DC Research Project

-Ravi Prakash

Paper 1: **Intelligent cryptography approach for secure distributed big data storage in cloud computing**

Problem description

One of major concerns in security and privacy is caused by the fact that cloud operators have chances to reach the sensitive data, which increases users' anxiety and reduces the adoptability of cloud computing in many fields, such as the financial industry and government agencies.

Solution

1. The paper proposes Security-Aware Efficient Distributed Storage(SAEDS).
2. Split those data packets which contain sensitive information based on labels and after encryption store on different cloud servers.
3. For normal data packets above rule doesn't get applied. They are encrypted and stored on single cloud server.
4. To retrieve, combine the fetched and decrypted data packets from different cloud servers.

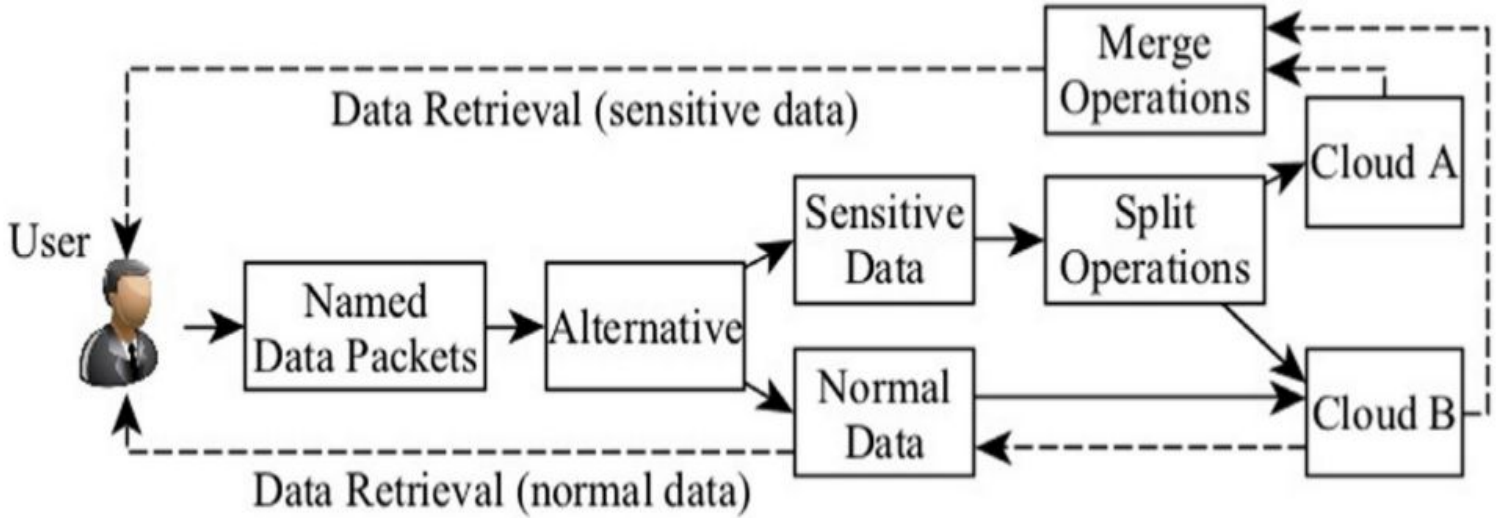

Fig. 1. The architecture of the proposed SA-EDS model.

Analysis

The comparison is between the proposed model and Advanced Encryption Standard(AES) application also between sending and receiving data at different stages on the basis of execution time. It was observed that :

1. the proposed model had a shorter execution time than AES, when EDS and AES were compared.
2. the decryption(EDCon) time required a longer time period when compared to encryption(SED2).
3. the data that needed decryptions were impacted by the data size. The execution time became longer when the data size increased.
4. when EDCon and AES decryption were compared, the execution time length of the proposed approach is slightly shorter than AES.
5. The proposed scheme consumed less computation time than AES, when simulated the data encryption process before the data were sent to cloud-side servers.

Conclusion

The paper focused on the problem of the cloud data storage and aimed to provide an approach that could avoid the cloud operators reaching out users' sensitive data. The paper proposed a model called Security-Aware Efficient Distributed Storage(SA-EDS) model. This model has used two algorithms, including Alternative Data Distribution (AD2), Secure Efficient Data Distributions (SED2) and Efficient Data Conflation(EDCon) algorithms. The proposed model could effectively defend major threats from cloud-side.

Achievements of the proposed model:

1. The achieved computation time was shorter than current approaches.
2. A novel cryptography approach for delivering mass distributed storage so that original data cannot be directly reached by cloud operators.
3. An efficient data split mechanism that does not produce big overheads, as well as ensures data retrievability.

Some things which are worth considering:

1. Data duplication so that data is available without failure.
2. Data centers must be independent so that operator of one data center must not gain access of others.
3. There should be a proper key management system.

Paper 2: **An algorithm for privacy-preserving distributed user statistics**

Problem description

This method has been proposed to determine the total number of distinct user IDs who connected to one or more entry points of a distributed internet service with multiple service operators in a privacy-preserving way.

Solution

1. Use FM sketches to collect data from each service points.
2. Perturbation Technique helps in confusing the attacker by setting the bits in FM sketches.
3. Combining the FM sketches from all service points while preserving privacy.
4. Get statistics based on the combined FM sketch e.g. counting users.

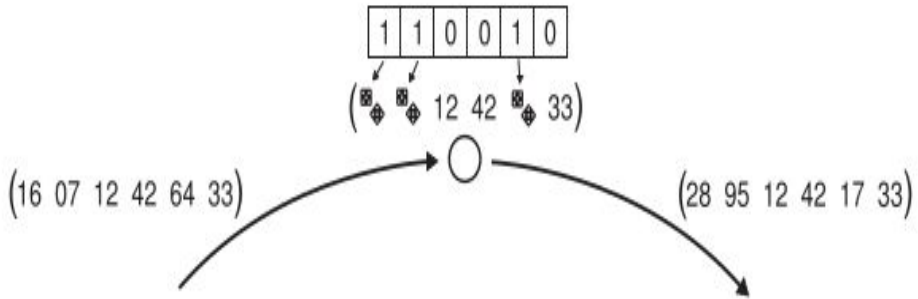

Fig. 6. Combining sketches in a privacy-preserving way without leaking any information in the presence of an honest-but-curious adversary.

Analysis

1. As for original FM sketches, increasing m rows reduces the standard error. For increasing r i.e perturbation standard error increases, but remains at reasonable levels even for relatively high values of r and also for higher m.
2. This algorithm is able to merge data from multiple entry points without sacrificing duplicate insensitivity. It does not estimate the total count based on a limited local view.
3. It also clear that the accuracy of estimation is independent of average number of client requests per day. Tor's current user count estimation based on data from a single entry point is very accurate iff average number of client requests per day is around 10; otherwise, they can be arbitrarily far off.

Hence, the proposed algorithm allows to obtain global, privacy-preserving, and duplicate insensitive user statistics without the need to make assumptions about user behaviour that cannot be checked.

Applications

The proposed method is useful in other areas also. For example:

1. Determining the number of users connecting from specific country.
2. Determining the number of users are searching for specific keyword
3. In case of healthcare analysis, number of patients with certain disease.