International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Secure Cloud Assisted Data Analysis Process for Smart Phone Tasks

Kumudashree H S[1], Triveni C L[2]

[1, 2]VTU University, Malnad College of Engineering, Salagame Road, 573202, Hassan

Abstract: ***The omnipresence of smart cell phones makes the versatile crowdsourcing conceivable, where the requester (task requester) can crowd source information from the smart cell phone user (cell phone clients) by utilizing their sensor-rich cell phones. In any case, information gathering, information total, and information examination have turned out to be testing issues for an asset obliged requester when information volume is to a great degree vast, i.e., enormous information. Specifically to information examination, set activities, including crossing point, association, and complementation, exist in most enormous information investigation for sifting repetitive information and preprocessing crude information. Confronting challenges as far as restricted calculation and capacity assets, cloud-helped methodologies may fill in as a promising method to handle huge information investigation issue. Cell phone users are not willing to provide their data. If the security of their detecting information and personality are not very much protected in the untrusted cloud. In this work, we propose to utilize cloud to process set task for the requester, in the meantime users' information security and characters protection are very much saved. Additionally, the requester can confirm the rightness of set activity comes about. We likewise stretch out our plan to help information preprocessing, with which invalid information can be barred before information examination. By utilizing clump check and information refresh strategies, the proposed conspire incredibly decreases the computational cost. Broad execution examination and investigation in view of genuine cloud framework have demonstrated both the achievability and productivity of our proposed conspire.***

Keywords: Crowdsourcing, Privacy, Verifiable Set Operation

1. Introduction

In recent times, mobile phones have been riding the wave of Moore's law with rapid improvements in mobile crowdsourcing. The mobile phones of today have evolved from merely being phones to fully-fledged computing, sensing and communication devices. It is thus hardly surprising that over 5 billion people globally have access to mobile phones. These advances in smart phone technology coupled with their ubiquity have paved the wave for an exciting new paradigm for mobile crowdsourcing.

Mobile Crowdsourcing[1] enables a task owner to get hold of data from a hefty number of smartphone users, and further act upon data analysis on the aggregated data[3]. The task owner is also known as the requester, while participating the smartphone users are mobile workers who will collect and/or sense the data for the requester. With the advance of the low cost sensing devices, many sensors have been embedded on mobile devices, such as GPS, accelerator, gyroscope, digital compass, temperature sensors, etc. More sensors used for measuring humidity, air quality, chemical, barometer, and biomedical information can be equipped into smartphones or allied via wireless technologies. These reasonably priced sensor-rich smartphones make them proficient of sensing the environment in the region of people and people's physiological data also. In mobile crowdsourcing, a requester can make use of the data crowd sourced from mobile workers to complete certain tasks. for instance, a transportation management bureau can make use of the speed data reported from the commuters to analyze the traffic condition. perceptibly, mobile crowdsourcing has many compensation: first, the omnipresent smartphone users can cover a large geographic area, which makes the statistics and information miscellaneous and rich; second, the requester does not need to install specific sensor networks or employees to amass the targeted data; third, workers can receive plunder such as status as well as profits from the Crowdsourcing chipping in.

Cloud computing[8] offers many benefits for companies, public institutions and individuals willing to store and process their data in the cloud, such as, dynamically scalable resources, improved agility and manageability, scalability, availability and universal data access independently of geographical location, thus providing computational power and flexibility. Most important, cloud computing generally implies costs savings since it reduces infrastructure and maintenance costs, thus providing cheap storage capacity and computing. However, security concerns about data loss or leakage since the lack of direct control over the storage and management of the outsourced data have proved to be a real threat that prevents many customers from migrating to the cloud. From another perspective, cloud users may have concerns about what Cloud Service Providers (CSPs) intend to do with their (potentially confidential) data. Cloud computing has given CSPs the opportunity to analyze and exploit large amounts of personal data. For example, a recent privacy policy in Google5 specifies that whatever information a user decides to share through any Google service can be used, reproduced, modified or distributed by Google with the aim of improving or promoting its services (e.g., the Gmail filtering system scans the content of our emails to serve personalized ads). Data collected by CSPs can be used to benefit the users (e.g., by providing personalized services) but, at the same time, they may raise privacy concerns.

