Literature Study on Data Protection for Cloud Storage

Dr. S. SRIDHAR
Ex. President and Vice Chancellor,
Dr. K. N. Modi University, Rajasthan

SAHANA S
Final year BE (CSE), SJIT, Chennai.

Abstract: **Many data security and privacy incidents are observed in today Cloud services. On the one hand, Cloud service providers deal with a large number of external attacks. In 2018, a total of 1.5 million Sing Health patients' non-medical personal data were stolen from the health system in Singapore. On the other hand, Cloud service providers cannot be entirely trusted either. Personal data may be exploited in a malicious way such as in the Face book and Cambridge Analytical data scandal which affected 87 million users in 2018. Thus, it becomes increasingly important for end users to efficiently protect their data (texts, images, or videos) independently from Cloud service providers. In this paper, we aim at presenting a novel data protection scheme by combining fragmentation, encryption, and dispersion with high performance and enhanced level of protection as Literature study.**

Keywords: **Data Security; Cloud Service Provides; Fragmentation; Encryption; Dispersion;**

I. INTRODUCTION

The Cloud-based services for individual end users are gaining popularity especially for data storage. Relying on large storage space and reliable communication channel, Cloud-based service providers such as Dropbox, Google Drive, or Amazon Drive just to name a few, are providing individual users with almost infinite and low-cost storage space. This situation raises the question of the trustworthiness of Cloud service providers. Many data security and privacy incidents are observed in today Cloud services. On the one hand, Cloud service providers deal with a large number of external attacks. In 2018, a total of 1.5 million SingHealth patients non-medical personal data were stolen from the health system in Singapore. On the other hand, Cloud service providers cannot be entirely trusted either. Personal data may be exploited in a malicious way such as in the Face book and Cambridge Analytical data scandal which affected 87 million users in 2018. Thus, it becomes increasingly important for end users to efficiently protect their data (texts, images, or videos) independently from Cloud service providers. In this paper, we aim at presenting a novel data protection scheme by combining fragmentation, encryption, and dispersion with high performance and enhanced level of protection. Fragmentation methods are introduced for data storage in a cost-effective manner using a public Cloud for the less confidential data fragments. For this matter, we expect the method to split data into a first *private fragment* that is small in size but important in content, while the other *public fragments* fulfill most of the storage space and leak little information related to the original data. Then, the *private fragment* should be well protected and stored in a trusted area such as an end user's personal computer in the end user to Cloud scenario. The *public fragments* are stored in public Clouds and cannot be used to rebuild the original data.

II. LITERATURE STUDY

The security level is always based on the design purpose. For instance, some multimedia SE methods are designed to only reduce the visual effects which are normally seen as low level considering security. More specifically, if the protection is only done on the private fragments, we consider it as low security level as there are many related works to show the direct recovery from the public fragments. Thus, the only previous works qualified high security levels. In this paper, intensive security analysis is performed to prove a high security level is achieved with protecting both the private fragments and public fragments. Data integrity is an important criteria but is always ignored in previous SE methods. For instance, ,a fractional Wavelet-based SE method is used to degrade the image quality. But data integrity cannot be guaranteed as the rounding errors of calculations between integers and floating point numbers are ignored which will cause serious issues. For the SE methods based on compression and coding, the data integrity could be guaranteed. However, SE methods designed based on compression and coding techniques are always relying on the details of specific compression and coding algorithms which lead to error propagation and format reliance. For instance, a protection method for JPEG2000 images is presented consisting in permuting the MQ lookup table. This will lead to error propagation in the decoding process when there are tiny errors in the transmission and also make this method only available when MQ coding is used. Such issue is a voided in our method with processing data as matrices of bytes in an agnostic manner and designing the allocation of data fragments according to communication channel status. Storage optimization is considered in this special use case of secure storage from end users to Clouds. For most SE methods, the fragmentation concept is not designed based on the storage usage of public

Clouds which optimize the storage space usage of the trusted area. In this brief review, only the work could be used to optimize the trusted storage area by uploading the public fragments to the Clouds. In this paper, we defined the confidential levels of the fragments and the public fragments are also protected. Thus, the small private fragment with high confidential level can be stored in an area trusted by the end users while the public and protected fragments can be stored on public Clouds with resistance to attacks.

III. DESIGN ANDPROPOSED METHOD

In this section, we first introduce the general concept of selective encryption and then combine it with the idea of fragmentation and dispersion. Afterward, several key design and implementation elements will be given in order to illustrate with our practical protection method.

General concept

The initial idea of existing SE methods is to protect a piece of information by encrypting only a part of it. This could increase the overall performance for multimedia contents by reducing the data amount that has to been crypted. In our design, additional concepts are introduced: fragmentation and dispersion. As shown, input data d is separated into two fragments, $d_1$and d_2: $d_1$must take important elements of original information and little storage space. $d_1$will become the *private fragment* after encryption. $d_2$is the *public fragment*. It is intended to take up most of the storage space, while carrying little pertinent information from d. Then, $d_1$can be stored in local and private storage whichcouldhavelimitedstoragespaceand$d_2$canbestore d in a public Cloud with light weight protection.

