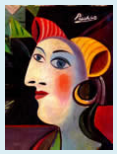
SPROV 2.0: A Highly-Configurable Platform-Independent Library for Secure Provenance

Ragib Hasan*#, Radu Sion+, Marianne Winslett* *University of Illinois at Urbana-Champaign, +Stony Brook University, #Johns Hopkins University

Will you spend $101 million to buy a fake painting?

Real Picasso, worth $**101.8** million



Fake, listed at eBay, worth nothing

Obviously, **No**! Buyers look for provenance, i.e. records of the paintings ownership, exhibition, and sales history. A painting without secure provenance is deemed fake.

When your data comes from/originates elsewhere, was processed by someone else, and owned and modified by many people, how do you know it is trustworthy?

Problem Definition: How to make digital history trustworthy?

Approach: Make data **provenance** tamper-evident.

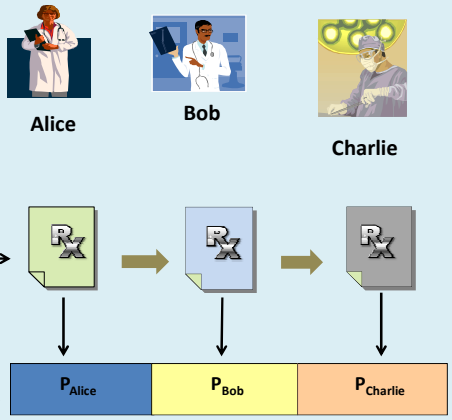
Provenance: from Latin *provenire* = '*come from*', defined as

"(i) *the fact of coming from some particular source or quarter; origin, derivation.*

(ii) the history or pedigree of a work of art, manuscript, rare book, etc.; a record of the ultimate derivation and passage of an item through its various owners" (Oxford English Dictionary)

In other words, **Who** owned it, **what** was done to it, **how** was it transferred …

So, we need provenance of provenance, i.e. a model for **Secure Provenance with Confidentiality, Integrity, and Privacy guarantees.**



Scenario

Dana goes to Dr. Alice, who refers her to Dr. Bob. Dr. Bob's misdiagnosis later causes Dr. Charlie to prescribe a wrong treatment.

Dana sues Charlie for malpractice. Dr. Charlie wants to show the provenance of Dana's medical records in his defense. Bob wants to hide his own misdiagnosis, by changing the provenance record P_{Bob}

Threat Model

Users: Edit **documents** on their machines

Auditors: semi-trusted principals. All auditors can verify chain integrity. Only certain auditors can read each entry

Adversaries: insiders or outsiders who

- Add or remove history entries
- Collude with others to add/remove entries
- Claim a chain belongs to another document
- Repudiate an entry

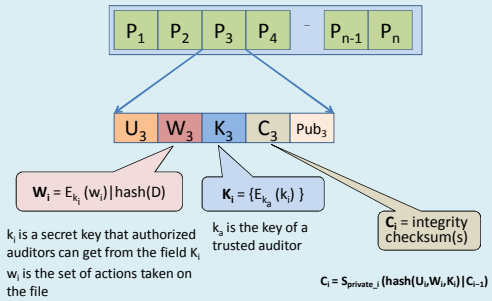
Our Guarantee: Ensure Plausible History

Plausible history: The version history that will result if a document were created and subsequently edited and transferred from user to user, with provenance information **correctly** and **indelibly** recorded all along the way.

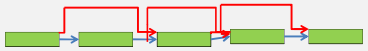
Guarantee:
If a provenance chain **does not** give a plausible history for its associated document, we will **detect** this.

Analogy: A forger's goal is to create a **plausible history** for a fake Prada bag (i.e. make it appear to come from real Prada distributors via real supply chain)
Forger's goal is NOT to replace a real Prada bag's plausible history, to show it was made elsewhere.

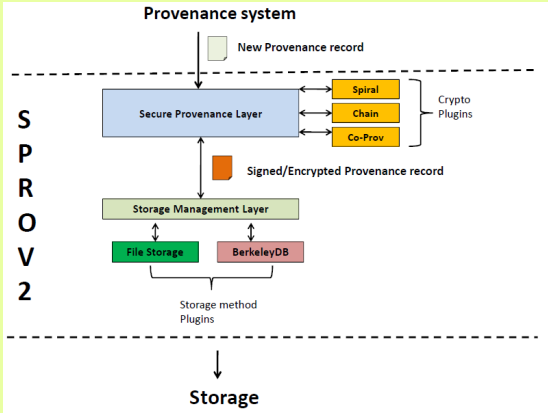
Solution



Augmented Provenance Chains: Use the Integrity spiral construct to provide Integrity-preserving summarization of provenance chains.
Construct the spiral checksum by using more than one previous checksums to create new cumulative checksum.



The SPROV 2.0 Architecture



Goals: Design a highly configurable architecture for providing secure provenance functionality to existing systems

Approach: Divide functionality into plugins for cryptographic operations and storage management.

How SPROV 2 Works

1. Provenance system (e.g. PASS) creates a new provenance record, and hands off the data to SPROV2
2. SPROV2 applies crypto-plugins to create checksums and encrypt the record
3. Storage method plugins handle chain storage / IO

Cryptographic Plugins :
- Linear provenance chains
- Spiral provenance chains
- Co-provenance chains
- Encryption

Storage Plugins :
- Berkeley DB storage
- Local file storage
- Cloud storage

Status
- Under development
- To be deployed with PASS

For further details http://dais.cs.uiuc.edu/provenance and http://tinyurl.com/secprov

See Also: Hasan et al., "The Case of the Fake Picasso: Preventing History Forgery with Secure Provenance, USENIX FAST 2009 | Secure Provenance: The Genealogy of Bits, USENIX ;login: June 2009.