To mitigate these problems and to regain the user's control over the protection applied to her confidential data prior outsourcing them to the cloud, several mechanisms have been proposed. They mainly apply a certain kind of data

protection on the client side so that only protected outcomes are outsourced to the cloud and so that only the data owner is able to correctly reconstruct the data retrieved from the cloud.

2. Related Work

Encryption is a natural solution for enforcing data protection in the cloud. Several solutions based on public and symmetric key cryptography have been presented for cloud storage in which data protection is provided via encryption. Encryption is performed before the data are transmitted and stored in the cloud, being decrypted only after having been returned to the data owner[13]. These solutions have been usually enforced as trusted encryption proxies. For example, CipherCloud provides a secure gateway located in a trusted environment (i.e., the user's local network) targeted at several popular Software-as-a-Service CSPs (e.g., Gmail, Salesforce, Amazon, etc.). It applies encryption to specific user data (e.g., mail bodies and subject, chat messages, etc.) before storing them in the cloud. Cloud services are replicated in the secure gateways (i.e., web interfaces, business logics) to provide coherent results to operations like searching or sorting. Encryption keys are managed and stored locally under the control of the user.

The communication between the CSP and the client device is captured and "reverse engineered" in order to enrich it with security features (data encryption) that are transparent to both the CSP and the client. However, the communication protocol may often change, which would require a continuous update of the provided security features, a difficult task that could seriously impair the reliability and availability of the service. In addition, the CSP itself might also implement specific measures to prevent this approach if it is uncomfortable with users that systematically upload encrypted contents to its servers. This is especially relevant for CSPs that offer their services free of charge because they expect to gain profits from the analysis of users' data (e.g., Google have legal possibilities to access these data, due to the terms of service agreement) and which may ban users that only provide encrypted –and thus, useless- data[9].

Encrypting the whole data uploaded to the CSP at the client side implies the loss of several degrees of magnitude in efficiency with regard to both storage and processing, which in the case of cloud computing it would mean defeating its own purpose, because one of the main motivation for moving to the cloud, in addition to the provided functionalities, is saving costs. Moreover, the management of encryption keys may add new security risks at the client side.

Because sensitive data are systematically encrypted and stored at the CSP, which is supposed to be unaware of this fact, functionalities offered by the CSP can yield gibberish outcomes. In this case, only a reduced set of functionalities can be preserved (basically, data storage and plain retrieval), or encryption can be applied only on data that are not processed in the cloud (e.g., binary files), or cloud services must be replicated at the trusted gateway. In the latter case, the gateway is forced to redundantly store unencrypted data and to re-implement and reverse-engineer some cloud services, thus defeating the whole purpose of data and computation outsourcing. Even though in recent years some cryptographic solutions have been proposed with a limited support for a number of operations over encrypted data (mainly searches), complex operations would require from solutions like homomorphic encryption[13], which are still far from being efficiently applicable in a real setting.

The vast majority of users are not familiar with the fundamental concepts of cryptography and many of them are not capable of properly managing keys or certificates; thus, the effectiveness and security of cryptographic solutions may be compromised because of a negligent management of cryptographic materials.

3. Problem Statement

In particular to the collected data, it might not be just a single value reported in a period of time. Instead, we consider a more general data type requested from the requester, which could be a range of data including multiple values or even a large set of elements without order. Set operations are often used in data processing. For example, a travel agency wants to know the most popular places that the tourists have visited during holidays. Here, the data from a worker (tourist) will be a set, and thus the requester (travel agency) needs to find the intersection of all sets. Set union may be used to merge different databases collected from different database owners. Set difference is useful when a requester wants to find the unique feature of one database compared to another. When the number of workers is very large, the requester requires a huge amount of storage space for storing the crowd sourced big data even if each worker's data is relatively small. As a result, a storage limited requester is not able to handle the above task. Taking a step further, even if the requester can store all collected "Big Data", the data processing and analysis may be another stumbling block when he/she lacks computation capability. Therefore, the set operation problem over the collected data might be overwhelming.

