Detection of Data Lineage with 1-out-of-n way

Sonali Vijay Patil[1]
Department of Computer Engineering, Shree Ramchandra CoE, Lonikand,Pune
sonaliptl663@gmail.com

B. H. Thombre[2]
Department of Computer Engineering, Shree Ramchandra CoE, Lonikand,Pune
babanthombre@gmail.com

Abstract - Cloud computing provides the hosted service through the internet. The authorize or unauthorized leakage of secret data is no doubt one of the most major security problems which organizations or systems face in this era. It also affects our personal day to day life: The personal information is available on social networks, or now-a-days it is also available on Smartphone is intentionally or unintentionally transferred to third party or hackers. Also a data distributor may give confidential data to some trusted agents or third parties. During this process some data is leaked or transferred to unauthorized place. We propose data allocation strategies that will give more probability of identifying leakages. I present LIME data lineage framework for data flow across various locations.

By using oblivious transfer, robust watermarking, and signature primitives we develop and analyze the data transfer protocol in a malicious environment between two entities. We are using the cloud computing concepts to do the high performance based computing which is sometimes used by military and research organization. At the end of I perform an experimental result and analysis of our framework.

Keywords: Lime; Watermarking; Data leakage; Oblivious transfer.

1. Introduction

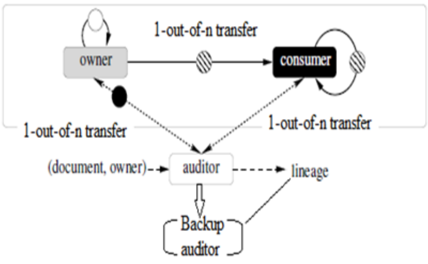
In this 21st century by many ways the information which keeps as a secret may get leaked through unintended vulnerability, some aggressive or dissatisfied employees and external spiteful entities. It also affects our personal day to day life. The Privacy Rights Clearinghouse in the California of United States maintains the chronology of data braches. They found that from 4355 data braches near about 868,045,823 records are breached which made public since 2005[1]. So the loss of data will cause loss to the organization or most of the companies. Also the organization have fear of losing customers confidence, their support or maybe they have to pay fine for data loss. For all these reasons, data leakage becomes headache to organization. Also the management of data is become crucial. In the metadata we can again and again use the same data for our purpose. Now I am using data provenance, which is like metadata concern to data product from original locations. Data leakage is defined as an unofficial, unauthorized transfer of data, information from computer to the outside world and data lineage is defined as chain of data which includes origin of data and where it goes over time. Data lineages do the survey of how this information is used and also track the data system. Data lineage gives the data source and intermediate data flow hops with backward data lineage which finally gives intermediate data flow hops with forward data lineage.

I work on the data transfer protocol in malicious environment between multiple entities and tabulating the problem identification of leakages. The life cycle of data from the origin of data and where it will transfer over time known as data lineage. It describes presentation on the data process with various changes from sender to receiver or from source to destination in the enterprise environment. Data lineage will show about what happens to data when it goes through various steps.

2. Related Work

I want a general case for data leakage in data transfer model, i propose detection of data lineage with 1- out –of N-way by using LIME model.

The LIME framework States that the documents are transferred from owner to consumer. Also the sending owner trusts the receiving owner to take responsibility if he should leak the document and this process is governed by a non repudiation assumption. The sender embeds the information while transferring document to consumer known as fingerprinting. Therefore if the document is leaked by consumer i can easily identify him with the help of the embedded information. The term honesty replies that the trust between the auditor and the owner to be honest. That means the owner does not leak a document and blame another party.



▨ : Fingerprinting ○ : Optional Non repudiation assumption ● : Optional Honesty Assumption

⟶ : Document Flow - - - -▶ : Auditing process ◀- - - -▶ : Communication.

Fig. 1: Proposed Architcture Of Detection Of Data Lineage With 1-Out-Of-N Way

In the LIME framework the data transfer takes place between owner and consumer. In our proposed model the data transfer takes place between owner and multiple receivers. Also i use the backup auditor for generating backup for owner. In case of transfer of data, if data crashed or hacked by someone then backup data from backup auditor transferred to consumer. The advantage of our model is that at the design stage if any data leakage happened then corresponding accountability constraints are considered by system designer. It will help to overcome the existing methodologies where most lineage methods are applicable only after a leakage has happened.

The auditor is not only involved in the transfer, but also he takes appropriate countermeasures when leakage occurs. Auditor is invoked by an owner and provided with leaked data. If in case the leaked data was transferred during this process, the identifying information embedded for each consumer who receives the leaked data. By this embedded information the auditor can create an ordered chain of consumers who received the document. This chain may known as lineage of leaked document. The leaker in this chain is the last consumer.

