International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

A Novel Accountable Data Transfer Protocol within A Malicious Environment

P.Indupriyanka Ray & P.Gangadhara
[1]M.Tech, Dept of CSE, Shri Shirdi Sai Institute of Science and Engineering, Affiliated to JNTUA, AP, India .
pindupriyankaray@gmail.com
[2]AssistantProfessor, Dept of CSE, Shri Shirdi Sai Institute of Science and Engineering, Affiliated to JNTUA, AP, India.
gangadhara115208@gmail.com

Abstract

In today's era, information leakage is one of the most serious threats to companies. A data owner sends secret or confidential information to a group of trusted agents. Some ofthe information is lost and found in an inappropriate place. Thus data is leaked. Data leakage means data distributed by the data owner is leaked by one or more agents. This causes a huge harm to the business. The distributor must assess whether data is leaked from one or more agents. To improve the probability of identifying leakages data allocation strategies (across the agents) are used. A data lineage framework is used for identifying a guilty entity. The digital watermarking is a technique in which vital information is kept hidden in the original data for protecting unauthorized duplication and distribution of data. An accountable data transfer protocol is built using oblivious transfer, robust watermarking, and signature primitives. In some occasions fake data records are injected in order to improve detecting data loss and identifying the guilty entity. The data sent by the data owner must be protected, secret and it must not be regenerated. The framework of data lineage is considered for transmission of data and is a key step towards achieving accountability.

I. INTRODUCTION

In the digital era, information leakage through unintentional exposures, or intentional sabotage by disgruntled employees and malicious external entities, present one of the most serious threats to organizations. Not only companies are affected by data leakage, it is also a concern to individuals. The rise of social networks and smart phones has made the situation worse. In these environments, individuals disclose their personal information tovarious service providers, commonly known as third party applications, in return for some possibly free services. In the absence of proper regulations and accountability mechanisms, many of these applications share individuals' identifying information with dozens of advertising and Internet tracking companies. Even with access control mechanisms, where access to sensitive data is limited, a malicious authorized user can publish sensitive data as soon as he receives it. Primitives like encryption offer protection only as long as the information of interest is encrypted, but once the recipient decrypts a message, nothing can prevent him from publishing the decrypted content. Thus it seems impossible to prevent data leakage proactively. Privacy, consumer rights, and advocacy organizations try to address the problem of information leakages through policies and awareness. Data Leakage is an important concern for the business organizations in this increasingly networked world these days. Illegitimate disclosure may have serious consequences for an organization in both long term and short term. Risks include losing clients and stakeholder confidence, tarnishing of brand image, landing in undesirable lawsuits, and overall losing goodwill and market share in the industry. To prevent from all these unwanted and nasty activities from happening, an organized effort is needed to control the information flow inside and outside the organization. Here is our attempt to demystify the jargon surrounding the data leakage prevention procedures which will help you to choose and apply the best suitable option for your own business. Leakage describes an unwanted loss of something which escapes from its proper location and Lineage describes as data flow across multiple entities that take two characteristic, principal roles (i.e., owner

and consumer). We define the exact security guarantees required by such a data lineage mechanism toward identification of a guilty entity, and identify the simplifying nonrepudiation and honesty assumptions. In the course of doing business, sometimes sensitive data must be handed over to supposedly trusted third parties. For example, a hospital may give patient records to researchers who will devise new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various other companies. The owner of the data can be called as distributor and the supposedly trusted third parties the agents. The goal is to detect when the distributors sensitive data have been leaked by agents, and if possible to identify the agent that crevice the data.

II. RELATED WORK

1) Multiple re-watermarking scenarios

AUTHORS: A. Mascher-Kampfer, H. St€ogner, and A. Uhl

The use of classical robust watermarking techniques for multiple re-watermarking is discussed. In particular we focus on a comparison of the usefulness of blind and non-blind algorithms for this type of applications. A surprisingly high number of watermarks may be embedded using both approaches, provided that additional data is recorded in the non-blind case.