In our work, we require that workers' data privacy and identity privacy should be protected. Specifically, the cloud should not know the plaintext of the data sets or the exact source of a data set[11]. We formulate this problem as a privacy-preserving set operation. The data privacy is preserved through ElGamal encryption and a keyed hash function. While the identity privacy is achieved through ring signature. The requester will get the computation result from the cloud together with proof information. Therefore, we formulate this problem as a Verifiable computation outsourcing problem.

4. Proposed Framework

The system proposes an efficient solution for the set operation in big data analysis based on the data collected from mobile crowdsourcing and introduce the cloud as an intermediate entity to the traditional mobile crowdsourcing, where worker's data privacy and identity privacy are well protected. For requesters, they can verify the correctness of

computation results retrieved from the cloud. Batch verification and data update also proposed for Data dynamism, integrated with batch auditing process and hence reduces computational costs of the system.

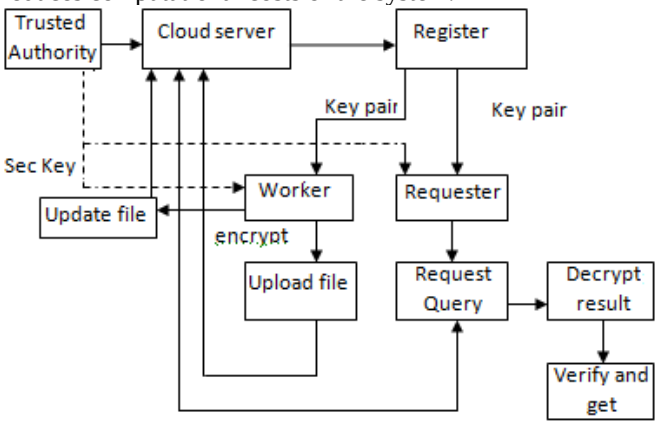


Fig 1: System Architecture

The following steps take place
1) System initialization
2) Crowd sourcing
3) Data Encryption
4) Set operation and verification
5) Data Dynamic and verification

System initialization

The system consists of Trust Authority (TA), Cloud, Requester, Workers. TA is responsible for initializing the whole system which includes registering workers, requesters and the cloud, generating public parameters, and distributing keys, and maintaining the system. TA may be offline unless a dispute arises. The requester wants to obtain the intersection set of the workers' data sets. However, due to his/her limitation on the storage and computation capability, the requester will delegate storage and most of the computation tasks to the cloud.

The cloud receives the delegation requests from the requester and the encrypted data sets from mobile workers, and then it computes the intersection set for the requester. The cloud also needs to provide some proof information to prove the correctness of the result.

Crowd sourcing

Under Crowdsourcing, users cover a large geographic area, which makes the data and information diverse and rich; second, the requester does not need to deploy specific sensor networks or employees to collect the targeted data. User uploaded data are stored in the cloud after encryption.

Data Encryption

The data privacy is preserved through encryption. The requester will get the computation result from the cloud together with a proof information. Every worker W_i generates his data set S_i, and encrypts it with p_k. The data will be signed with ring signature before sending to the cloud. After receiving encrypted data sets from all workers, the cloud verifies the authenticity of each of them, and computes the intersection set based on the encrypted data sets. Then the cloud sends the result together with its corresponding proof information to the requester. Finally, the requester decrypts the result and checks its correctness.

Set operation and verification

Set operation is performed when the workers are uploading data to cloud. Supposing the range limit set defined by the requester is S_R, which means all valid data should be within S_R. Worker W_i has data set S_i. There are four possible relationships between S_R and S_i. When the requester delegates the set intersection computation to the cloud, the cloud needs to excludes set S_i if the relationship between S_i and S_R, which means S_i contains at least an element that's not in S_R.

