Research Article

Symbiotic view of Provenance in Cyber Infrastrcuture and Information Security

Kukatlapalli Pradeep Kumar*[1], Cherukuri Ravindranath Chowdary[2], Vinay Jha Pillai[3], Sarath Chandra K[4], Boppuru Rudra Prathap[5]

[*1] Assistant Professor, Computer Science and Engineering, School of Engineering and Technology, Christ University, Bangalore - 560074, India
[2]Associate Professor, Computer Science and Engineering, School of Engineering and Technology, Christ University, Bangalore - 560074, India
[3]Assistant Professor, Electronics and Communication Engineering, School of Engineering and Technology, Christ University, Bangalore - 560074, India
[4]Assistant Professor, Civil Engineering, School of Engineering and Technology, Christ University, Bangalore - 560074, India
[5]Assistant Professor, Computer Science and Engineering, School of Engineering and Technology, Christ University, Bangalore - 560074, India
kukatlapalli.kumar@christuniversity.in*[1],cherukuri.ravindranath@christuniversity.in[2]
vinay.pillai@christuniversity.in[3], sarathchandra.k@christuniversity.in[4], boppuru.prathap@christuniversity.in[5]

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract— Access control is one of the important elements in providing confidentiality to the secured data. Access specifiers helps us understand degree of rights given to the users in utilizing data records in a right manner. Tampering the records by unauthorized parties is a high concern in secure communication. Tamper detection plays an important role in trouble shooting an issue associated with network/ host intrusion scenario. The advances in computer technology has driven the contemporary world to focus on the World Wide Web for digital data and associated information. People across the globe rely on the internet for all the data from local information to distribution of personal data through heterogeneous networks. Technology and software tools has grown so broad to an extent, where almost all of the financial transactions are taking place through online portals. On the other hand, there has been high growth in the security coercions towards user's sensitive data. This information is however shared by the online users while performing financial transaction in e-commerce portals. In order to maintain security mechanism over the untrusted networks various authentication techniques available in this regard. All these security procedures are said to be stubborn and adequate on contextual basis, on the other hand over a period of time the intruders find out means to break into systems. Data theft and intrusion into the information systems would increase on a daily basis if defensive measures are not in place. We integrate concepts of secret sharing and data provenance to provide an indigenous solution for parameters of information security namely confidentiality, integrity and availability.

Keywords—Data Provenance, Secret Sharing, Information Security, Access control.

1. Introduction

When data is transmitted between a point to another point or point to multi point, it can be visualized as data communication scenario. Sender, receiver, protocol followed, communication channel and packets of data are said to be components of communication. In this context, data or message is transmitted through a communication channel in an encrypted fashion for providing security for the same. At the receiver end, the data or message is decrypted and processed further.

Fig. 1 demonstrates the concept of communication channels in the networks. One is a high-risk channel with less safety aspects and another is a low-risk system, coupled with good security features [1]. Along with the communication channels with their risk factors, source and destination aspects are also depicted in the above illustration. The low risky communication station is denoted with X1 Y1 Z1 and high risky communication station is with X2 Y2 Z2. The naming convention of the channels depicts the parameters with regards to fundamental security aspects namely confidentiality, integrity and availability. It is with the availability of exceptional wireless internet access in mobile motivated situations, customers and their data has turned out to be massive with respect to media. For example, financial related operations carried out over online platforms by users in many ways were found insecure and unauthenticated.

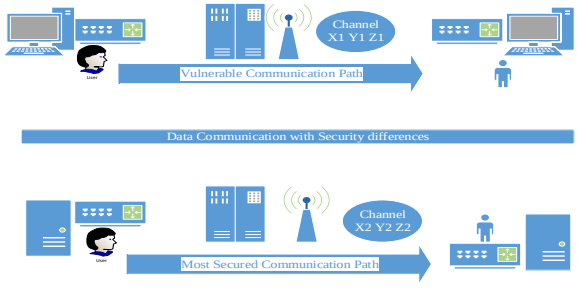

Fig. 1 Scenario of communication channels over data transmission.

