http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (23): 7568-7573, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7568.7573

Research on Provenance Security of E-commerce Information

[1]Fengying Wang,[2]Xiumei Li, [1]Caihong Li and [1]Hui Zhao
[1]College of Computer Science and technology, Shandong University of Technology,
Zibo, Shandong, 255049, China
[2]Zibo High School, Zibo,Shandong, 255000, China

Abstract: Information provenance is a new research field that can be used to determine the source, quality and reliability of the Information. We focus on the unique security needs of the provenance information in electronic commerce and study regrowth method of broadcast encryption tree for the insufficiencies of encryption scheme of information provenance. We propose a new security provenance model with the introduction of time-stamp technology in order to increase the security of signatures which protect the security of provenance information.

Key words: Provenance confidentiality, security provenance model, broadcast encryption

INTRODUCTION

Information provenance is a new research field that can be used to determine the source, quality and reliability of the Information. Recently, the application of digital documents with provenance in the financial, commercial, medical, scientific and legal environment has become very important. Up to now, research on the provenance has mainly focused on the works of modeling, computing, storage, querying. Yet, few works focus on ensuring the security for provenance. With the increasing importance of credibility of electronic data, the guarantees of provenance security are more important than ever. Data provenance has the unique characteristics that are different from the data. However, the existing security models are not well applied to the provenance, so the need of safety protection for provenance is different from that of data.

As provenance continuously used for the areas of digital copyright protection, DNA testing, drug testing, corporate financial accounting and national intelligence, provenance information is also faced with security threats growing more serious, including active attacks from the adversaries. The significant incentives of the attackers is that change the provenance records according to the value of scientific data, where the value of scientific data rests upon the provenance information by which the data was created and by whom. Users need to trust the provenance information associated with the data that can be accurately reflected the data being created and the process of data being transformed. But without proper protection measures with the information through a different application layer or untrusted environment, provenance information associated with it can be subject to accidental damage and they can even be vulnerable to malicious forgery.

In e-commerce, information is a core element of e-commerce systems. The bottom-up information of the enterprise is that the enterprise's daily operational information is gathered from the grassroots to the senior levels. Information is derived from the lowest level of the organization and then passed upward for making a strategic decision. For example, when the business of collecting information on structure of the buyer comes up, the officer A of company gathers the buyer's information in accordance with the purpose and need and then passes on to the officer B, who will analyze and make judgments, forming a result, improve the information value. Eventually the officer B transfers resulting analysis to the manager C for reasonable decisions. In this process, the provenance chain ($P_A|P_B|P_C$) of the information is generated, shown in Fig 1. The quality of information collected, that is the authenticity, reliability, accuracy, confidentiality of the information, will determine whether it achieve the intended purpose and enhance economic efficiency of enterprises. Thus, internal or external attackers may have a clear incentive to alter the history of data records. If the staff B made irrational judgments for commodity purchases, will result in the wrong sales strategy. In order not to affect his performance evaluation, the staff B may want to hide his wrongdoing by tampering with the provenance record matches with his action.

With the data and its provenance information passing through the different users and tasks in the untrusted environment, the provenance information are vulnerable to unauthorized alter, therefore, providing

Corresponding Author: Caihong Li, College of Computer Science and technology, Shandong University of Technology, Zibo, Shandong, 255049, China

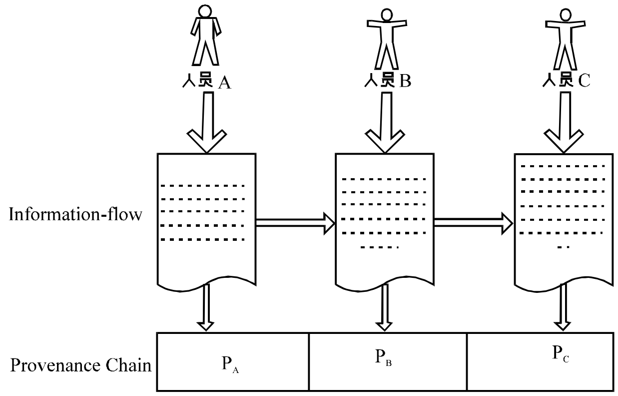

Fig. 1: Example of provenance

integrity, confidentiality, security for the provenance has become very important. This paper discusses the existing security issues associated to provenance and further investigates the problem of provenance security model of information and security threats model. A new provenance security model is put forward by introducing regrowth method of broadcast encryption tree for the insufficiencies of encryption scheme.

