THREAT MODEL AND DEFENSE SCHEME FOR SIDE- CHANNEL ATTACKS IN CLIENT SIDE DEDUPLICATION

V.Savithri[1], Y.Mounika[2]

[1]Assistant Professor,Dept. of MCA,Audisankara College of Engineering & Technology (AUTONOMOUS), Gudur, AP, India.

[2]PG Scholar,Dept. of MCA,Audisankara College of Engineering& Technology (AUTONOMOUS), Gudur, AP, India.

ABSTRACT:

In cloud storage, client-side deduplication is widely used to reduce storage and communication costs. In client-side deduplication, if the cloud server detects that the user's outsourced data have been stored, then clients will not need to reupload the data. However, the information on whether data need to be uploaded can be used as a side-channel, which can consequently be exploited by adversaries to compromise data privacy. In this paper, we propose a new threat model against side-channel attacks. Different from existing schemes, the adversary could learn the approximate ratio of stored chunks to unstored chunks in outsourced files, and this ratio will affect the probability that the adversary compromises the data privacy through side-channel attacks. Under this threat model, we design two defense schemes to minimize privacy leakage, both of which design interaction protocols between clients and the server during deduplication checks to reduce the probability that the adversary compromises data privacy. We analyze the security of our schemes, and evaluate their performances based on a real-world dataset. Compared with existing schemes, our schemes can better mitigate data privacy leakage and have a slightly lower communication cost.

1 INTRODUCTION

The rapid growth of data volume has required cloud service providers to use the data deduplication to reduce storage and communication costs[1–3]. After the deduplication, the cloud server couldidentify data redundancy and only store a single copy of user data. Based on the deduplication location, deduplication can be classified as a server or client side deduplication[2]. In the server-side deduplication[4,5], clients always upload data to the cloud server. After receiving the uploaded data, the cloud server performs deduplication to save

storage space. In the client-side deduplication, clients compute hash values for user data as data tags and send them to the cloud server. After receiving data tags, the cloud server checks whether the data have been stored based on data tags and returns deduplication responses to clients. For example, if the data have been stored, then the deduplication response will be set to 1. Otherwise, it will be set to 0, as shown in Fig. 1. When a client receives a response of 0, it should upload the data; otherwise, it does not need to upload the data. Compared with server-side deduplication, client-side deduplication can reduce storage and communication costs and has been widely used in cloud storage[6]. However, deduplication responses in the client-side deduplication could be exploited by adversaries to launch side-channel attacks[7] to violate data privacy, because the data transmission between clients and the server can be monitored by adversaries and data tags could be used to detect the data existence in the server.

2. LITERATURE SURVEY

2.1 Data deduplication

Data deduplication is an effective method to save storage overhead for cloud storage systems[2,10,11]. The cloud server can detect redundant data in cross-user uploaded data by deduplication, and only store unique data. Based on the data granularity, deduplication can be divided into file-level or chunk-level deduplication. In the file-level deduplication[11,12], the user file is treated as the basic unit for deduplication. By contrast, in the chunk-level deduplication[5,13–16], clients divide user files into chunks and the chunk is the basic unit. Compared with file-level deduplication, chunk-level deduplication usually has a higher deduplication ratio.

In the chunk-level client-side deduplication, we suppose that the outsourced file is F , the client first splits F into multiple chunks fmg, where m denotes the data chunk. Then, the client computes the hash values fhg for chunks as data tags and sends them to the cloud server. The cloud server uses data tags for deduplication check. If the data tag hi is not found, then the server will set the deduplication response to 0 and return it to the client. Then, the client needs to upload the chunk mi and the server stores .hi D H.mi /; mi /, where H.·/ denotes a cryptographic hash function. Otherwise, if hi has been found in the server, then the client will receive the response of 1 and does not need to upload mi .

2.2 Side-channel attacks

Harnik et al. found that the client-side deduplication can be used as a side-channel. In cross-user client-side deduplication, the adversary can establish a side-channel using deduplication responses and violate the privacy of user data. For

example, adversaries can use the side- channel to launch the following attacks:

- Identifying the existence of specific files: Suppose an adversary wants to learn whether a specific file F has been uploaded to the cloud server by other users, it can observe deduplication responses when uploading F . If the server asks the adversary to upload F , then it learns that F is not stored on the server. Otherwise, F has already been uploaded by other users.