Procedure with appropriate algorithms are available for safe data communication in various modes, however lacks to attain high accuracy and performance with regards to the fundamental objectives of information security (the CIA traid) at a significant extent. Security is the main aspect of any communications among untrusted networks in the current world [2]. Sincere gratitude to many researchers for their tremendous contributions to effective security algorithms despite various threats that compromise the computer systems vulnerabilities. The source of the information, i.e., by which the online operation was created, is the pertinent query to be countered while the transaction is finalized. This definition of 'data antiquity' has received decent interest from investigators in different fields for many decades and is often termed as data provenance [3-5]. However, security in provenance has made some progress with recent research, particularly in the field of cyber security. The below sections explains about the aspects of data provenance and visual encryption followed by literature analysis with results and discussions.

2. Literature related to Data Provenance and VisuaL Encryption

Description and representation about the genesis, lineage and pedigree of an object is mentioned as provenance. With respect to data object and its associations the same is characterized as data provenance. Provenance data is delicate and a slight disparity tips to alteration in the complete sequence. This creation desired to be safeguarded and admission should be approved for sanctioned party. Data provenance provides a record of lineage with regards to transactions, events, processes and systems that influence the data of concern. This kind of record or lineage data provides better understanding on aspects related to data dependency and associated relationships. A multiple entity access control is showcased using secret sharing mechanism. Secret sharing is connected via visual encryption process. However these security mechanism are deployed for safe guarding sensitive genesis data called as provenance data. The proposed work sheds knowledge on a background allied to safeguarding provenance through secret sharing security concept.

A. Literature aspects of Data Provenance

Data provenance is the record or collection of events and transitions that a data item has experienced in its life cycle. Provenance information is often called as genesis data and is sensitive information. A small modification made to genesis data changes the entire lineage concerning to its variables and functions. A detailed literature is shown in this sub section on data provenance and security.

Visual cryptography is a strong development in the research area of information security, which allowed the encryption methodology with little mathematical calculations. The visual cryptography concept of shares [6] is however the heart of the design. In a few instances, the data transmitted using this form were also targeted at the attackers end. The public key encryption for the created image shares is embedded to render the method robust [7]. With time, the importance and relevance of video streaming data decreases, and this decline is also rightly proportional to the reliability of the multi-media data. The Burrow Wheeler Encoding Algorithm [8] was based on images with the suggested conditional transposition technique for short-period visual safety, taking into account graphical encoding of large-scale data and security unification. Analyzing the creation and sources of the genuine data has become an vital aspect of latest study following developments in the media with the public exchanging data on the internet. It is the term, protection that is catching the interest of common internet users. In an untrusted contexts, consumers are more concerned about the security of their knowledge that they exchange [9]. Establishing a structured arrangement for authentication of data provenance. Developed a basic paradigm that describes provenance and correlates it with the essential consequences for security fundamentals namely integrity, confidentiality and availability.

McDaniel, P. stated in his paper [10] the enormous amounts of information from both inner procedures and the far-flung, untrusted, unfamiliar foundations that societies and different users have provided. So, data provenance and information security are performed in a similar style. This information has to be polished to verify with any of the weaknesses allied. Provenance confirmation is also an significant activity that must be accomplished with data security susceptibility tests. Fig. 2 Illustrates the interaction between provenance, information security priorities and related fields of operation. With the arrival of the WWW (world wide web) and its popularity across the globe reaching enormous speeds, people have become very close to access data over the internet. Data streaming on images capture and audiovisual processing, particularly for the sensor systems, yields large amounts of information. As far as attribution is associated in this field, secure communication of provenance data is a challenging. There is an innovative method to data transmission [11]; which embeds provenance into the inter-packet scheduling domain based with regards to sensor networks. The data receiver extracts provenance using an optimized threshold-based process which reduces the possibility of deciphering mistakes from provenance.