3. Implementation Process

3.1 Data Lineage Detection Process:

3.1.1 Data Lineage Generation:

The auditor is the entity that is used to find the guilty party in case of a leakage. He is invoked by the owner of the document and is provided with the leaked document. In order to find the guilty party, the auditor proceeds in the following way:

(1) The auditor initially takes the owner as the current suspect.

(2) The auditor appends the current suspect to the lineage.

(3) The auditor sends the leaked document to the current suspect and asks him to provide the detection keys k_1 and k_2 for the watermarks in this document as well as the watermark σ. If a non-blind watermarking scheme is used, the auditor additionally requests the unmarked version of the document.

(4) If, with key k_1, σ cannot be detected, the auditor continues with 9.

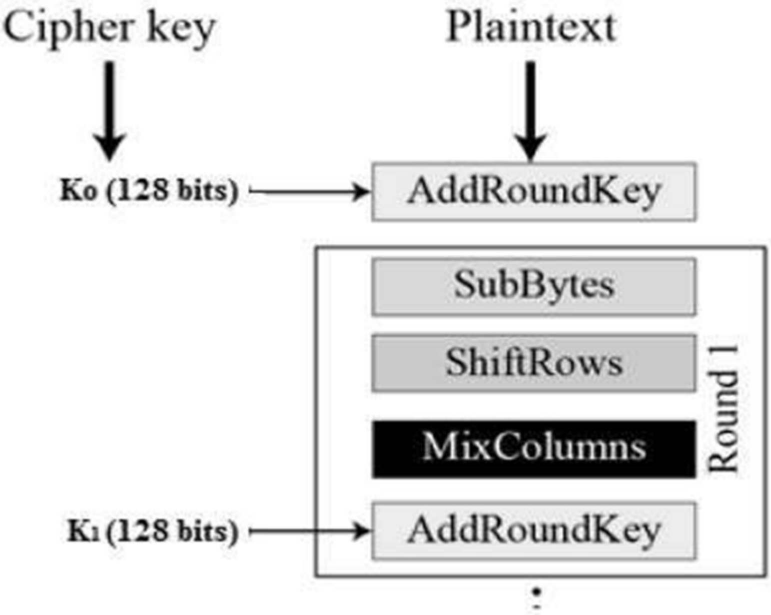
(5) If the current suspect is trusted, the auditor checks that σ is of the form (C_S,C_R, τ) where C_S is the identifier of the current suspect, takes C_R as current suspect and continues with 2.

(6) The auditor verifies that _ is of the form $[C_S, C_R, \tau]_{skcR}$ Where C_S is the identifier of the current suspect. He also verifies the validity of the signature.

(7) The auditor splits the document into n parts and for each part he tries to detect 0 and 1 with key k_2. If none of these or both of these are detectable, he continues with 9. Otherwise he sets b_i' as the detected bit for the ith part. He sets $b'^{-} = b_1' \ldots b_n'$.

(8) The auditor asks C_R to prove his choice of $b^{-} = b_1 \cdots b_n$ for the given timestamp τ by presenting the m_i,b_i $=[\tau, i, b_i]_{skC_S}$. If C_R is not able to give a correct proof (i.e., m_i,b_i is of the wrong form or the signature is invalid) or if $b^{-} = b'$, then the auditor takes C_R as current suspect and continues with 2.

(9) The auditor outputs the lineage. The last entry is responsible for the leakage.



- User has to select User U to send document.
- Watermarked that document with Keys K
- k ←GenKeyWM(1k)
- σ= (CS,CR, Γ)
- Dw= W(D, σ, k)
- AES Algorithm for encryption and decryption

Interestingly, AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix.Unlike DES, the number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key.

3.2 1- Out –of –N Oblivious Transfer:

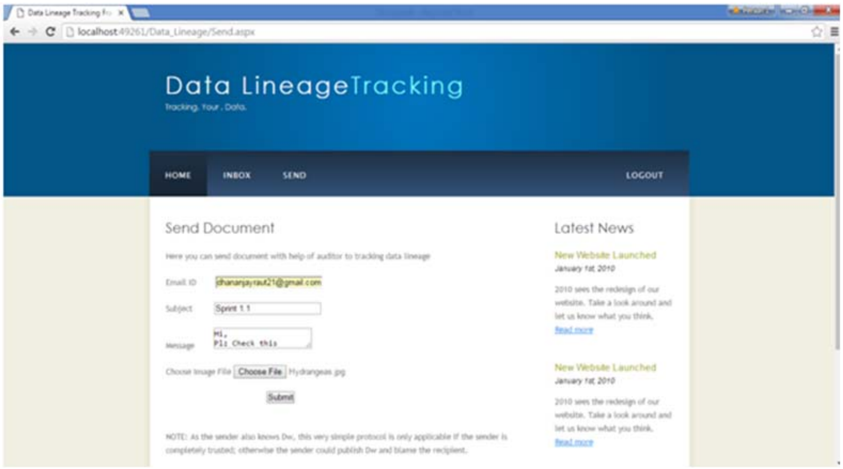
- Alice randomly picks n secrets $s_{1,}$ …..s_n and calculate t_i as:

