ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computerand Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 6.577 |

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

A Very Safe and Effective Way to Protect Privacy in Cloud Data Storage Configurations

Pankit Arora[1*], Sachin Bharadwaj[2]

Data Analyst, Innovaccer Analytics Pvt. Ltd., India[1]

Assistant Manager (GRC), EXL Service Pvt Ltd. India[2]

ABSTRACT: Pay-as-you-go cloud data outsourcing has becoming more and more popular. But there are privacy concerns with this approach. The typical approach to data privacy is to encrypt sensitive data before outsourcing. When data is encrypted, a trade-off between security and speedy query processing must be negotiated. Existing solutions that employ different encryption techniques cause a large overhead in terms of query performance and data storage, and thus are not suitable for cloud data warehousing. In this study, we propose an efficient Shamir's secret sharing-based additive encryption method (S4) for cloud data warehouse security. S4 overcomes the shortcomings of earlier approaches by reducing overhead and preserving robust data privacy.

KEY WORDS: Cloud Computing, Cloud Storage

I. INTRODUCTION

Data warehouses (DWs) provide a consolidated view of organizations and businesses' data, optimized for reporting and analysis. Greatly enhance decision making. DWs consolidate historical data from direct sources and allow on- line analytical processing (OLAP). Nowadays, data outsourcing scenarios tremendously grow with the advent of cloud computing that offers both cost savings and service benefits. One of the most notable cloud outsourcing services is Database-as-a-Service, where individuals and organizations outsource data storage and management to a Cloud Service Provider (CSP) [1-13]. Naturally, such services allow outsourcing a DW and running OLAP queries. Yet, data outsourcing brings out privacy concerns since sensitive data are stored, maintained and processed by an external third party that may not be fully trusted [14-27]. A typical solution to preserve data privacy is encrypting data locally be-fore sending them to an external server. Secure database management systems (SDBMSs) such as CryptDB implement cryptographic schemes. Paillier's partially homomorphic encryption scheme is notably used in CryptDB to provide high security. However, it induces a highstorage and computation over-head [28-39]. Hence, in this paper, we propose a new Secure Secret Splitting Scheme (S4) that aims at replacing Paillier's scheme in systems such as CryptDB. S4is based on the idea of secret sharing and is efficient both in terms of storage and computing, without sacrificing privacy too much [40-47]. In the remainder of this paper, Section 2 discusses related works about SDBMSs, homomorphic encryption and secret sharing. Section 3 details and discusses S4. Section 4 provides an experimental validation of S4 against Paillier's scheme. Finally, section 5 concludes the paper and hints as future research.

II. RELATED WORKS

2.1. Secure Database Management Systems

CryptDB brings together powerful cryptographic tools to handle query processing on encrypted data without decryption. Encryption in CryptDB is like onion layers that store multiple ciphertexts, i.e., encrypted data, within each other. Each onion layer enables certain kind of query processing and a given security level provided by one encryption scheme. For instance, order-preserving encryption (OPE) enables range queries and additive hemimorphic encryption enables addition over encrypted data. Yet, CryptDB is not perfectly secured since schemes such as OPE reveal some statistical information about plaintext.

MONOMI builds upon CryptDB to allow the execution of analytical work-loads over encrypted data outsourced to the cloud. MONOMI aims at im-proving CryptDB's query processing capability and efficiency based on split client/server

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computerand Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 6.577 |

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

execution. A designer also optimizes physical data layout. Eventually, using a local trusted hardware at the CSP's, such as TrustedDB and CipherBase, is an alternative approach to query encrypted data. How-ever, trusted hardware is limited in computation ability and memory capacity, and also very expensive.

2.2. Homomorphic Encryption

Fully homomorphic encryption (FHE) allows performingarbitrary arithmetic operations over encrypted data without decryption. FHE provides semantic security, i.e., it is computationally impossible to distinguish two cipher texts encrypted from the same plaintext. However, FHE requires so much computing power that it cannot be used in practice. Partially homomorphic encryption (PHE) is more efficientthan FHE.

Paillier's the most efficient additive FHE.With Paillier's scheme, multiplying the encryption of twovalues results in an encryption of the sum of the values, i.e.,$Enc_k(x)$ $Enc_k(y) = Enc_k(x + y)$, where the multiplication isperformed modulo some public-key k. Paillier's scheme is, however, still computation-ally intensive and induces as large ciphertext sizes as 2048 bits. Additionally, modular multiplications become computationally expensive on a large number of records, such as in the fact table of a DW.