SECURITY PROVENANCE MODEL

In this section, we introduce the fundamental concepts concerning the provenance and propose threat model by analyzing internal or external attacks possibly occurred in information of the intra-business.

Provenance model: We use the term document to refer to the data object for which provenance information is collected, such as a file, database tuple and information or network packet. The provenance information is the record of modification taken on the document over its lifetime. Accessing a document D each time may produce a provenance record P. A provenance chain for document D is a non-empty time-ordered sequence of provenance records $P_1|\cdots|P_n$(Hasan *et al.*, 2009). When a user changes a document, a new provenance record that describes this change is added to the provenance chain and the user allowed the auditors or its subset to read the new record. According to the provenance chain we can roll back the evolution of the document, track the source of a document and the modifications in the document lifecycle. Each record of the provenance chain describes the current state of the document, such as the user ID accessing to the document, process ID, the access actions (read or write), the relevant data (documents bytes), as well as the description of the environment in which action occurred (including the host ID, IP, date and time) and integrity and confidentiality-related security components, such as checksum, cryptographic signatures, key materials and digital time-stamp. Of the document to the origin of each operation is recorded in the form of being recorded, the document is deleted it's the origin of information is no longer meaningful. Each operation is documented in the form of a provenance record. After an document has been deleted it's provenance chain has no meaning anymore.

Threat model: When information crosses application and organizational boundaries and passes through untrusted environments its associated provenance information is vulnerable to malicious forgery. Access control is insufficient to prevent this tampering, as unauthorized users may have physical control over a machine where the information resides. If there is no special security guarantees for provenance, unauthorized users can easily modify the data and tamper with its associated provenance and even remove the corresponding provenance record of the chain or store the forgery provenance records in the chain, these operations are difficult to be found. The purpose of this paper is to construct an effective scheme for detecting illegal tampering with provenance records.

We propose a threat model and the ideal security guarantees based on the model described by (Ragib etc). Suppose in a security organization, the user are principals who read and write documents and their provenance. Each organization has one or more auditors, who are principals authorized to access and verify the integrity of provenance records associated with documents. Attackers are principals from inside or outside an organization who have access to a document and its provenance chain and who want to alter them inappropriately.

Suppose that the provenance records P accurately described the transmission process of the information but an attacker may want to forge history by modifying the document or the provenance records of P. Therefore, we have listed the following guarantees required by security provenance:

- **S1:** An attacker can not selectively modify the records of any users (including himself) provenance chain. As shown in Fig. 1, the staff B can not hide his wrongdoing by amending his provenance records which would be detected
- **S2:** An attacker can not selectively remove the records of any users (including himself) provenance chain
- **S3:** An attacker can not insert the records into the beginning or middle of the provenance chain.
- **S4:** Users can not repudiate adding the provenance records
- **S5:** An attacker can not claim that the record of a document belongs to other documents
- **S6:** An attacker can not only modify a document without add the correct provenance records describing this modification to the chain
- **S7:** Provenance chain itself can not be modified, that is an attacker can not destroy the sequence of provenance records
- **S8:** Two colluding attackers cannot insert provenance records of non-colluding users between them
- **S9:** Two colluding attackers cannot selectively remove the non-colluding users' record between them
- **S10:** The auditors can verify the integrity of provenance chain without access to any confidential components, unauthorized auditors accessing to confidential provenance record will be detected

It is worth noting that if the attacker has actual control over the machine, then he can remove the provenance chain completely to prevent the information from being used properly, thus we should reply on trusted hardware. On the other side, an attacker claims to be the original creator by copying a document manually or automatically and thus forged the identity of the creator. Therefore, the model we constructed is to detect tampering, the main security factors we considered is to prevent malicious attacker from damaging part of the chain.