$$\varphi_i \in \{1,\ldots n\} : t_i = s1,\ldots\ s_i\text{-}1 \in M_i$$

- For each I ∈ {1,…n}, Alice and John are engaged in a 1- Out –of OT where John's first message will be t_i and the second message is $s_{i.}$ Alice picks t_i to receive if she wants s_i and M_i otherwise.
- After Alice receives N components,she has $t_i = s1,\ldots\ s_i\text{-}1 \in M_i$ for the i she wants and s_k for k≠ i, she can recover the M_i by

$$M_{i\,=}\,t_i\ \in s_i\text{-}1 \in s_i\text{-}1\ \in s_i\text{-}1 \in \ldots\ s_1$$

4. Experimental Result



Send Document

Here you can send document with help of auditor to tracking data lineage

Email ID dhananjayraut21@gmail.com

Subject Sprint 1.1

Message Hi,
Pl: Check this

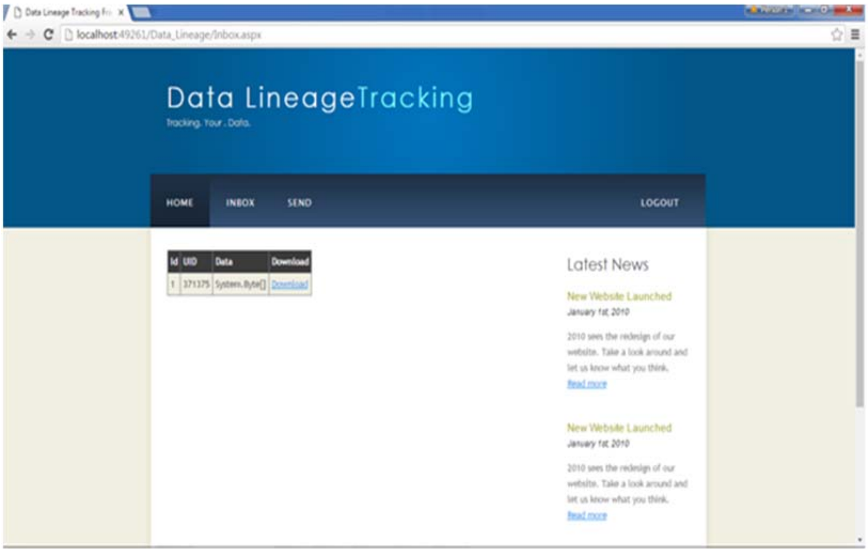
Choose Image File Hydrangeas.jpg

NOTE: As the sender also knows Dw, this very simple protocol is only applicable if the sender is completely trusted; otherwise the sender could publish Dw and blame the recipient.

Latest News

New Website Launched
January 1st, 2010
2010 sees the redesign of our website. Take a look around and let us know what you think.

New Website Launched
January 1st, 2010
2010 sees the redesign of our website. Take a look around and let us know what you think.



Data Lineage Tracking Fr...

localhost:49261/Data_Lineage/Inbox.aspx

Data LineageTracking

Tracking. Your . Data.

HOME INBOX SEND LOGOUT

Id	UID	Data	Download
1	371375	System.Byte[]	Download

Latest News

New Website Launched

January 1st 2010

2010 sees the redesign of our website. Take a look around and let us know what you think.

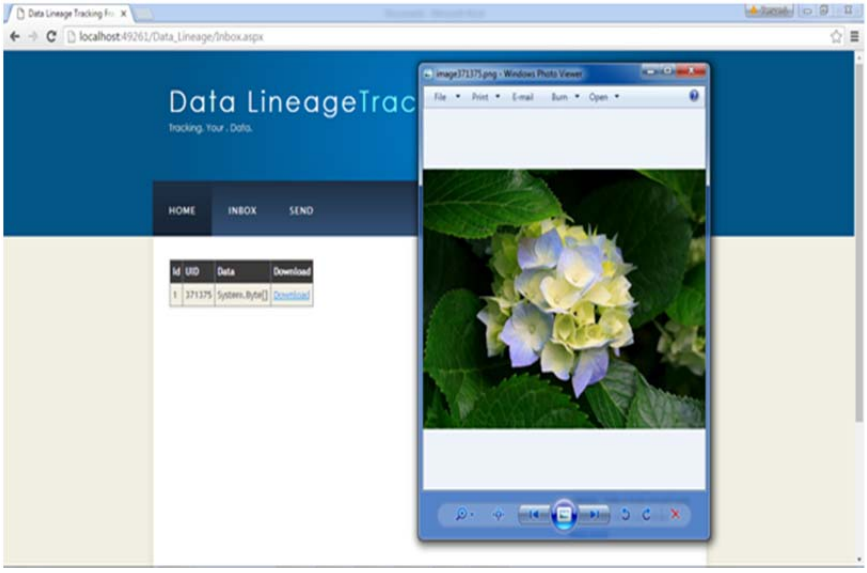
Read more

New Website Launched

January 1st 2010

2010 sees the redesign of our website. Take a look around and let us know what you think.