2.3. Secret Sharing

Secret sharing divides a secret piece of data into so-called shares that are stored at n participants'. A subset of k n participants is required to reconstruct the secret. In Shamir's, the rst secret sharing scheme, to share a secret v_j, a random polynomial P_{vj} (x) of degree k 1 is rst built. The owner of the secret chooses a prime $p > v_j$ and k 1 random numbers a_1; a_2; :::; a_{k1} from F_p; and sets $a_0 = v_j$ (Equation 1). P_{vj} (x) passes through the point (0; v_j).

$$P_{vj}(x) = a_{k1}x^{k1} + ::: + a_1x + a_0 \quad \text{mod } p \qquad (1)$$

To build n points over P_{vj} (x), a set of n distinct elements inF_p, X = fx_1; x_2; : : : ; x_ng, is chosen such that x_i 6= 0 8i = 1; :::; n: For each participant i, the corresponding share is $v_{i;j} = P_{vj}(x_i)$. For each secret v_j, there are n points (x_i; $v_{i;j}$) through which the polynomial P_{vj} (x) passes. Any k shares form k points (x_i; $v_{i;j}$) i = 1; : : : k, from which polynomial P_{vj} (x) canbe reconstructed using Lagrange interpolation.

S4's driving idea is based on secret sharing, but instead of sharing secrets to n participants' or CSPs', they are stored at one single CSP's. Thus, we avoid the high storage overhead of secret sharing. In S4, each secret v_j is divided into n = k splits $v_{1;j}$; :::; $v_{k;j}$. k 1 splits, $v_{1;j}$; :::; $v_{k1;j}$, are stored at the CSP's and $v_{k;j}$ is stored in a trusted machine, e.g., at the user's. In order to reduce storage overhead at the user's, $v_{k;j}$ is set to be the same for all secrets.

2.4. Splitting and Reconstruction Processes

First, x_k and v_k are randomly set up from F_p, where p is a big prime number, i.e., greater than the greatest possible query answer. For any secret v_j, a random polynomial P_{vj} (x) is builtthat passes through (0; v_j) and (x_k; v_k). To this end, k--2 points (a_i; b_i); i = 1; :::; k 2 are chosen randomly from $F_p$such that a_i 6= x_k and a_i 6= 0 8i = 1; :::; k 2. Given k points (a_1; b_1); (a_2; b_2); :::; (a_{k2}; b_{k2}), (0; v_j) and (x_k; v_k), polynomial P_{vj} (x) is built.

Storing the k 2 ran-dom points is unnecessary because they are not needed for secret reconstruction.To divide v_j into k 1 splits (since (x_k; v_k) is already xed), a set of k 1 distinct elements X = fx_1; x_2; : : : ;$x_{k\ 1}$g is chosen from F_p such that x_i 6= 0and x_i 6= x_k 8i = 1; :::; k 1. Then, splits are $v_{i;j} = P_{vj}(x_i)$. K=(X; (x_k; v_k)) is considered as a private key for S4 and must be kept hidden from the CSP. Tore construct secret v_j, its k 1 splits must be retrieved from the CSP. Given points (x_i; $v_{i;j}$), i = 1; :::; k 1 and (x_k; v_k), which is stored at the user's, polynomial P_{vj} (x) can be reconstructed.

2.5 Summation Queries

Let a relational table T consist of one attribute A (additional attributes, if any, can be processed similarly). Suppose T has m records. We denote by v_j the j^{th} value of A. For attribute A in T , k 1 attributes A_i, i = 1; :::; k 1 are created in table T^0 atthe CSP's, where each attribute A_i stores the i^{th} splits. Without loss of generality, we assume integer data type for other data types can be transformed into integers before splitting. S4 allows summation queries to be computed directly at the CSP's. Consider a query that sums q values of A.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computerand Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 6.577 |

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

III. ANALYSIS

Paillier's PHE is semantically secure, but it is too expensivein terms of cipher-text storage space and query response time. S4 proposes a classical trade-o with a lower level of security, but better storage and response time efficiency. Let us consider a scenario where the CSP is said honest but curious, which is a widely used adversary model for cloud data outsourcing . Such a CSP faithfully complies to any service-level agreement and, in our particular case, storesdata, runs queries and provides results without alteration, malicious or otherwise. Yet, the CSP may access data and infer information from queries and results.

Privacy in S4 relies on the fact that a secret value is only retrievable by the user via private key K. As in secret sharing, it is indeed guaranteed that at least k splits and X are necessary to reconstruct a secret, while the CSP has access to only k 1 splits.

