International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Prevention of Pollution Attack in Cloud

Shanmugavel. S [1], Hariharan. J [2], Daniel Thomas Abraham[3], Abiram[i]. M [4]
[1, 2, 3]Student IV Year CSE, [4]Assistant Professor Panimalar Institute of Technology

Abstract: ***The security of user's data to cloud is the major impact of all stakeholder. The user, System designer and cloud storage provider has to manage their data. Pollution attacks, whereby a set of malicious entities attempt to corrupt stored data, are one of the many risks that affect cloud data security. Those cloud storage provider should not be hacked and it should be safe to maintain their data. In our proposed technique we were implementing how to prevent pollution attack and how to store our file to be safe. The following techniques and algorithm are implemented in our projects. The first technique is to the user send a file to admin with encrypted file by encryption we are using RC5 encryption algorithm. The user only able to upload the file format (.jpg, .txt). The technique is Checksum Method, a user send an encrypted file to admin by using this method it generate a key for the particular file. The encrypted file has to verified by the admin. By this verification the checksum method generates a key to the file if the file has same key means the file not hacked or not attached. After verification the admin has to upload the file to cloud. The file should be stored safely If the user need a file means send a request to the admin after getting the response key from admin the user able to download the file. Our Approach is to very robust and highly isolated to polluters and also to improve performance and computational cost.***

Keywords: Pollution attacks, Cloud storage, coding, security, integrity, performance.

I. INTRODUCTION

Cloud computing is an information technology (IT) paradigm, a model for enabling ubiquitous access to shared pools of configurable resources (such as computer networks, servers, storage, applications and services), which can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing allows users and enterprises with various computing capabilities to store and process data either in a privately-owned cloud, or on a third-party server located in a data center - thus making data-accessing mechanisms more efficient and reliable. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility. Advocates note that cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. As well, third-party clouds enable organizations to focus on their core businesses instead of expending resources on computer infrastructure and maintenance. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and that it enables IT teams to more rapidly adjust resources to meet fluctuating and unpredictable business demand. Cloud providers typically use a "pay-as-you-go" model. This could lead to unexpectedly high charges if administrators are not familiarized with cloud-pricing models In 2009 the availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing led to a growth in cloud computing. Companies can scale up as computing needs increase and then scale down again when demands decrease. In 2013 it was reported that cloud computing had become a highly demanded service or utility due to the advantages of high computing power, cheap cost of services, high performance, scalability, and accessibility - as well as availability. Some cloud vendors experience growth rates of 50% per year, but while cloud computing remains in a stage of infancy, it has pitfalls that need to be addressed to make cloud-computing services more reliable and user-friendly.

II. EXISTING SYSTEM

As Network coding allows intermediate (possibly malicious) nodes to actively mix packets it introduces new challenges in the detection of corrupted (or polluted) packets, it. This is vulnerable and widely used to corrupt the data. Unfortunately, the very nature of packet mixing makes network coding systems vulnerable to a severe security threat known as pollution attacks, in which attackers inject corrupted packets into the network. The server also gets infected or victim for the pollution attack and some authorities (i.e., coercers) may force cloud storage providers to reveal user secrets or confidential data on the cloud, thus will lead to the breach of user data integrity and confidentiality. The security of user's data to cloud is the major impact of all stakeholder. The user, System designer and cloud storage provider has to manage their data. Pollution attacks, whereby a set of malicious entities attempt to corrupt stored data, are one of the many risks that affect cloud data security.

A. Disadvantage

There is a malicious entities has to corrupt the stored file or data. This is the major risk factor that affect the data

III. PROPOSED SYSTEM

To provide solutions for the above pollution attack and user data breach, we propose an early pollution detection algorithm able to spot the presence of an attack while fetching the data from cloud storage during the normal disk reading operations. Thus each fragment is check for malicious content and passed over to the server for processing. The alarm triggers a procedure that locates the polluting nodes in early stage of processing to avoid the intrusion. Also to detect the polluted content hashing methods are used. In this way, after network peers receive the transmitted data, hash value is calculated by hash functions and then the result is compared with the received hash value from the content distribution network (CDN) servers to determine whether the transmitted data is polluted or not. Our proposed system also addresses in preserving user data confidentiality and integrity after outsourcing. For this concern, we present our design for a new cloud storage encryption scheme that enables cloud storage providers to create convincing fake user secrets to protect user privacy. Since coercers cannot tell if obtained secrets are true or not, the cloud storage providers ensure that user privacy is still securely protected. For encryption we have proposed RC5 encryption algorithm.