The "Chain-structure" provenance system [12] points out the protection features of the data explicitly in a hierarchical setting providing three-dimensional meta-data provenance.

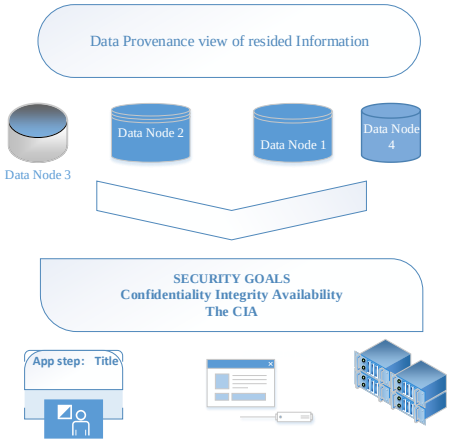

Fig. 2 Proposed model in connection with security aspects and provenance

Cloud storage is one of the latest technology developments, and its infrastructure and retrieval features have evolved to a larger extent. Organizations are expanding activities at various sites without caring about their expenditures in the network, with cloud computing and applications linked to virtualization. For the customers who pay for those services, protection for the same is an evolving problem. A novel technique for tracing entire lineage of data provenance on the descent of data history is available [13]. This uses data provenance algorithms based on rules to track customer data for cloud-based leakage risks.

B. Model of Provenance activities

Provenance basic understanding and its purest form of definition can be observed in W7 model [14]. Any information entity can be analyzed with seven Ws viz. Who, What, Why, Where, When, Which and How. Rationale for understanding issues and troubleshooting the same becomes easier with this model. A semantic repository is created and saved for all the data records with available parameters. Illustration of the afore mentioned model is generalized in Fig. 3 We are focusing on securing the provenance with a unique security mechanism in secret sharing.

C. Literature aspects of Visual Cryptography

Visual encryption is a cryptographic system using images to cover up the input message. In this technique deciphering will not need calculations. The input is reflected as group of black and white pixels. The input is broken down into two parts. Every share contains a series of black pixels and white sub pixels. Each share is printed distinctly, in which pixels are arranged. One is the cipher text published page and the other is the transparency in printed form. When the message is decoded, the encrypted image is put in a particular arrangement for getting the actual message as the text.

Fig. 3 Category of shares in black and white pixels

Fig. 3 illustrates an overview of the black and white pixel division of 4 sub pixels. The first element is horizontal, the second is upright and the third is crosswise.

The Fig. 4 elucidates about base concept of visual encryption in a pictorial form. The elements in middle of the diagram are the communication entities. Switch, Hubs and communication links are taken into consideration.

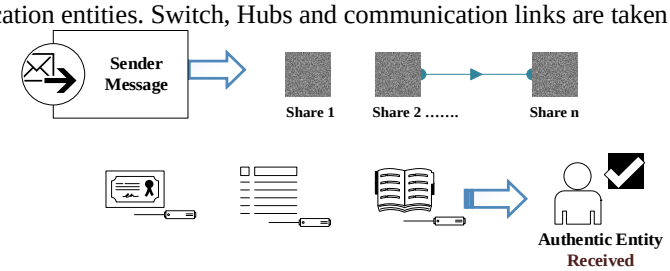

Fig. 4 The base concept of visual encryption

D. Application aspects of secret shares

To boost the picture effectiveness after implementation of visual cryptography the following approaches were suggested. They are termed as XOR constructed visual cryptography for GAS (General access structure) and adaptive section implementation using exclusive OR. In the conservative VC (visual cryptography), the procedures used OR for image rebuilding. These 2 techniques use exclusive OR based procedures in VC to deliver decent visual precision for the rebuilt image and safety to the transparencies of the secret image [15].