Both X and the k^{th} split, i.e., K, are stored at the user's. However, the CSP still has access to linear combinations of splits, which provide some information. Still, the higher k is, the more di cult it is to interpret linear combinations of splits. Thus, k is the prime securityparameter in S4. Experiments in Section 4 provide hints for choosing k.

Moreover, if some secrets are known by the CSP, e.g., through public communication of a company to its shareholders. For example, if the CSP knows secrets v_1; :::; v_{k1}. Also knowing the correspond-ing splits $v_{1;j}$; :::; $v_{k1;j}$ 8j 2 [1; k 1], the CSP can recover the Lagrange basis polynomials $`_i(0)$ 8i 2 [1; k] and recover all secrets. However, the CSP must know at least k 1 secrets to do so. Moreover, we also propose leads to address this problem in next section.

IV. EXPERIMENTAL EVALUATION

4.1Experimental Setup

We implement S4 in C using compiler gcc 4.8.2. S4's source code is freely available on-line[1]. Experiments related to Paillier's PHE exploit the libpaillier standard C library. All mathematical computations use the GNU Multiple Precision Arithmetic Library (GMP). Eventually, we conduct our experiments on an Intel Core i7 3.10 GHz PC with 16 GB of RAM running Linux Ubuntu 15.05.

We compare S4 and Paillier's PHE using simple synthetic datasets, i.e., 32-bit unsigned integers generated uniformly at random from the integer range [103; 104]. We scale up the number of records m such that m 2 (103, 104, 105, 106), forming four distinct datasets.

In S4, we vary k from 8 to 64, higher values of k inducing too long execution Pm times. Prime p must be greater than the greatest query answer, e.g., p > j=1 vj. In Paillier's PHE, we use a key size of 1024 bits, which induces ciphertexts of 2048 bits. Such key size is the absolute minimum to achieve security.

4.2 Encryption and Decryption Time

Figure 1 plots the time of secret splitting in S4 and secret encryption in Paillier's scheme with respect to m. It shows that encryption time in S4 is lower than Paillier's when k 16, and then becomes higher when k 16. Secret splitting consists in building a random polynomial by randomly choosing k 2 points. Hence, splitting time increases with k. Figure 1 actually illustrates the tradeo between S4's security and encryption efficiency with respect to Paillier's PHE.

Figure 2 plots the time of secret reconstruction in S4 and secret decryption in Paillier's scheme with respect to m. With the selected values of k, decryption is faster with S4 than with Paillier's PHE. This is mainly because Paillier's scheme needs m expensive modular multiplications of large, 2048-bitnumbers for decryption, while secret reconstruction in S4 works by polynomial interpolation over k points andevaluating the polynomial in one single point.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computerand Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 6.577 |

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

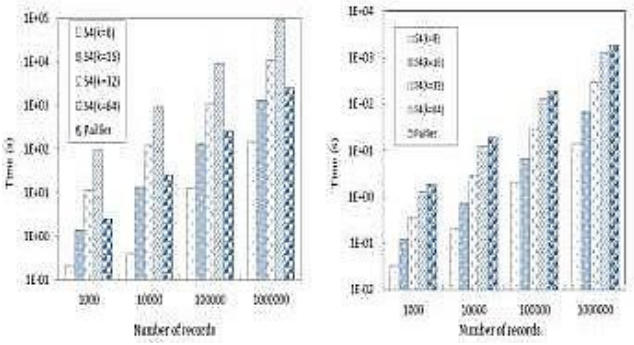

Figure 1. Splitting/encryption time; Figure 2 Reconstruction/decryption time

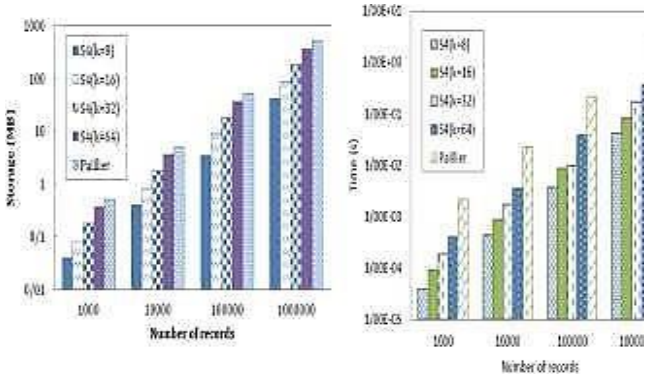

Figure 3. Storage overhead; Figure 4. Summation execution time

4.3 Space Overhead

Figure 3 plots the storage required by S4 and Paillier's PHE with respect to m. With the selected values of k, S4's storage overhead is always much smaller than that of Paillier's PHE since Figure 3's y axis follows a logarithmic scale. Paillier's scheme indeed produces 2048-bit cipher texts. Thus, its storage overhead is m 2048. With S4, each value is split into k 1 values. Thus, S4's storage overhead is m (k 1) times plaintext size.