Data Dynamic and verification

To reduce the cost of processing the operation on collected data, we need to carefully exam the reported data. Normally, the requester has specific range requirements on the data set. The requester may determine that only sets of a specific range of tourist sites are eligible for the computation of intersection. This is especially useful for improving efficiency and accuracy in big data analysis, because it will greatly reduce the unnecessary raw data for data processing. The requester needs to compute a hashing set, based on relationships the cloud finds out all sets S_i which satisfy requester conditions.

A. Algorithm for ELGAMAL ENCRYPTION

DA (D, I, k ,m)

1. scan D and create count-tree
2. Initialize C_{out}
3. for each node v in preorder count-tree transversal do
4. if the item of v has been generalized in C_{out} then
5. backtrack
6. if v is a leaf node and v.count < k then
7. J:= item set corresponding to v
8. find generalization of items in J that make J k-anonymous
9. merge generalization rules with C_{out}
10. backtrack to longest prefix of path J, wherein no item has been generalized in C_{out}
11. Return C_{out}
12. for i :=1 to C_{out} do
13. initialize count=0
14. scan each transactions in C_{out}
15. Separate each item in a transaction and store it in p
16. Increment count
17. for j:=1 to count do
18. for all g belongs C_{out} do
19. compare each item of p with that of C_{out}
20. if all items of i equal to C_{out}
21. Increment the r
22. if ka equal to r then backtrack to i
23. else if r greater than ka then get the index position of the similar transactions
24. make them NULL until ka equal to r else update the transactions in database

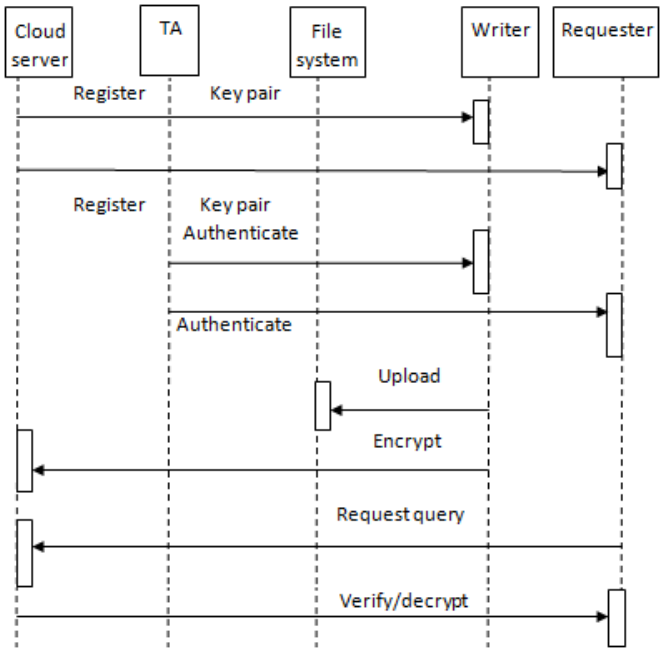


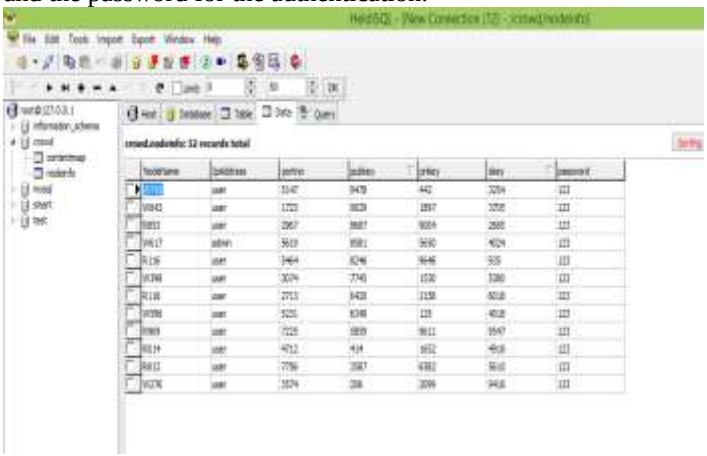
Figure 2: Sequence Diagram for process

5. Performance Evaluation

In this section, we evaluate the computation and observation of the proposed framework.