A. Advantages

1) Should have enough data redundancy.
2) Very Robust approach.
3) Able to affectively isolate the polluters.
4) Efficient to store data with highly secured.

IV. PROPOSED ARCHITECTURE

A. Pollution Detection Algorithm

A pollution detection algorithm detects, with high probability if a set of untrusted storage resources provides at least one polluted coded fragment. The algorithm is based on a modified version of the LT decoding algorithm exploiting Gaussian Elimination; since an analytical model for decoding (and detection) performance is unavailable in the literature we resort to simulations to estimate the detection probability.

B. Identification Algorithm

An identification algorithm that identifies the storage resources that are polluters with high probability. The algorithm we propose is not based on cryptographic checksums or digital signatures (hence it does not rely on the existence of a PKI or preestablished secure channels) and it only exploits coding redundancy and efficient decoding algorithms that require the solution of systems of linear equations.

C. Bp Core Algorithm

Belief propagation, also known as sum-product message passing, is a message-passing algorithm for performing inference on graphical models, such as Bayesian networks and Markov random fields. It calculates the marginal distribution for each unobserved node, conditional on any observed nodes. Belief propagation is commonly used in artificial intelligence and information theory and has demonstrated empirical success in numerous applications including low-density parity-check codes, turbo codes, free energy approximation, and satisfiability. The algorithm was first proposed by Judea Pearl in 1982, who formulated this algorithm on trees, and was later extended to poly trees. It has since been shown to be a useful approximate algorithm on general graphs If $X=\{X_i\}$ is a set of discrete random variables with a joint mass function p, the marginal distribution of a single X_i is simply the summation of p over all other variables: However, this quickly becomes computationally prohibitive: if there are 100 binary variables, then one needs to sum over $2^{99} \approx 6.338 \times 10^{29}$ possible values. By exploiting the polytree structure, belief propagation allows the marginals to be computed much more efficiently. Polluter identification can be cast as a statistical inference problem as follows. The main idea is to characterize each SNs i 2 AS by an unknown (hidden) binary state i, where i = 1 is used to identify a polluter and i = 0 is used to identify an honest SN. The goal is then to infer 8i 2 AS, p(i = 1).