4.4 Query Processing Time

Figure 4 plots summation query processing times over all records in each dataset, for both S4 and Paillier's PHE, with respect to m. It shows that, with the selected values of k, query execution time in S4 is lower than that of Paillier's scheme. This is because Paillier's scheme requires m expensive modular multiplications to compute a sum, while S4 computes only (k 1) m simple modular additions.

V. CONCLUSION

In this paper, we introduce S4, a new cryptographic scheme that supports summation queries in cloud-based OLAP. We experimentally show that S4 is much more efficient than Paillier's PHE in terms of query response time and space overhead. Thus, replacing Paillier's scheme with S4 in secure DBMSs such as CryptDB and MONOMI can improve analytical query processing in cloud DWs. Moreover, we also plan a variant of S4 for computing multiplications. However, we achieve performance gains through a slight degradation of se-curity, especially when an adversary has knowledge of secret values. Although it is definitely acceptable in some cloud DW and OLAP scenarios, e.g., public aggregate data might not actually yield secrets, i.e., ne-grained data, we will devote future research to strengthen S4 against such threats. More precisely, we plan to introduce noise, as in many cryptographic problems such as approximate-GCD or LWE. For instance, instead of sharing vj, we could share 10r vj + noise. By doing so, security is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computerand Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 6.577 |

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

intuitively enhanced while the whole process remains correct, provided r is sufficiently large and noise sufficiently small.

REFERENCES

1. A. Serageldin, H. Alturkostani, and A. Krings, "On the reliability of DSRC safety applications: a case of jamming," in International Conference on Connected Vehicles and Expo, 2013, pp. 501–506.
2. B.K. Chaurasia, R.S. Tomar, S. Verma, G.S. Tomar, Suitability of manet routing protocols for vehicular ad hoc networks, in: 2012 International Conference on Communication Systems and Network Technologies, IEEE, 2012, pp.334–338.
3. C. Laurendeau and M. Barbeau, "Threats to security in DSRC/WAVE," in Proc. 5th International Conference on Ad-Hoc Networks & Wireless, LNCS 4104, 2006, pp. 266–279.
4. E. Hamida, H. Noura, and W. Znaidi, "Security of cooperative intelligent transport systems: standards, threats analysis and cryptographic countermeasures," Connected Vehicles, V2V Communications, and VANET, vol. 4, no. 3, pp. 380–423, 2015.
5. F.K. Karnadi, Z.H. Mo, K.-c. Lan, Rapid generation of realistic mobility models for vanet, in: 2007 IEEE Wireless Communications and Networking Confer-ence, IEEE, 2007, pp.2506–2511.
6. H. Hasbullah, I. Soomro, and J. Ab Manan, "Denial of service (DOS) attack and its possible solutions in VANET," International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, vol. 4, no. 5, pp. 813–817, 2010.
7. H.-M. Zimmermann, I. Gruber, C. Roman, A Voronoi-based mobility model for urban environments, in: 11th European Wireless Conference 2005-Next Generation Wireless and Mobile Communications and Services, VDE, 2005, pp.1–5.
8. J. Härri, F. Filali, C. Bonnet, M. Fiore, Vanetmobisim: generating realistic mo-bility patterns for vanets, in: Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks, ACM, 2006, pp.96–97.
9. J. Harri, M. Fiore, Vanetmobisim–vehicular ad hoc network mobility exten-sion to the canumobisim framework, Institut Eurécom Department of Mobile Commu 6904 (2006) 1–19.
10. J. Zhao, G. Zucchelli, and M. Roggero, "Design of FMCW radars for active safety applications," http://embedded-computing.com/articles/design-fmcw-radars-active-safety-applications/, 2015.
11. L. Bononi, M. Di Felice, M. Bertini, E. Croci, Parallel and distributed simulation of wireless vehicular ad hoc networks, in: Proceedings of the 9th ACM In-ternational Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, ACM, 2006, pp.28–35.
12. M. Brooker, "Mutual interference of millimeter-wave radar systems," IEEE Transactions on Electromagnetic Compatibility, vol. 49, pp. 170–181, 2007.
13. M. Piorkowski, M. Raya, A.L. Lugo, P. Papadimitratos, M. Grossglauser, J.-P. Hubaux, TRANS: realistic joint traffic and network simulator for vanets, Mob. Comput. Commun. Rev. 12 (2008) 31–33.
14. N. Li and Y. Zhang, "A survey of radar ECM and ECCM," IEEE Trans. Aerospace and Electronic Systems, vol. 31, no. 3, pp. 1110–1120, 1995.
15. N. Wisitpongphan, O.K. Tonguz, J.S. Parikh, P. Mudalige, F. Bai, V. Sadekar, Broadcast storm mitigation techniques in vehicular ad hoc networks, IEEE Wirel. Commun. 14 (2007) 84–94.
16. Q. Chen, T. Roth, T. Yuan, J. Breu, F. Kuhnt, M. Zollner, M. Bogdanovic, C.Weiss, J. Hillenbrand, and A. Gern, "DSRC and radar object matching for cooperative driver assistance systems," in IEEE Intelligent Vehicles Symposium, 2015, pp. 1348–1354.
17. Q. Xu, T. Mak, J. Ko, and R. Sengupta, "Vehicle-to-vehicle safety messaging in DSRC," in Proceedings of the 1st ACM International Workshop on Vehicular ad hoc Networks, 2004, pp. 19–28.
18. R. Chauhan, "A platform for false data injection in frequency modulated continuous wave radar," http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=4983&context=etd, 2014.
19. R. Fernandes, P.M. d'Orey, M. Ferreira, Divert for realistic simulation of hetero-geneous vehicular networks, in: The 7th IEEE International Conference on Mo-bile Ad-Hoc and Sensor Systems (IEEE MASS 2010), IEEE, 2010, pp.721–726.
20. R. Mangharam, D.S. Weller, D.D. Stancil, R. Rajkumar, J.S. Parikh, Groovesim: a topography-accurate simulator for geographic routing in vehicular networks, in: Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks, ACM, 2005, pp.59–68.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computerand Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 6.577 |

