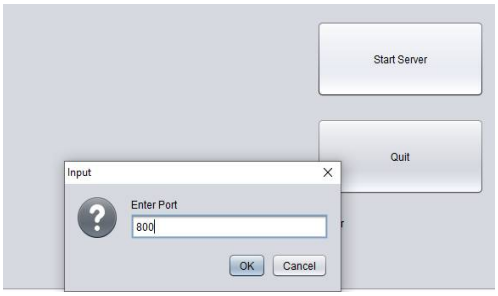

Fig 2 Setting Storage

If the file or document is stored for the first time it shows uploaded successfully else it maps to the current location without again saving the same file or folder.

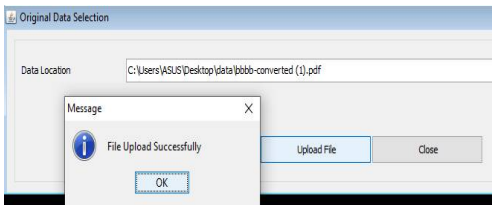

Fig 3 Upload File

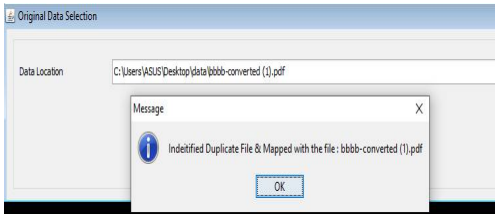

Fig 4 Mapping file

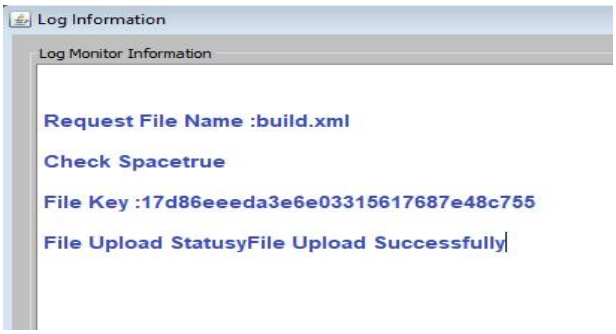

Fig 5 Information of the uploaded file

A. Result Analysis

The performance of the secure encrypted algorithm is measured by using the formula O (c+ x n…). The file size is taken as (n) and the run time is calculated by means of getting the output from the initial padding process(c) and output from Compression function (x)

TABLE I. FORMULA DESCRIPTION

Notation	Semantic
O	Limiting behavior of a function
c	Functions with initialization and finalization are considered as single constant
x	Compression function in a single input
n	length of the input

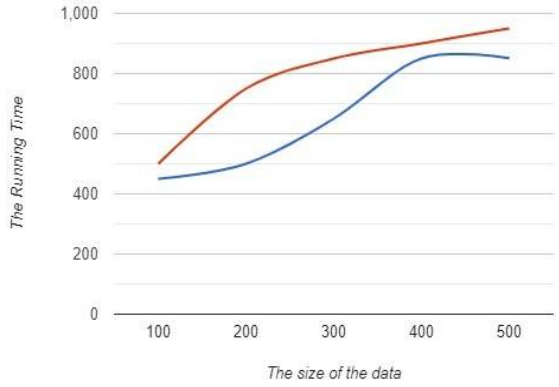

Fig 6 Speed Analysis

V. CONCLUSION

The various deduplication techniques are surveyed. Among them, it has been concluded that Secure hash algorithm in data deduplication is well and good when compared to other strategies by comparing the hash of each and every chunk. Hence, this technique improves storage efficiency and thereby improve the performance by enabling storage resources to transfer and handle more data. In future, more research works could be focused on variable size chunking method toreduce processing time, and optimize of largescale data storage. And also to develop an efficient method to reduce fragmentation and obtain high write and read throughput.

ACKNOWLEDGEMENT

This work is performed at Dr. Mahalingam College of Engineering and Technology as a part project work titled "Secure Encrypted data with authorized deduplication" supported by Department of Computer Science and Engineering.