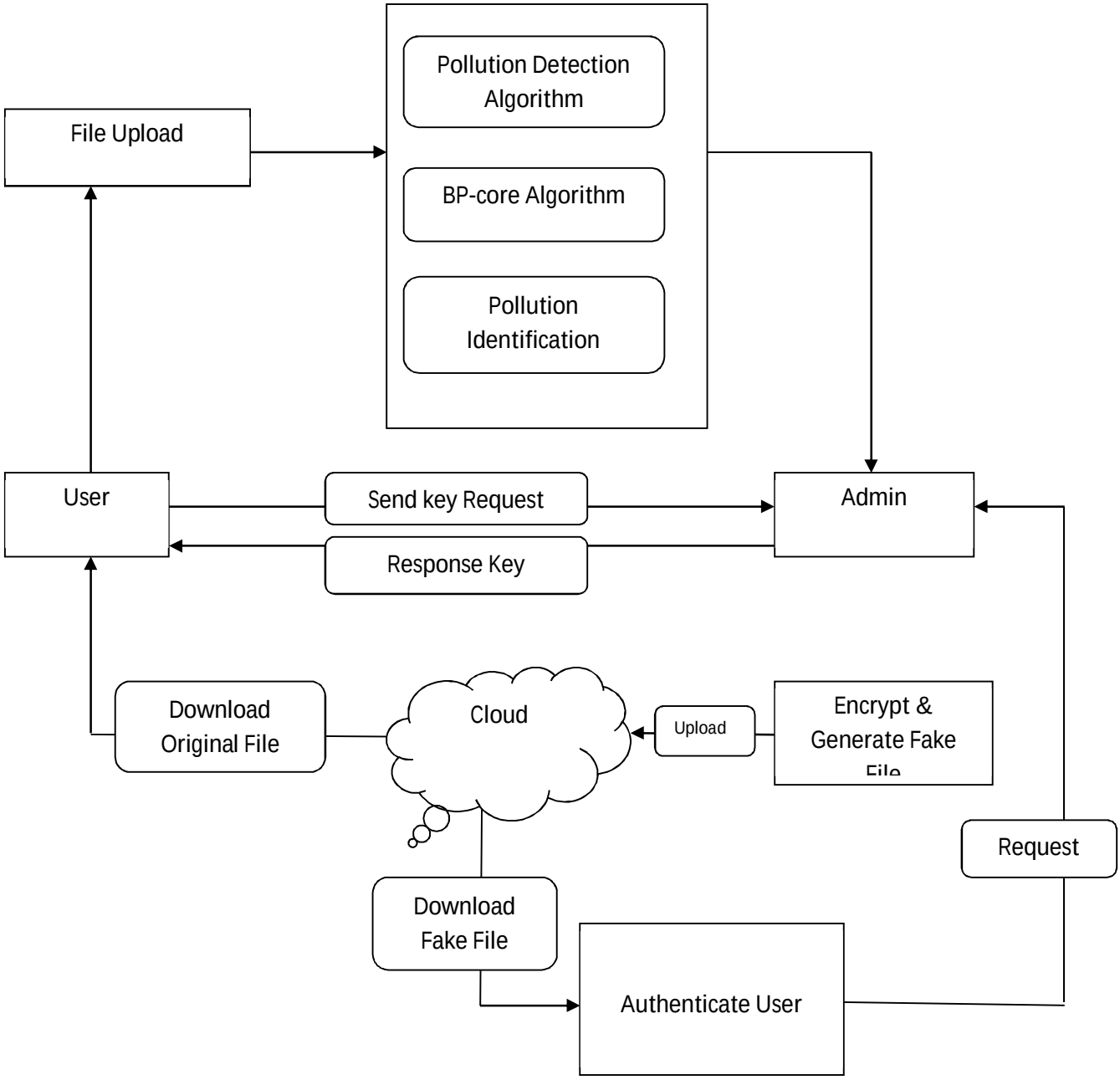

Fig 1: Pollution Attack Architecture Diagram

D. Source Code

Belief propagation(W,H)

1) for all a 2 W d
2) P(_a) =
3) if a 2 H the
4) p(_a = 1) = 0
5) else
6) p(_a = 1) = 0.5
7) end if
8) end fo
9) for i = 1 to BPt d

10) G Build random factor graph(W
11) {p(_a)} BP inference(G)
12) for all a 2 W d
13) P(_a) = P(_a) + p(_a = 1
14) end fo
15) end for
16) for all a 2 W d
17) P(_a) = P(_a)/BPt
18) end for
19) return {P(_a)}

V. CONCLUSION

In this project, pollution detection mechanism is used to check data integrity during the normal read operations of a cloud-based storage system. Also the proposed system identify the malicious nodes. Integrating hash values during transmission and then the result is compared with the received hash value from the content distribution network (CDN) servers to determine whether the transmitted data is polluted or not provides high security. Creation of convincing fake user secrets to protect user privacy.

REFERENCES

[1] H. Dewan and R. Hansdah, "A survey of cloud storage facilities," in IEEE SERVICES, jul 2011, pp. 224 –231.
[2] C. Anglano, R. Gaeta, and M. Grangetto, "Exploiting rateless codes in cloud storage systems," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1313–1322, May 2015.
[3] L. Buttyan, L. Czap, and I. Vajda, "Detection and recovery from pollution attacks in coding-based distributed storage schemes," IEEE Transactions on Dependable and Secure Computing, vol. 8, no. 6, pp. 824–838, 2011.
[4] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT codes-based secure and reliable cloud storage service," in IEEE INFOCOM, 2012, pp. 693–701.
[5] L. Buttyn, L. Czap, and I. Vajda, "Pollution attack defense for coding based sensor storage," in Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC), 2010.
[6] M. N. Krohn, M. J. Freedman, and D. Mazieres, "On-the-fly verification of rateless erasure codes for efficient content distribution," Security and Privacy, IEEE Symposium on, 2004.
[7] C. Gkantsidis and P. Rodriguez, "Cooperative security for network coding file distribution," in IEEE INFOCOM, 2006.
[8] Z. Yu, Y. Wei, B. Ramkumar, and Y. Guan, "An efficient signature based scheme for securing network coding against pollution attacks," in INFOCOM 2008. The 27th Conference on Computer Communications. IEEE, 2008.
[9] E. Kehdi and B. Li, "Null keys: Limiting malicious attacks via null space properties of network coding," in INFOCOM 2009, IEEE.
[10] Z. Yu, Y. Wei, B. Ramkumar, and Y. Guan, "An efficient scheme for securing xor network coding against pollution attacks," in INFOCOM 2009, IEEE.