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

21. S. Senthilkumar, R. Nithya, P. Vaishali, R. Valli, G. Vanitha, & L. Ramachanndran, "Autonomous navigation robot", International Research Journal of Engineering and Technology, vol. 4, no. 2, 2017.
22. S. Jaap, M. Bechler, L. Wolf, Evaluation of routing protocols for vehicular ad hoc networks in typical road traffic scenarios, 2005, pp.584–602.
23. S. Maddio, A. Cidronali, M. Passafiume, G. Collodi, and G. Manes, "Interference cancellation for the coexistence of 5.8 GHz DSRC and 5.9 GHz ETSI ITS," in IEEE MTT-S International Conference on Microwaves for Intelligent Mobility, 2015, pp. 1–4.
24. S. Roome, "Digital radio frequency memory," Electronics & Communication Engineering Journal, vol. 2, no. 4, pp. 147–153, 1990.
25. S.-Y. Wang, C.-L. Chou, Nctuns Simulator for Wireless Vehicular Ad Hoc Net-work Research, Ad Hoc Networks: New Research, Nova Science Publishers, 2009.
26. T. Fujiki, M. Kirimura, T. Umedu, T. Higashino, Efficient acquisition of local traffic information using inter-vehicle communication with queries, in: 2007 IEEE Intelligent Transportation Systems Conference, IEEE, 2007, pp.241–246.
27. T. Jeyaprakash, R. Mukesh, A survey of mobility models of vehicular adhoc networks and simulators, Int. J. Electron. Inf. Eng. 2 (2015) 94–101.
28. T. Zhang, H. Antunes, and S. Aggarwal, "Defending connected vehicles against malware: challenges and a solution framework," IEEE Internet of Things Journal, vol. 1, pp. 10–21, 2014.
29. V. Richard, Millimeter wave radar applications to weapons systems, USA Ballistic Research Laboratories, 1976.
30. V.D. Khairnar, S. Pradhan, Comparative study of simulation for vehicular ad-hoc network, preprint, arXiv:1304 .5181, 2013.
31. W. Zhang, H, Zeng, Y, Li, and X. Wang, "Polarimetric radar performance test of signal processing for anti-active jamming," in IET International Radar Conference, 2009, pp. 1–4.
32. X. Qiao, T. Jin, X. Qi, M. Zhang, S. Yuan, and Q. Zhang, "Anti-millimeter wave polarization agile active jamming," in Proceedings of the International Conference on Microwave and Millimeter Wave Technology, 2007, pp. 1–4.
33. Y.P. Fallah, C. Huang, R. Sengupta, H. Krishnan, Congestion control based on channel occupancy in vehicular broadcast networks, in: 2010 IEEE 72nd Ve-hicular Technology Conference-Fall, IEEE, 2010, pp.1–5.