Lin et.al paper focuses on executing secret sharing through the combination of various twofold encryption techniques and visual cryptography methods [16]. It emphases on hiding information in image based. First, two shares information are designed through the pixels of the original data with a distribution matrix. Then a data encryption rule is used to interpret these two matrixes into an image which converts to a cover key. As per the authors, this technique is simple to use and is safer and does not require an input image for the decryption process..

Main issue of the traditional visual cryptography is usage of pixel expanding for encryption. Lee et.al suggested a methodical and non-sophisticated scheme to eliminate the use of pixel extensions. [17].

Daisy et.al defines the more powerful visual cryptography technique that addresses most of the problems previously existing, such as pixel expansion, degraded contrast, decreased visibility, etc. [18]. Huang et. al. presents a non-expanded visual cryptography scheme where there is not much improvement in the resulting picture compared to the original one. [19]. Here you can see the original picture of 'Division, Sharing and Superimposition.' This avoids pixel augmentation.

Jung-San Lee et.al. and others dealt with 2 notions of network communication protection, i.e. the visual secret sharing technique and use of CAPTCHA[20]. This deals with the simple Visual Secret Exchange scheme using the traditional method of using 4 (2, 2) subpixel classes in Visual Cryptography. The definition of CAPTCHA is also used because only HVS (Human Visual System) will decode the stacked picture and not any machine source. Three types of images are considered here. Type I has twisted black and white contextual typescripts. Form II consists of warped and vivid characters. The background was detached in type III, which helped achieve an accuracy rate of 0.96. The above outcomes were showed up based on the comparison.

Nitty et. al. listed the process of diffusion of errors in the visual cryptography of halftones [21]. Node error diffusion substitutes a pixel with a reference node in classical error diffusion so that coarse quantization errors are minimized. In a pixel row, the correct proportions of the quantization error at each pixel are diffused to the chosen pixels in the neighboring lines. A half-tone VC image share is divided into non-overlapping half-tone cells of the size q = v1 x v2. In each share a secret pixel of the image is encrypted into one halftone cell. Kumar et.al suggested a definition that aims to improve the already existing methods of visual cryptography [22].

3. Results and Discussion

At the core of the approach suggested is the preservation of the origin or genesis of the database. Fig. 5 demonstrate the same with three shares available at primary power, secondary power, third authority. A legal share checking process is in place before the access is provided for the concern [23]. Producing the shares for access control is performed by visual encryption mechanism. Keyword for retrieving the vital file is known by peculiar combination of shares.

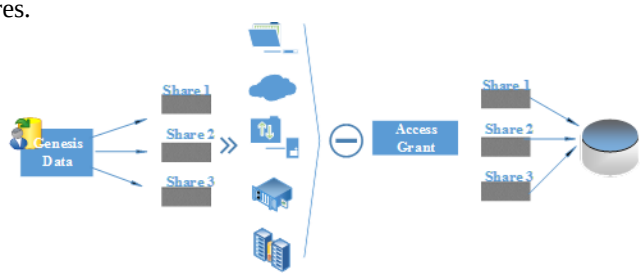

Fig. 5 Architecture of the proposed methodology for accessing secret shares in a database.

This is basically provided to the user for entry into proposed provenance based application. Timestamp data is collected with regards to users entry into the system during a particular period. Corresponding customer Id is also captured here. This captured information will be useful in cyber-attack investigations, troubleshooting, risk

management etc. [24-25]. Fig. 6 illustrates the 3-D Scatter plot representation of OTP, Operations and Timestamp. Operations referred here are as follows.

- Delete
- Alter
- Update and
- Add

One time password (OTP) values across time stamp of users login is also depicted in the graph.