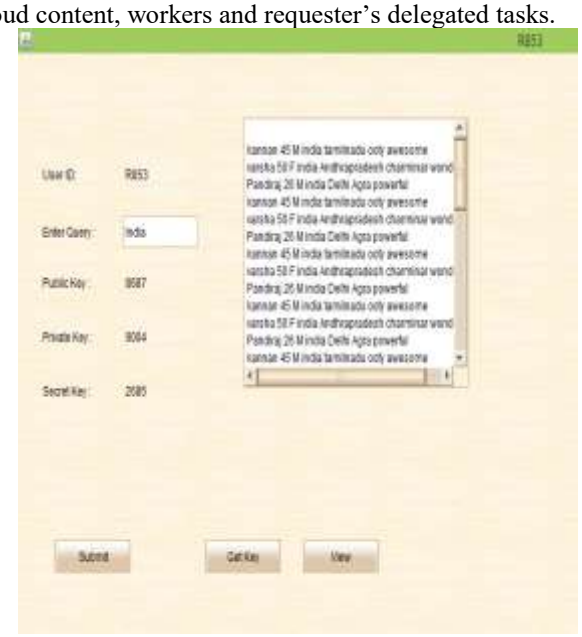
The data collected from mobile crowdsourcing should be well protected. So an efficient solution for the set operation in big data analysis is proposed. Here we are providing worker's data privacy and identity privacy by introducing cloud as an intermediate entity to the traditional mobile crowdsourcing. On the other side the requester can retrieve the information from the cloud and can check for the correctness of the sent data. And further it is implemented for applications such as data pre-processing, batch verification and efficient data update in big data analytics.

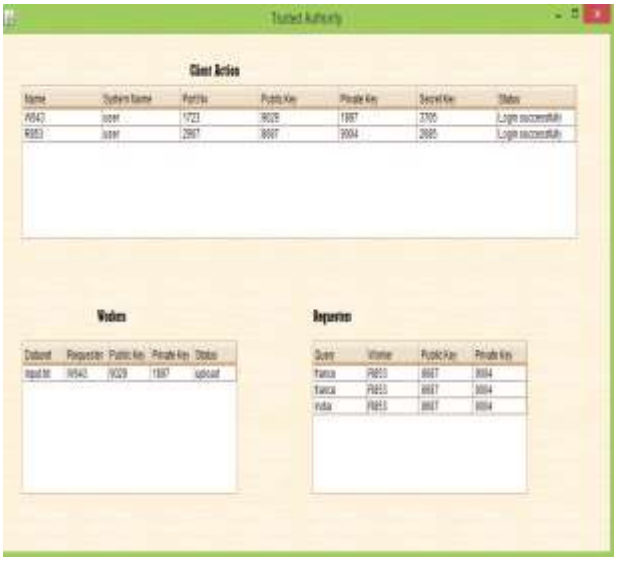
Initially the worker and the requester should get register to the cloud. Once they register they get a node name and the password for the authentication.



HeidiSQL act as a trusted authority that will provide the necessary information such as workers and registers node name, Ip address, public and private key, secret key and password for each of the worker and register during the system initialization. Then each of them should get login by giving their respective password and port number. Based on the requester task the input data set must be given to the worker. This task is done by cloud and is shown in the above result.

Then the register will login to receive the requested data from the worker. The requested data that has been sent from the worker will be in the form of encrypted data sets. Set operation is computed based on the encrypted data sets. Then the cloud sends the private key to the requester. Using that key the requester can decrypt the data and can check for the correctness of the information. Finally the trusted authority gives the complete information about client action, cloud content, workers and requester's delegated tasks.





6. Conclusion

Mobile crowdsourcing is a term that describes crowd sourcing activities that are processed on smart phones. However data collection, data aggregation, and data analysis have become challenging problems for a resource constrained requester when data volume is extremely large, i.e., big data. In particular to data analysis, set operations, including intersection, union, and complementation, exist in most big data analysis for filtering redundant data and preprocessing raw data. Facing challenges in terms of limited computation and storage resources, cloud-assisted approaches may serve as a promising way to tackle big data analysis issue.