Discrete wavelet transform

In previous works, Discrete Cosine Transform (DCT) was used to support fragmentation decision before performing SE for bitmap protection. As pointed out in, DCT cannot guarantee the loss due to conversions between integers and floating point numbers which will result in rounding errors. These rounding errors can be reduced with a precise design with more storage space, but Performance of DWT must be considered based on comparing the execution time against full encryption. In some scenarios, the preprocessing steps of SE can legitimately be ignored, as SE and compressions are integrated, such transformation is being used by both applications. However, our use case that encompasses any kind of data, will have to take into account the entire process when it comes to performance evaluation. Any kind of data can be seen as a sequence of data chunks D_i were each chunk D_i will be defined in turn as a square matrix with a tunable size (512 512 or 1024 1024bytes), depending on the accommodation of transformation or the hardware implementation. Every element of D_i is a byte which can be seen as an 8-bit integer. Then every chunk $D_i$will simply be processed using the SE method block by block with a block size of 8 8 as shown in Figure 3. The chunk size can be changed according to the implementation details, especially the GPGPU hardware configuration. The block size is supposed to be always 8 8, which fits bestour design as indicated in Section 3.4. This tiling step is usednot only for fitting with the GPGPU architecture but also forthe best design of the three fragments. For every 8 8 block, the first step is to perform the 2D, Discrete Wavelet Transform (DWT-2D). In our work, two successive levels of the DWT-2D were performed with the Le Gall 5/3 filter. The low frequency coefficients ($2^{nd}LL$ coefficients) are considered as the private fragment. This fragment takes only 4 out of 64 coefficients but carries most of the information according to an energy viewpoint. The AES-128 bit will be used to protect this fragment. In our design, the code is structured such that another cipher algorithm can easily replace AES-128 if needed. The other two coefficients levels are considered as the two" public and protected fragments" (PPFs). The private fragment of each 8 8 block will then be used to generate a 256-bit sequence, by using SHA 256. This will guarantee the generation of different bit sequences, even when the corresponding private fragments in the neighboring blocks are very similar (encryption key is also involved to guarantee the key sensitivity). This bit sequence is used to protect the 1^{st}PPFs (the remaining coefficients of 2^{nd}level DWT are shown) by performing an XOR operation. This fragment is defined as the 1^{st}PPF. For the 2^{nd}PPF which contains the remaining DWT coefficients, protection is done by XORingit with a bit sequence generated from SHA-512 based on the inputs of 1^{st}PPF and the encryption key. The protection of the PPFs is provided by XOR operations and is based on the randomness guaranteed by the SHA functions. For example, in some cases like bitmaps, as long as there are redundancies due to similar neighboring pixels, the frequency coefficients could be very similar, especially between neighbor blocks.

Numerical precision

The preprocessing step used to separate data may lead to integrity problems, where data before and after reversing protection could be different. In previous SE methods, this is normally due to rounding errors of conversions between integers and floating point numbers. One way to solve this problem is described, where the authors proposed to declare all variables with a *double* precision based on their bit-length of 64 bits. This leads to a large increase in the usage of storage space without being able to totally avoid such rounding errors. However, it is not optimal in terms of footprints to use larger storage space in order to float precision numbers, especially if the input data is stored as an integer with a bit-length of 8 bits, where the resulting storage space will require 8 times the storage space

of the original data. An optimized value representation in terms of the footprint was designed, yet it was not possible to totally avoid rounding errors caused by the DCT. In this paper, the preprocessing step is the DWT based on "LeGall 5/3" filter which is designed to be an integer-to- integer map, such that this DWT is lossless. As a result, on one hand, any rounding error is avoided; on the other hand, the extra storage space usage caused by the *int* to *float* conversion does not exist either. The only possible extra storage usage can be caused by the different range value of the input 8-bit *int*, and the output *int* coefficients. The output value range can be calculated as long as the input values are always stored Byte by Byte. The input value range (seen as unsigned value) will be then from 0 to 255 which can be considered as from128 to +127 (the range is seen as from 128 to +128 during the following calculation). Then the storage methods can be designed according to the value range distribution.

IV. CONCLUSIONS

Thus, it becomes increasingly important for end users to efficiently protect their data (texts, images, or videos) independently from Cloud service providers. In this paper, we aim at presenting a novel data protection scheme by combining fragmentation, encryption, and dispersion with high performance and enhanced level of protection as Literature study.