2) Data leakage detection

AUTHORS: P. Papadimitriou and H. Garcia-Molina

We study the following problem: A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some of the data are leaked and found in an unauthorized place (e.g., on the web or somebody's laptop). The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. We propose data allocation strategies (across the agents) that improve the probability of identifying leakages. These methods do not rely on alterations of the released data (e.g., watermarks). In some cases, we can also inject "realistic but fake" data records to further improve our chances of detecting leakage and identifying the guilty party.

3) Secure spread spectrum watermarking for multimedia

AUTHORS: I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon this paper presents a secure (tamper-resistant) algorithm for watermarking images, and a methodology for digital watermarking that may be generalized to audio, video, and multimedia data. We advocate that a watermark should be constructed as an independent and identically distributed (i.i.d.) Gaussian random vector that is imperceptibly inserted in a spread-spectrum-like fashion into the perceptually most significant spectral components of the data. We argue that insertion of a watermark under this regime makes the watermark robust to signal processing operations (such as lossy compression, filtering, digital-analog and analog-digital conversion, requantization, etc.), and common geometric transformations (such as cropping, scaling, translation, and rotation) provided that the original image is available and that it can be successfully registered against the transformed watermarked image. In these cases, the watermark detector unambiguously identifies the owner. Further, the use of Gaussian noise, ensures strong resilience to multiple-document, or collusional, attacks. Experimental results are provided to support these claims, along with an exposition of pending open problems

4) Asymmetric fingerprinting for larger collusions,

AUTHORS: B. Pfitzmann and M. Waidner

Fingerprinting schemes deter people from illegal copying of digital data by enabling the merchant of the data to identify the original buyer of a copy that was redistributed illegally. All known fingerprinting schemes are symmetric in the following sense: Both the buyer and the merchant know the fingerprinted copy. Thus, when the merchant finds this copy somewhere, there is no proof that it was the buyer who put it there, and not the merchant.

We introduce asymmetric fingerprinting. where only the buyer knows the fingerprinted copy, and the merchant, upon finding it somewhere, can find out and

prove to third parties whose copy it was. We present a detailed definition of this concept and constructions.

5) A digital signature scheme secure against adaptive chosen-message attacks

AUTHORS: S. Goldwasser, S. Micali, and R. L. Rivest We present a digital signature scheme based on the computational difficulty of integer factorization. The scheme possesses the novel property of being robust against an adaptive chosen-message attack: an adversary who receives signatures for messages of his choice (where each message may be chosen in a way that depends on the signatures of previously chosen messages) cannot later forge the signature of even a single additional message. This may be somewhat surprising, since in the folklore the properties of having forgery being equivalent to factoring and being invulnerable to an adaptive chosen-message attack were considered to be contradictory. More generally, we show how to construct a signature scheme with such properties based on the existence of a "claw-free" pair of permutations--a potentially weaker assumption than the intractibility of integer factorization. The new scheme is potentially practical: signing and verifying signatures are reasonably fast, and signatures are compact.

III.EXISTING SYSTEM

- The data provenance methodology, in the form of robust watermarking techniques or adding fake data, has already been suggested in the literature and employed by some industries.
- Hasan et al. present a system that enforces logging of read and write actions in a tamper-proof provenance chain. This creates the possibility of verifying the origin of information in a document.
- Poh addresses the problem of accountable data transfer with untrusted senders using the term fair content tracing. He presents a general framework to compare different approaches and splits protocols into four categories depending on their utilization of trusted third parties, i.e., no trusted third parties, offline trusted third parties, online trusted third parties and trusted hardware. Furthermore, he introduces the additional properties of recipient anonymity and fairness in association with payment.