PROVENAN CESECURITY COMPONENTS

Structure of provenance chain: According to analysis of the security provenance model, this section gives more detailed description for the structure of provenance chain. Provenance record is the basic units of the chain and each record P_i summarizes a sequence of one or more actions taken by a user. Provenance record can be defined as the following formation:

$$P_i = \langle U_i, M_i, \text{hash}(D), K_i, TS_i, C_i, Pub_i \rangle$$

The descriptions of each field are as follows:

- U_i is the plaintext or ciphertext identifier of user

For example, in the employee's performance assessment employees are allowed and even encouraged to read their performance assessment in order to improve themselves. But employees can not read who had input to their performance assessment. Therefore, employees can read only the document rather than the provenance of the document. In this case, the user identifier describing the current state of document should be encrypted ciphertext using the session key:

- M_i is ciphertext or plaintext representation of a series of operations (change log) by a user
- hash (D) is the one-way hash value of the current contents of the document.
- K_i is key material, including the keys that can be used by auditor to decrypt the encrypted fields
- TS_i is a timestamp
- C_i contains the integrity checksum of the provenance records signed by the user.
- Pub_i is the encrypted or plaintext public key certificate for user U_i

Next are the security components related to the field.

Integrality of provenance chain: When the user modify a document, simultaneously he generate hash(D) by applying one-way hash to the document. And further hash this hash(D), modification M_i, key material K_i, the user identity U_i, as well as the user public key certificate Pub_i. Then we sign the resulting hash, the timestamp TS_{i-1} and checksum C_{i-1} of previous provenance record P_{i-1} using the user's private key. Timestamp is typical of the uniqueness and nonreversibility, so it can not be verified for the altered provenance records. The integrity checksum field is defined as follows:

$$C_i = S_{\text{privatei}} (\text{hash}(U_i, M_i, \text{hash}(D), K_i, Pub_i)|C_{i-1}|TS_{i-1})$$

In order to increase the security of signatures, we introduce the timestamp technology. In a simple hash-and-sign time stamping, the Digital Time Stamp Service receives the digest of a user identity U_i hashed by user,

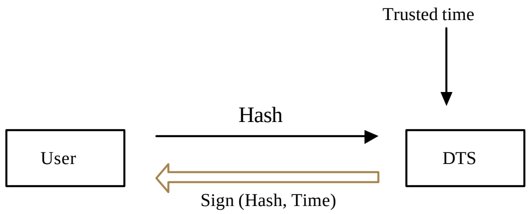

Fig. 2: Digital time stamp service

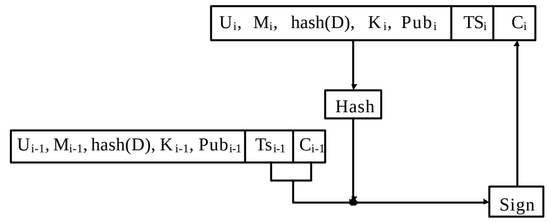

Fig. 3: Integrality of provenance chain

appends the current date and time to it and signs the resulting data structure(digital signature) and then returns it back to the user, as shown in Fig. 2. The timestamp and the checksum are stored in the corresponding fields of provenance records. The form of timestamp is defined as follows:

$$TS_i = S_{DTS}(hash(_{Ui}), Time_i)$$

The auditor obtains the information from provenance records, determine the source of data by computation of checksum and verify whether the data has been maliciously modified in the process of transmission. To verify chain integrity, starting from the first record of the chain, the auditor extracts the user identity U_i and the field of Pub_i from the record and verify that the Pub_i is public key certificate of the user U_i. Then the auditor decrypt the checksum using the user's public key from the Pub_i and get hash(U_i, M_i, hash(D), K_i, Pub_i). The auditor then hash the U_i, M_i, hash(D), K_i, Pub_i of current record, denoted as hash'. The provenance information has not been modified if the hash = hash', as shown in Fig. 3.

Because each modified information is cryptographically one-way hashed. For example, in Fig 1, the user B wanting to hide evidence of his misaction, changes the provenance records or information in the document which will give rise to change the value of hash(U_i, M_i, hash(D), K_i, Pub_i), that will be detected. If the external attacker would like to modify the provenance records, he needs to obtain the signature of corresponding user, or find a hash collision, so the security of S1 is guaranteed.