V. REFERENCES

[1] F. Hu, M. Qiu, J. Li, T. Grant, D. Taylor, S. McCaleb, L. Butler,and R. Hamner, "A review on cloud computing: Design challenges in architecture and security," Journal of computing and information technology, vol. 19, no. 1, pp. 25–55,2011.

[2] H. Li, K. Ota, and M. Dong, "Virtual network recognition and optimization in SDN-enabled cloud environment," IEEE Transactions on Cloud Computing, 2018.

[3] Y. Li, W. Dai, Z. Ming, and M. Qiu, "Privacy protection for preventing data over-collection in smart city," IEEE Transactions on Computers, vol. 65, no. 5, pp. 1339–1350, 2016.

[4] L. Kuang, L. Yang, J. Feng, and M. Dong, "Secure tensor decomposition using fully homo morphic encryption scheme," IEEE Transactions on Cloud Computing, 2015.

[5] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "Big data analysis based secure cluster management for optimized control plane in software-defined networks," IEEE Transactions on Network and Service Management, vol. 15, no. 1, pp. 27–38,2018.

[6] K. Gai, K.K. R. Choo, M. Qiu, and L. Zhu, "Privacy-preserving content-oriented wireless communication in Internet-of-Things," IEEE Internet of Things Journal, vol. 5, no. 4, pp. 3059–3067,2018.

[7] S. Hambleton et al., "A glimpse of 21st century care," Australian Journal of General Practice, vol. 47, no. 10, pp. 670–673, 2018.

[8] A.Massoudi, F.Lefebvre, C.DeVleeschouwer, B.Macq, and J.J. Quisquater, "Overview on selective encryption of image and video: challenges and perspectives," EURASIP Journal on Information Security, vol. 2008, no. 1, p. 1, 2008.

[9] T. Xiang, J. Hu, and J. Sun, "Outsourcing chaotic selective image encryption to the cloud with steganography," Digital Signal Processing, vol. 43, pp. 28–37, 2015.

[10] H.Qiu, G.Memmi, X.Chen, and J.Xiong, "DC coefficient recovery for JPEG images in ubiquitous communication systems," Future Generation Computer Systems, 2019.

[11] G. O. Karame, C. Soriente, K. Lichota, and S. Capkun, "Securing cloud data under key exposure," IEEE Transactions on Cloud Computing, 2017.

[12] H. Qiu and G. Memmi, "Fast selective encryption methods for bitmap images," International Journal of Multimedia Data Engineering and Management (IJMDEM), vol. 6, no. 3, pp. 51–69, 2015.

[13] H.Qiu, "Phd thesis: An efficient data protection architecture based on fragmentation and encryption," arXiv preprintar Xiv: 1803.04880, 2018.

[14] H.Li, K.Ota, M.Dong, A.Vasilakos, and K.Nagano, "Multimedia processing pricing strategy in GPU-accelerated cloud computing," IEEE Transactions on Cloud Computing, 2017.

[15] L. Krikor, S. Baba, T. Arif, and Z. Shaaban, "Image encryption using DCT and stream cipher," European Journal of Scientific Research, vol. 32, no. 1, pp. 47–57,2009.

[16] A. Pommer and A. Uhl, "Selective encryption of wavelet-packet encoded image data: efficiency and security," Multimedia Systems, vol. 9, no. 3, pp. 279–287, 2003.

[17] N. Taneja, B. Raman, and I. Gupta, "Selective image encryption in fractional wavelet domain," AEU-International Journal of Electronics and Communications, vol. 65, no. 4, pp. 338–344, 2011.

[18] A. Belazi, A. A. A. El-Latif, R. Rhouma, and S. Belghith, "Selective image encryption scheme based on DWT, AES S-box and

chaotic permutation," in Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International. IEEE, 2015, pp. 606–610.

[19] H. Qiu, N. Enfrin, and G. Memmi, "A case study for practical issues of DCT based bitmap selective encryption methods," in 2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC). IEEE, 2018, pp.1–7.

[20] T. Xiang, C. Yu, and F. Chen, "Secure MQ coder: An efficient way to protect JPEG 2000 images in wireless multimedia sensor networks," Signal Processing: Image Communication, vol. 29, no. 9, pp. 1015–1027,2014.

[21] Z. Shahid and W. Puech, "Visual protection of HEVC video by selective encryption of CABAC bin strings." IEEE Trans. Multimedia, vol. 16, no. 1, pp. 24–36,2014.

[22] Y. Sadourny and V. Conan, "A proposal for supporting selective encryption in JPSEC," IEEE Transactions on Consumer Electronics, vol. 49, no. 4, pp. 846–849,2003.

[23] W. Puech and J. M. Rodrigues, "Crypto-compression of medical images by selective encryption of DCT," in Signal Processing Conference, 2005 13th European. IEEE, 2005, pp.1–4.