- Establishing a covert channel: Multiple Badversaries can establish a covert channel to communicate with one another through deduplication responses

3. PROPOSED WORK

- We propose a new threat model against side- channel attacks in the client-side deduplication. Different from existing defense schemes[8,9], our threat model considers a stronger adversary, which can learn the approximate ratio of stored chunks to unstored chunks in outsourced files. We assume that the adversary could maliciously construct outsourced files with a certain number of stored chunks (uploaded by it before) and unstored chunks (random chunks). Then, it places a specific target chunk in maliciously constructed files and performs side-channel attacks by constantly uploading constructed files with different ratios of stored chunks to unstored chunks. The adversary can observe deduplication responses and data transmissions during the deduplication check, and try to learn the existence of the target chunk.

- Under our threat model, we propose two defense schemes against side-channel attacks, namely basic and enhanced schemes. We argue that the reason why the adversary could launch side-channel attacks is that the deduplication responses leak the information of data existence. Therefore, both our schemes design interaction protocols between clients and the server to disturb the correlation between deduplication responses and data existence.

- We analyze the security for our basic and enhanced schemes and two existing schemes[8,9] under our threat model, and then evaluate the computational and communication overheads in these four schemes based on a real-world dataset. The results of the security analysis and performance evaluation show that our schemes can effectively mitigate data privacy leakage and reduce the communication cost for the system

3.1 Architecture

Our schemes consist of two entities: clients and a cloud server.

- Clients: To outsource a user file F to the cloud server, the client divides F into fix-sized chunks fmg and computes the hash values fhg for these chunks as data tags. The client sends data

tags to the server to ask for deduplication responses, which determine how the client uploads data.

- Cloud server: The cloud server provides data storage services for multiple users and performs cross- user chunk-level client-side deduplication to minimize storage and communication costs. After receiving the data tags uploaded by clients, the cloud server checks whether they were stored before. Then, it sends deduplication responses back to clients.

4. CONCLUSION

Although the client-side deduplication can be used to save storage and communication costs for cloud storage systems, deduplication responses are easily to be used as a side-channel by the adversary to violate data privacy. We argue that the threat models in existing defense schemes against side-channel attacks need to be strengthened. Thus, we propose a new threat model, that considers an adversary that could construct files containing a certain number of stored and unstored chunks to launch side-channel attacks. We propose basic and enhanced defense schemes against this kind of attack. The security analysis and performance evaluation show that the proposed schemes can effectively mitigate the privacy leakage of user outsourced data, and can effectively reduce the communication cost for the system.

5. REFERENCES

[1] W. Xia, H. Jiang, D. Feng, F. Douglis, P. Shilane, Y. Hua, M. Fu, Y. C. Zhang, and Y. K. Zhou, A comprehensive study of the past, present, and future of data deduplication, Proceedings of the IEEE, vol. 104, no. 9, pp. 1681–1710, 2016.

[2] Y. Shin, D. Koo, and J. Hur, A Survey of secure data deduplication schemes for cloud storage systems, ACM Computing Surveys, vol. 49, no. 74, pp. 1–38, 2017.

[3] D. T. Meyer and W. J. Bolosky, A study of practical deduplication, presented at 9th USENIX Conference on File and Storage Technologies, San Jose, CA, USA, 2011.

[4] J. Li, Z. Yang, Y. Ren, P. Lee, and X. Zhang, Balancing storage efficiency and data confidentiality with tunable encrypted deduplication, presented at 15th EuroSys Conference on Computer Systems, Heraklion, Greece, 2020.

[5] J. Li, P. P. C. Lee, Y. Ren, and X. Zhang, Metadedup: Deduplicating metadata in encrypted deduplication via indirection, presented at 35th Symposium on Mass Storage Systems and Technologies (MSST), Santa Clara, CA, USA, 2019.

[6] M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber, and E. Weippl, Dark clouds on the horizon: Using cloud storage as attack vector

and online slack space, presented at 20th USENIX Security Symposium, San Francisco, CA, USA, 2011.

[7] D. Harnik, B. Pinkas, and A. Shulman-Peleg, Side-channels in cloud services: Deduplication in cloud storage, IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010.

[8] Z. Pooranian, K. Chen, C. Yu, and M. Conti, RARE: Defeating side-channels based on data-deduplication in cloud storage, presented at IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, HI, USA, 2018.

[9] C. Yu, S. P. Gochhayat, M. Conti, and C. Lu, Privacy aware data deduplication for side-channel in cloud storage, IEEE Transactions on Cloud Computing, vol. 8, no. 2, pp. 597– 609, 2020.