Read more



Data Lineage Tracking Fr...

localhost:49261/Data_Lineage/Inbox.aspx

Data LineageTrac

Tracking. Your . Data.

HOME INBOX SEND

Id	UID	Data	Download
1	371375	System.Byte[]	Download

image371375.png - Windows Photo Viewer

5 Performance analyses

I can measure the execution time for watermarking, encryption, signature creation, detection and oblivious transfer.

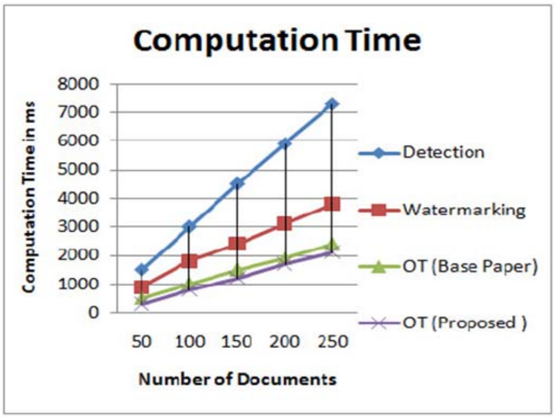


Fig. 2 Computation time for different documents

6 Conclusion

In spite of the fact that LIME system does not effectively avert information spillage, it presents receptive responsibility. Hence it will prevent vindictive gathering from releasing private archives and will energize genuine parties give the obliged insurance to touch information. LIME is adaptable as i separate between trusted senders and untrusted senders. On account of the trusted senders, an exceptionally basic convention with minimal overhead is conceivable. The untrusted senders require more confounded protocol but the outcomes are not in view of trust suppositions and in this way they ought to have the capacity to persuade a neutral entity. Our work spurs further research on information spillage location methods for different document types and situations.

Acknowledgement

I am profoundly grateful to Prof. B. H. Thombre for his guidance and continuous encouragement throughout to see that this paper rights its target since its commencement to its completion. I would like to express deepest appreciation towards his invaluable guidance and support me in completing this paper. Also I must express my sincere gratitude to all the staff members of Computer Department who helped me directly or indirectly during this course of work.

References

[1] http://www.computerworld.com/s/article/109938/Offshore outsourcing cited in Florida data leak.

[2] A. Mascher-Kampfer, H. St ¨ogner, and A. Uhl, "Multiplere-watermarking scenarios, in Proceedings of the 13th International Conference on Systems, Signals, and Image Processing (IWSSIP 2006).Citeseer, 2006, pp. 53–56.

[3] P. Papadimitriou and H.Garcia-Molina "Data leakage detection, Knowledge and Data Engineering, IEEE Transactions on, vol. 23, no. 1, pp. 51–63, 2011.

[4] Pairing-Based cryptographyLibrary(PBC),http://crypto.stanford.edu/pbc.

[5] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia, Image Processing, IEEE Transactions on, vol. 6, no.12, pp. 1673–1687, 1997.

[6] Bhamare Ghanashyam,Desai Kiran,Khatal Supriya, Mane Vinod,Prof. Hirave K.S.," A Survey Paper on Data Lineage in Malicious Environments" Multidisciplinary Journal of Research in Engineering and Technology, Volume 2, Issue 4,Pg.720-724

[7] "Chronology of data breaches,"http://www.privacyrights.org/data-breach.

[8] "Data breach cost," http://www.symantec.com /about/ news/release /article .jsp?prid=20110308 01.

[9] "Privacy rights clearinghouse," http://www.privacyrights.org.

[10] Michael Backes, Niklas Grimm, and Aniket Kate," Data Lineage in Malicious Environments" DOI 10.1109/TDSC.2015.2399296, IEEE Transactions on Dependable and Secure Computing

[11] M. Handley and J. Crowcroft. Network Text Editor (NTE): A scalable shared text editor for the MBone. In ACM SIGCOMM, pages 197–208, Cannes, France, 1997.

[12] F.Hartung and B. Girod. Fast public-key watermarking of compressed video. In IEEE International Conference on Image Processing, pages 528– 531, Santa Barbara, USA, 1997.