If an attacker wanting to insert or delete provenance records can also be detected because of the checksum C_i of each record containing the checksum C_{i-1} of previous record which is sufficient to ensure the security of S2, S3, S8, S9.The non-repudiation of S4 is guaranteed by the signature on provenance checksum of the chain. We can verify the timestamp of each provenance record to gurantee the sequence (S7) of the records not destroyed.

Confidentiality of provenance record: In the systems that the provenance is more sensitive than the data, in order to ensure the confidentiality of the provenance information,

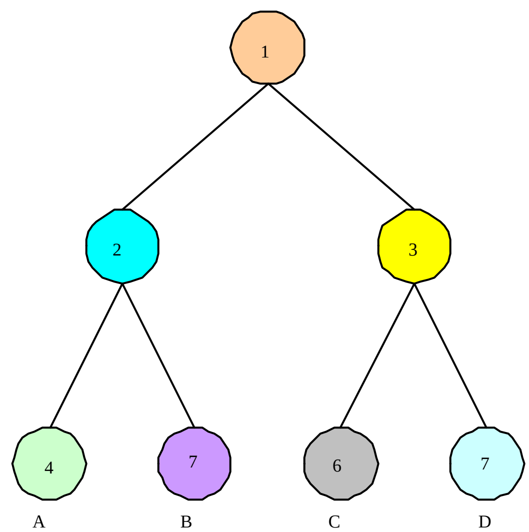

Fig. 4: Broadcast encryption tree

]the modification log m_i should be encrypted using the session key k_i of the user, that is $M_i = E_{ki}$ (m_i) as defined in 3.1 and multiple copies of the key k_i maked using the broadcast encryption tree in accordance with the number of auditors. Then we encrypt the each copy with the public key k_a of a deifferent trusted auditor, that is the key material fields $K_i = E_{ka}(k_i)$ in 3.1. In this process it is only a certain trusted auditor that can decrypt a copy of the corresponding user and use it to decrypt sensitive fields. Broadcast encryption tree is a binary tree with nodes of the keys, in which each node contains the public/private key pair in PKI and the leaf nodes correspond to the auditors. Each auditor knows the private key from the leaf node to the root node while give all the public key of the tree to each user and auditor. Fig. 4 shows the broadcast encryption tree of leaf nodes of A, B, C, D for auditors, if the user only trusts auditor A, the user encrypts the session key only using the public-key of node 4; if the user confidence in auditor C and D, only using the public key of node 3; Correspondingly, if the user confidence in all of the auditors, then using the root node simply.

Since the number of the auditor may increase dynamically while the conventional broadcast encryption mechanism can not extend the number of auditors and therefore need to add new auditors in batches without influencing the previous users.

RELATED WORK

The provenance is the description for the current state of the data as well as the process of creation, modification, transformation of data over its lifetime. The computation of provenance is not a new problem. Cui *et al.* (2000) first proposed the problem of tracing the data provenance which was the first study of the reverse query that considered as why-provenance according to (Buneman *et al.*, 2001). Who put forward the where-provenance at the same time. It is the where-provenance type of provenance that we use for determining where the annotations are propagated from and how. Data provenance is closely related to the problem of update views and annotation is becoming increasingly the most useful approach for scientific computing. Literature (Bhagwat *et al.*, 2004) designed and implemented an annotation management system for relational databases according to idea that annotations can be made on relational data This was first proposed in (Buneman *et al.*, 2002; Tan, 2004). This was the first implementation of an annotation management system for relational databases that would allow a user to specify how annotations should propagate. The quality or security level of a piece of data can also be described in annotations. Since annotations are propagated along as a query is executed, the annotations on the result of a query can be aggregated to determine the quality or degree of sensitivity of the resulting output. However, the practical operation focus on collecting and storing the information, rather than the security and credibility of provenance. This does not satisfy the challenges of confidentiality, integrity and privacy of provenance information.