DISADVANTAGES OF EXISTING SYSTEM:

- In some cases, identification of the leaker is made possible by forensic techniques, but these are usually expensive and do not always generate the desired results.
- Most efforts have been ad-hoc in nature and there is no formal model available.
- Additionally, most of these approaches only allow identification of the leaker in a non-provable manner, which is not sufficient in many cases.
- An attacker is able to strip of the provenance information of a file, the problem of data leakage in malicious environments is not tackled by their approach.

PROPOSED SYSTEM:

- We point out the need for a general accountability mechanism in data transfers. This accountability can be directly associated with provably detecting a transmission history of data across multiple entities starting from its origin. This is known as data provenance, data lineage or source tracing.
- In this paper, we formalize this problem of provably associating the guilty party to the leakages, and work on the data lineage methodologies to solve the problem of information leakage in various leakage scenarios.
- This system defines LIME, a generic data lineage framework for data flow across multiple entities in the malicious environment.
- We observe that entities in data flows assume one of two roles: owner or consumer. We introduce an additional role in the form of auditor, whose task is to determine a guilty party for any data leak, and define the exact properties for communication between these roles.
- In the process, we identify an optional non-repudiation assumption made between two owners, and an optional trust (honesty) assumption made by the auditor about the owners.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

- As our second contribution, we present an accountable data transfer protocol to verifiably transfer data between two entities. To deal with an untrusted sender and an untrusted receiver scenario associated with data transfer between two consumers, our protocols employ an interesting combination of the robust watermarking, oblivious transfer, and signature primitives.

ADVANTAGES OF PROPOSED SYSTEM:

- The key advantage of our model is that it enforces accountability by design; i.e., it drives the system designer to consider possible data leakages and the corresponding accountability constraints at the design stage. This helps to overcome the existing situation where most lineage mechanisms are applied only after a leakage has happened.
- We prove its correctness and show that it is realizable by giving micro benchmarking results. By presenting a general applicable framework, we introduce accountability as early as in the design phase of a data transfer infrastructure.

SYSTEM ARCHITECTURE

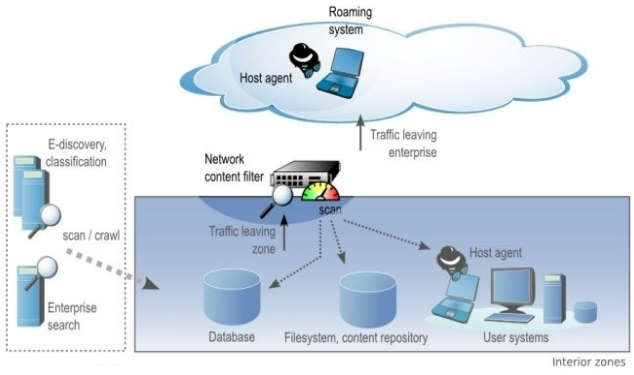

Fig 1: system Architecture

IMPLEMENTATION

LIME System Model

- In the first module, we develop the LIME System Model, which consists of system entities data owner, data consumer and auditor. There are three different roles that can be assigned to the involved parties in LIME: data owner, data consumer and auditor.
- The data owner is responsible for the management of documents and the consumer receives documents and can carry out some task using them.
- The auditor is not involved in the transfer of documents, he is only invoked when a leakage occurs and then performs all steps that are necessary to identify the leaker.
- All of the mentioned roles can have multiple instantiations when our model is applied to a concrete setting. We refer to a concrete instantiation of our model as scenario.
- When documents are transferred from one owner to another one, we can assume that the transfer is governed by a non-repudiation assumption. This means that the sending owner trusts the receiving owner to take responsibility if he should leak the document. As we consider consumers as untrusted participants in our model, a transfer involving a consumer cannot be based on a non-repudiation assumption. Therefore, whenever a document is transferred to a consumer, the sender embeds information that uniquely identifies the recipient. We call this fingerprinting. If the consumer leaks this document, it is possible to identify him with the help of the embedded information.

