**Data Lineage in Malicious Environments**

**ABSTRACT:**

Intentional or unintentional leakage of confidential data is undoubtedly one of the most severe security threats that organizations face in the digital era. The threat now extends to our personal lives: a plethora of personal information is available to social networks and smartphone providers and is indirectly transferred to untrustworthy third party and fourth party applications. In this work, we present a generic data lineage framework LIME for 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 non-repudiation and honesty assumptions. We then develop and analyze a novel accountable data transfer protocol between two entities within a malicious environment by building upon oblivious transfer, robust watermarking, and signature primitives. Finally, we perform an experimental evaluation to demonstrate the practicality of our protocol and apply our framework to the important data leakage scenarios of data outsourcing and social networks. In general, we consider LIME , our lineage framework for data transfer, to be an key step towards achieving accountability by design.

**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.
* 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 REQUIREMENTS**

# Hardware Requirements:

# Processor - Pentium –IV

* Speed - 3.5 Ghz
* Ram - 2 gb
* Hard Disk - 40 Gb
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - SVGA

**Software Requirements:**

* Operating System : Windows XP/ Later Version
* Coding Language : Java/J2EE
* Front End : Java
* Database : MySql