So far, various system architectures have been proposed for collecting and maintaining provenance records: Some gather modification of information and attach provenance to the data itself as a form of annotation (Buneman *et al.*, 2006) and others deposit provenance in one or more repositories (Chapman *et al.*, 2008; Davidson *et al.*, 2007). Thus, we need to take different measures of security guarantee for different memory models, based on their different needs for security. Related research of Rigib etc. has focused on security (integrity and confidentiality) of tracing and storing provenance for file system (Cui *et al.*, 2000). They consider the provenance as a type of metadata. When documents move from one user to another user, provenance information and other metadata move with the documents. On this basis, (Zhang *et al.*, 2009) have the first in-depth study of integrity and tamper-evidence for database provenance and propose the security model providing integrity for database provenance. This paper investigates provenance security model for information of e-commerce based on the Rigib's research.

CONCLUSION

Provenance generally refers to recording the source of data and its entire processing steps of the subsequent conversion, reflecting the static information in a state and the dynamic characteristics in the transformation processes of data. Therefore, provenance can be used to determine the origin, quality and reliability of the data. It is the basic requirement to tracing the data provenance in terms of the protection of the rights, the management for intelligence and medical data, message authentication and etc. This requires a well provenance mechanism to guarantee the integrity and confidentiality of provenance information. This paper analyzes the existing security issues related to provenance, improves the threat model to ensure that an attacker can not change the order of the provenance records; introduces regrowth method of broadcast encryption tree for the insufficiencies of encryption scheme of information provenance; and a new provenance security model is put forward with the introduction of time-stamp technology in order to increase the security of signatures. But this model will also be continually improved.

ACKNOWLEDGMENT

This research is supported by the Science and Technology Development Plan Program of Shandong Province under Grant 2013GGX10116 and by the Higher Educational Science and Technology Program of Shandong Province under Grant J13LN27 and by the Natural Science Foundation Program of Shandong Province under Grant ZR2013FM013.

REFERENCES

Bhagwat, D., L. Chiticariu, W.C. Tan and G. Vijayvargiya, 2004. An annotation management system for relational databases. Proceedings of the 30th International Conference on Very Large Data Bases, August 29-September 3, 2004, Toronto, Canada, pp: 900-911.

Buneman, P., A. Chapman and J. Cheney, 2006. Provenance management in curated databases. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 27-29, 2006, Chicago, IL., USA., pp: 539-550.

Buneman, P., S. Khanna and W.C. Tan, 2001. Why and where: A characterization of data provenance. Proceedings of the 8th International Conference on Database Theory, January 4-6, 2001, London, UK., pp: 316-330.

Buneman, P., S. Khanna and W.C. Tan, 2002. On propagation of deletions and annotations through views. Proceedings of the 21st ACM Symposium on Principles of Database Systems, June 2-6, 2002, Madison, WI., USA., pp: 150-158.

Chapman, A.P., H.V. Jagadish and P. Ramanan, 2008. Efficient provenance storage. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 9-12, 2008, Vancouver, Canada, pp: 993-1006.

Cui, Y., J. Widom and J. Wiener, 2000. Tracing the lineage of view data in a warehousing environment. ACM Trans. Database Syst., 25: 179-227.

Davidson, S., S. Cohen-Boulakia, A. Eyal, B. Ludascher and T. McPhillips *et al*., 2007. Provenance in scientific workflow systems. IEEE Data Eng. Bull., 30: 44-50.

Hasan, R., R. Sion and M. Winslett, 2009. The case of the fake Picasso: Preventing history forgery with secure provenance. Proceedings of the 7th USENIX Conference on File and Storage Technologies, February 24-27, 2009, San Francisco, CA., USA., pp: 1-14.

Tan, W.C., 2004. Containment of relational queries with annotation propagation. Proceedings of the 9th International Workshop on Database and Programming Languages, September 6-8, 2003, Potsdam, Germany, pp: 37-53.

Zhang, J., A. Chapman and K. LeFevre, 2009. Do you know where your data's been?-Tamper-evident database provenance. Proceedings of the 6th VLDB Workshop on Secure Data Management, August 28, 2009, Lyon, France, pp: 17-23.