Attackers Module

- In this module, we develop attackers in our model as consumers that take every possible step to publish a document without being held accountable for their actions. As the owner does not trust the consumer, he uses fingerprinting every time he passes a document to a consumer. However, we assume that the consumer tries to remove this identifying information in order to be able to publish the document safely.
- As already mentioned previously, consumers might transfer a document to another consumer, so we also have to consider the case of an untrusted sender. This is problematic because a

sending consumer who embeds an identifier and sends the marked version to the receiving consumer could keep a copy of this version, publish it and so frame the receiving consumer.

- Another possibility to frame other consumers is to use fingerprinting on a document without even performing a transfer and publish the resulting document.

Data Lineage Generation Module

- 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 such that the auditor initially takes the owner as the current suspect.
- The auditor appends the current suspect to the lineage. The auditor sends the leaked document to the current suspect and asks him to provide the detection keys k1 and k2 for the watermarks in this document as well as the watermark. The auditor outputs the lineage. The last entry is responsible for the leakage.

Outsourcing Module

- In this module, we develop a typical outsourcing scenario. An organization acts as owner and can outsource tasks to outsourcing companies which act as consumers in our model. It is possible that the outsourcing companies receive sensitive data to work on and as the outsourcing companies are not necessarily trusted by the organization, fingerprinting is used on transferred documents.
- The outsourcing company itself can outsource tasks to other outsourcing companies and thus relay the documents, again using fingerprinting. It is important to notice that a single organization can outsource to may different outsourcing companies in parallel, thus creating a tree-shaped transfer diagram.
- If now at any point one of the involved outsourcing companies leaks a confidential document, the organization can invoke the auditor to find the responsible party. The auditor then asks the organization to reveal the first set of fingerprints in the leaked document, which leads the auditor to one of the outsourcing companies. This outsourcing company can in turn reveal additional fingerprints in the leaked document in order to point to the next outsourcing company and to prove its own innocence.
- Finally, the auditor creates the complete lineage and is able to determine the guilty party. The responsible party can be clearly found using LIME.

VI. CONCLUSION

The chances that a data consumer is culpable for data loss is checked on the basis of overlay of his data with the exposed data and the data of other consumers, and based on the possibility that data items can be presumed by other modes. The lineage approach appliances a wide range of data circulation methodologies that can boost the owner's likelihood of finding leakage and diagnosing a data leaker. Thus the data provenance model is effective than the existing watermarking model. Data provenance model caters protection to data at the time of circulation or transmission of data and can also find if that gets leaked. Watermarking safeguards the data using techniques like encryption, whereas data provenance model provides prevention plus guilt identification. This model proves to be advantageous to enterprise, where data is disbursed using any public or private medium and shared with the outsider (third party). Now, enterprise, numerous organizations can entrust or depend on this data provenance mode.

REFERENCES

[1] Chronology of data breaches [Online]. Available: http://www.privacyrights.org/data-breach, 2014.
[2].Pairing-based cryptography library (PBC) [Online]. Available: http://crypto.stanford.edu/pbc, 2014.
[3] Privacy rights clearinghouse [Online]. Available: http://www.privacyrights.org, 2014.
[4] Facebook in privacy breach [Online]. Available: http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html, 2010.

[5] Offshore outsourcing [Online]. Available: http://www.computerworld.com/s/article/109938/Offshore_outsourcing_cited_in_Florida_data_leak, 2006.
[6] A. Mascher-Kampfer, H. St€ogner, and A. Uhl, "Multiple re-watermarking scenarios," in Proc. 13th Int. Conf. Syst., Signals, Image Process., 2006, pp. 53–56.
[7] P. Papadimitriou and H. Garcia-Molina, "Data leakage detection," IEEE Trans. Knowl. Data Eng., vol. 23, no. 1, pp. 51–63, Jan. 2011.
[8] (1994). Electronic privacy information center (EPIC) [Online].Available: http://epic.org, 1994.
Data breach cost [Online]. Available: http://www.symantec.com/about/news/release/article.jsp?prid=20110308_01, 2011