[24] A. Massoudi, F. Lefebvre, C. De Vleeschouwer, and F.-O. Devaux, "Secure and low cost selective encryption for jpeg 2000," in Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on. IEEE, 2008, pp.31–38.

[25] M. Ayoup, A. H. Hussein, and M. A. Attia, "Efficient selective image encryption," Multimedia tools and applications, vol. 75, no. 24, pp. 17 171–17 186,2016.

[26] H. Qiu and G. Memmi, "Fast selective encryption method for bitmaps based on GPU acceleration," in Multimedia (ISM), 2014 IEEE International Symposium on. IEEE, 2014, pp. 155–158.

[27] S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to wavelets and wavelet transforms: a primer,"1997.

[28] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," IEEE Transactions on circuits and systems for video technology, vol. 11, no. 3, pp. 301–317, 2001.

[29] J.Franco, G.Bernabé, J.Fernández, and M.E.Acacio, "A parallel implementation of the 2D wavelet transform using CUDA," in Parallel, Distributed and Network-based Processing, 2009 17th Euro micro International Conference on. IEEE, 2009, pp.111–118.

[30] A. Webster and S. E. Tavares, "On the design of S-boxes," in Conference on the Theory and Application of Cryptographic Techniques. Springer, 1985, pp.523–534.

[31] S. Xu, Y. Wang, J. Wang, and M. Tian, "Cryptanalysis of two chaotic image encryption schemes based on permutation and XOR operations," in Computational Intelligence and Security, 2008.CIS'08. International Conference on, vol. 2. IEEE, 2008, pp. 433–437.

[32] G. Zhang and Q. Liu, "A novel image encryption method based on total shuffling scheme," Optics Communications, vol. 284, no. 12, pp. 2775–2780, 2011.

[33] N. Veyrat-Charvillon and F.-X. Standaert, "Mutual information analysis: how, when and why?" in Cryptographic Hardware and Embedded Systems-CHES 2009. Springer, 2009, pp. 429–443.

[34] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," Electronics letters, vol. 44, no.13, pp. 800–801, 2008.

[35] S. Li and X. Zheng, "Cryptanalysis of a chaotic image encryption method," in Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on, vol. 2. IEEE, 2002, pp. II–708.

[36] Kirk, "NVIDIA CUDA software and GPU parallel computing architecture," in ISMM, vol. 7, 2007, pp.103–104.

[37] S. Mittal and J. S. Vetter, "A survey of methods for analyzing and improving GPU energy efficiency," ACM Computing Surveys (CSUR), vol. 47, no. 2, p. 19, 2015.

[38] Daubechies, "Orthonormal bases of compactly supported wavelets," Communications on pure and applied mathematics, vol. 41, no. 7, pp. 909–996, 1988.

[39] A. Alcantara, A. Sharf, F. Abbasinejad, S. Sengupta, M. Mitzenmacher, J. D. Owens, and N. Amenta, "Real-time parallel hashing on the GPU," ACM Transactions on Graphics (TOG), vol. 28, no. 5, p. 154,2009.

[40] A. Bogdanov, M. M. Lauridsen, and E. Tischhauser, "Comb to pipeline: Fast software encryption revisited," in International Workshop on Fast Software Encryption. Springer, 2015, pp.150–171.

[41] Gregg and K. Hazelwood, "Where is the data?

why you cannot debate CPU vs. GPU performance without the answer," in Performance Analysis of Systems and Software (ISPASS), 2011 IEEE International Symposium on. IEEE, 2011, pp.134–144.

[42] Gong, T. Wang, J. Chen, H. Wu, F. Ye, S. Lu, and J. Cong, "An efficient and flexible host-FPGA PCIe communication library," in Field Programmable Logic and Applications (FPL), 2014 24th International Conference on. IEEE, 2014, pp.1–6.

[43] W. Wei, H. Gu, K. Wang, X. Yu, and X. Liu, "Improving cloud based IoT services through virtual network embedding in elastic optical inter-DC networks," IEEE Internet of Things Journal,2018.

[44] K.-T. Cheng and Y.-C. Wang, "Using mobile GPU for general purpose computing–a case study of face recognition on smart phones," in VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on. IEEE, 2011, pp. 1–4.

[45] Zhao, "Fast filter bank convolution for three-dimensional wavelet transform by shared memory on mobile GPU computing," The Journal of Supercomputing, vol. 71, no. 9, pp. 3440–3455, 2015.

[46] Y. Li, K. Gai, L. Qiu, M. Qiu, and H. Zhao, "Intelligent cryptography approach for secure distributed big data storage in cloud computing," Information Sciences, vol. 387, pp. 103–115,2017.

[47] Gai, M. Qiu, Z. Ming, H. Zhao, and L. Qiu, "Spoofing-jamming attack strategy using optimal power distributions in wireless smartgridnetworks,"IEEETransactionsonSmartGrid,vol.8,no. 5, pp. 2431–2439,2017.