In the proposed system, a scheme is proposed to enable the requester to delegate set operations over crowdsourced big data to the cloud. Meanwhile, worker's data and identity privacy are preserved, and the requester can verify the correctness of the set operation result. Additional to that data preprocessing is computed with which invalid data can be excluded before data analysis. Batch verification and data update methods are also proposed so that the computational costs of the system are also reduced.

References

[1] S. S. Kanhere, "Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces," in Mobile Data Management (MDM), 2011 12th IEEE International Conference on, vol. 2. IEEE, 2011, pp.3–6.
[2] Q. Li and G. Cao, "Privacy-preserving participatory sensing."
[3] "Efficient and privacy-preserving data aggregation in mobile sensing," in Network Protocols (ICNP), 2012 20th IEEE International Conference on, Oct 2012, pp. 1–10.
[4] Q. Li, G. Cao, and T. La Porta, "Efficient and privacy-aware data aggregation in mobile sensing," Dependable and Secure Computing, IEEE Transactions on, vol. 11, no. 2, pp. 115–129, March 2014.
[5] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "Anonysense: privacy-aware people-centric sensing," in Proceedings of the 6th international conference on Mobile systems, applications, and services. ACM, 2008, pp. 211–224.
[6] R. Zhang, J. Shi, Y. Zhang, and C. Zhang, "Verifiable privacy-preserving aggregation in people-centric urban sensing systems," Selected Areas in Communications, IEEE Journal on, vol. 31, no. 9, pp. 268–278, September 2013.
[7] S. S. Kanhere, "Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces," in Distributed computing and internet technology. Springer, 2013, pp. 19–26.
[8] H. Yue, L. Guo, R. Li, H. Asaeda, and Y. Fang, "Dataclouds: Enabling community-based data-centric services over the internet of things," IEEE Internet of Things Journal, vol. 1, no. 5, pp. 472–482, Oct 2014.
[9] K. Hara, S. Azenkot, M. Campbell, C. L. Bennett, V. Le, S. Pannella, R. Moore, K. Minckler, R. H. Ng, and J. E. Froehlich, "Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis," ACM Transactions on Accessible Computing (TACCESS), vol. 6, no. 2, p. 5, 2015.
[10] B. Liu, Y. Jiang, F. Sha, and R. Govindan, "Cloud-enabled privacypreserving collaborative learning for mobile sensing," in Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems. ACM, 2012, pp. 57–70.
[11] G. Zhuo, Q. Jia, L. Guo, M. Li, and Y. Fang, "Privacy-preserving verifiable proximity test for location-based services," in 2015 IEEE Global Communications Conference (GLOBECOM). IEEE, 2015, pp. 1–6.
[12] G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li, "Privacy-preserving verifiable data aggregation and analysis for cloud-assisted mobile crowdsourcing," in INFOCOM, 2016 Proceedings IEEE. IEEE, 2016.
[13] H. Li, D. Liu, Y. Dai, and T. H. Luan, "Engineering searchable encryption of mobile cloud networks: when qoe meets qop," Wireless Communications, IEEE, vol. 22, no. 4, pp. 74–80, 2015.
[14] H. Li, Y. Yang, T. Luan, X. Liang, L. Zhou, and X. Shen, "Enabling finegrained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data," IEEE Transactions on Dependable and Secure Computing, vol. PP, no. 99, pp. 1–1, 2015.
[15] X. Chen, X. Wu, X.-Y. Li, Y. He, and Y. Liu, "Privacy-preserving highquality map generation with participatory sensing," in INFOCOM, 2014 Proceedings IEEE. IEEE, 2014, pp. 2310–2318.

Author Profile

Ms. Kumudashree H S is a P.G Scholar in Digital Electronic and Communication System at Malnad College of Engineering, Hassan.

Mrs. Triveni C L is an Assistant Professor in Department of Electronics & Communication at Malnad College of Engineering, Hassan.