10] V. Srikanth. "ANALYZING THE TWEETS AND DETECT TRAFFIC FROM TWITTER ANALYSIS" v srikanth | INTERNATIONAL JOURNAL OF MERGING TECHNOLOGY AND ADVANCED RESEARCH IN COMPUTING, 20 MARCH. 2017. http://ijmtarc.in/Papers/Current%20Papers/IJMTARC-170309.pdf

[11] V. Srikanth. "A NOVEL METHOD FOR BUG DETECTION TECHNIQUES USING INSTANCE SELECTION AND FEATURE SELECTION" v srikanth | INTERNATIONAL JOURNAL OF INNOVATIVE ENGINEERING AND MANAGEMENT RESEARCH, 08 DECEMBER. 2017. https://www.ijiemr.org/public/uploads/paper/976_approvedpaper.pdf

[12] V. Srikanth. "SECURED RANKED KEYWORD SEARCH OVER ENCRYPTED DATA ON CLOUD" v srikanth | INTERNATIONAL JOURNAL OF INNOVATIVE ENGINEERING AND MANAGEMENT RESEARCH, 08 Febraury. 2018. http://www.ijiemr.org/downloads.php?vol=Volume-7&issue=ISSUE-02

[13] V. Srikanth. "WIRELESS SECURITY PROTOCOLS (WEP,WPA,WPA2 & WPA3)" v srikanth | Journal of Emerging Technologies and Innovative Research (JETIR), 08 mAY. 2019. https://www.jetir.org/papers/JETIRDA06001.pdf

[14] V. Srikanth, et al. "Detection of Fake Currency Using Machine Learning Models." Deleted Journal, no. 41, Dec. 2023, pp. 31–38. https://doi.org/10.55529/ijrise.41.31.38.

[15] V. Srikanth, et al. "A REVIEW ON MODELING AND PREDICTING OF CYBER HACKING BREACHES." 25 Mar. 2023, pp. 300–305. http://ijte.uk/archive/2023/A-REVIEW-ON-MODELING-AND-PREDICTING-OF-CYBER-HACKING-BREACHES.pdf.

[16] V. Srikanth, "DETECTION OF PLAGIARISM USING ARTIFICIAL NEURAL NETWORKS." 25 Mar. 2023, pp. 201–209. http://ijte.uk/archive/2023/DETECTION-OF-PLAGIARISM-USING-ARTIFICIAL-NEURAL-NETWORKS.pdf.

[17] V. Srikanth, "CHRONIC KIDNEY DISEASE PREDICTION USING MACHINELEARNINGALGORITHMS." 25 January. 2023, pp. 106–122. http://ijte.uk/archive/2023/CHRONIC-KIDNEY-DISEASE-PREDICTION-USING-MACHINE-LEARNING-ALGORITHMS.pdf.

[18] Srikanth veldandi, et al. "View of Classification of SARS Cov-2 and Non-SARS Cov-2 Pneumonia Using CNN". journal.hmjournals.com/index.php/JPDMHD/article/view/3406/2798.

[19] Srikanth veldandi, et al. "Improving Product Marketing by Predicting Early Reviewers on E-Commerce Websites." Deleted Journal, no. 43,

Apr. 2024, pp. 17–25. https://doi.org/10.55529/ijrise.43.17.25.
[20] Srikanth veldandi, et al.“Intelligents Traffic Light Controller for Ambulance.” Journal of Image Processing and Intelligent Remote Sensing, no. 34, July 2023, pp. 19–26. https://doi.org/10.55529/jipirs.34.19.26.
[21] O. Heen, C. Neumann, L. Montalvo, and S. Defrance, Improving the resistance to side-channel attacks on cloud storage services, presented at 5th International Conference on New Technologies, Mobility and Security(NTMS), Istanbul, Turkey, 2012.

[22] Y. Shin and K. Kim, Differentially private client-side data deduplication protocol for cloud storage services, Secur. Commun. Networks., vol. 8, no. 12, pp. 2114–2123, 2015.

AUTHOR’SPROFILE

Ms.V. SAVITHRI currently she is working as Assistant professor in Audisankara college of Engineering andTechnology (AUTONOMOUS),NH-5,BypassRoad,Gudur,Tirupati(Dt.),Andhra Pradesh, India.

Y. MOUNIKAispursuing MCA fromAudisankara Collegeof Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.