REFERENCES

[1] L Chen, Z. Wei, Z. Cui, M. Chen, H. Pan,(2019) ' CMD: Classification-based Memory Deduplication through Page Access' Proceedings of IEEE, vol 413, no 43 ,pp.709-821.
[2] H. Xiong, H. Zhang, and J.M. Sun,(2018), 'Attribute-based privacy-preserving data sharing for dynamic groups in cloud computing,' IEEE Systems Journal, no. 99, pp. 1–22.
[3] D. Wu, H. Shi, H. Wang, R. Wang, H. Fang., (2017),'A feature-based learning system for internet of things applications,' IEEE Internet of Things Journal, vol 85, pp. 445-512
[4] J. Xiong, Y. Zhang, X. Li et al.,(2018), 'Rse-pow: a role symmetric encryption scheme with authorized deduplication for multimedia data' Mobile Networks Applications, vol. 23, no. 3, pp. 650–663.
[5] W. Xia, H. Jiang, D. Feng, F. Douglis, P. Shilane, Y. Hua, M. Fu, Y. Zhang, Y.Zhou, (2016), 'A comprehensive study of the past, present, and future of Deduplication', Proceedings of the IEEE, vol. 104, no. 9, pp. 1681–1710.
[6] J. Li, Qin, P. P. Lee, and X. Zhang, (2017),'Information leakage in encrypted deduplication via frequency analysis,' in the 47

Annual IEEE International Conference on Dependable Systems and Networks.
[7] Chen, S. S. Chow, Q. Huang, D. S. Wong, and Z. Liu,(2018),'Multiauthority fine-grained access control accountability and application cloud,'Journal of Network and Computer Applications, vol. 112, pp. 89–96.
[8] D.Wu, Q. Liu, H.Wang, Q.Yang, and , (2018) 'Cache less for more Exploiting cooperative video and delivery in d2d communications,' IEEE Transactions Multimedia, vol 65, pp. 71-89
[9] H. Xiong,Q. Mei, and Y. Zhao,(2019) 'Efficient and provably secure certificate parallel key-insulated signature without pairing for iiot environments,' IEEE Systems Journal, vol. 412, pp. 554-875
[10] R. H. Deng, K. K. R. Choo, and J. Weng,(2016) 'An efficient privacy preserving outsource calculation toolkit with multiple key,' IEEE Transactions Information Forensics and Security, vol. 11, pp. 2401–2414.
[11] H. Cheng, C. Lin, J. Li, and C. Yang,(2014)' Memory Latency Reduction via Thread Throttling' ,Journal of Network and Multimedia vol. 179, pp. 46–61. 42
[12] J. Liu, N. Asokan, and B. Pinkas,'Secure deduplication of encrypted without additional Independent servers,'(2015) in Proceedings of the 22nd ACM Conference on Computer and Communications Security. ACM, pp. 874–885.
[13] D. Meyer andW. J. Bolosky,(2012) 'A study of practical deduplication,' ACM Transactions on Storage (TOS), vol. 7, no. 4, pp. 1–20.
[14] Lin.Y. K. Li, X. Chen, P. Lee, and W. Lou, (2015).'A hybrid cloud approach for Secure authorized deduplication,'
[15] IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1206–1216
[16] J. Ni, K. Zhang, Y. Yu et al,J Chen.,(2018) 'Providing task allocation and secure deduplication for mobile crowd sensing via fog computing,' IEEE Trans. on Dependable and Secure Computing, vol. PP, no. 99, pp. 1–1
[17] Y. Zhang, X. Chen, J. Li, D. S. Wong, H. Li, and I.You,(2017),'Ensuring attribute privacy protection and fast decryption for outsourced data security mobile cloud computing,' Information Sciences, vol. 379, pp. 42–61.
[18] J. Xiong, J. Ren, L. Chen et al.,(2019) 'Enhancing privacy and availability for data clustering intelligent electrical service of IOT,' IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1530–1540.
[19] K.-H. Yeh,(2018) 'A secure transaction scheme with certificateless cryptographic primitives for iot- based mobile payments, 'IEEE Systems Journal, vol. 12, 2, pp-. 2027–2038.
[20] C.-M. Chen, B. Xiang, Y. Liu, and K.-H. Wang,(2019), 'A secure authentication protocol for internet of vehicles,' IEEE Access, vol. 7, pp. 12 047–12 057.
[21] Y. Zhang, R. Deng, G. Han, D. Zheng,2018 'Secure smart health with privacy aggregate authentication and access control flow in internet of things,' Journal of Network and Computer Applications, vol. 123, no. 12, pp. 89–100.
[22] Z. Mo, Y. Zhou, and S. Chen,(2012)'A dynamic proof of retrievability scheme with o (log n) complexity,' in Communications IEEE International Conference,. pp. 912–916.
[23] L. Ramos, T.F. Wenisch, and R. Bianchini,(2011) MemScale: Active Low-Power Modes for Main Memory by Q.
[24] Deng, D. Meisner, Conference on communication Security pp.86-90
[25] C.-M. Chen, B. Xiang, Y. Liu, and K.-H. Wang, , (2019)'A secure authentication protocol for internet of vehicles,' IEEE Access, vol. 7, pp. 12 047–12 057.