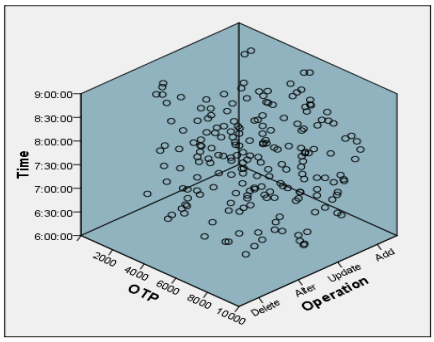

Fig. 6 3-D Scatter plot representation of OTP, Operations and Timestamp

The 3-D Scatter plot representation of OTP, Cust_Id and Reason descriptions are illustrated in Fig. 7. Reason parameters here include; workflow, additional records, data exchange and redundancy. These are the causes or actions performed by the user in accessing the database i..e., the genesis provenance data.

Scatter dot variations on Time parameter across Id of the customer with reference to operations performed by the user in the application are illustrated in Fig. 8

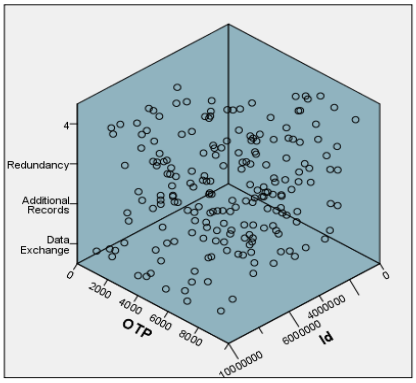

Fig. 7 3-D Scatter plot representation of OTP, Cust_Id and Reason descriptions

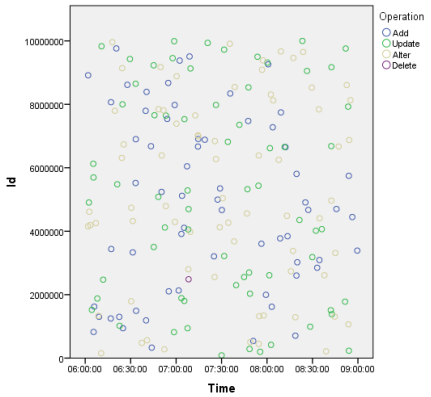

Fig. 8 Scatter dot variations on Time parameter across Id of the customer with reference to operations performed.

Scatter dot variations on Time parameter across Id of the customer with reference to reason mentioned for accessing application related to provenance data are illustrated in Fig. 9. Their variation can be observed in the above mentioned graphical representation. The recorded results helps in knowing the identity in accessing the system. It helps in troubleshooting issues in cyber-attack scenarios. Perhaps the tracking of this lineage corresponds to the definition of data provenance in the digital world. Graphical illustrations are simulated from SPSS statistical tool [26].

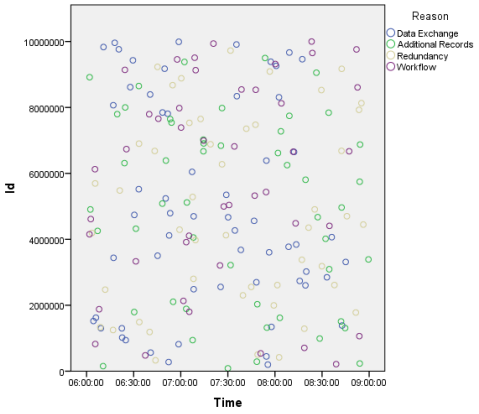

Fig. 9 Scatter dot variations on Time parameter across Id of the customer with reference to reason a user accessed the application.

4. Conclusion

As a consequence of experimentation on the suggested approach, simulations reveal improved outcomes in CIA parameters by preserving provenance with secret sharing. This is compliant with secret sharing for access to the genesis of a specific archive. Confidentiality will be preserved by the permitting of access by multiple entities and, unlike other cryptographic mechanisms, is a exclusive aspect of secret sharing. This work provides an interpretation of the concept of a secret data protection on provenance aspects. Literature is described for both the approaches in a comprehensive and extensive way. Provenance has several solicitations in various fields. Computer Information Technology is one such area which is closely related in various software implementations and data repositories. Need for appropriate security and relevant access control mechanisms is required for provenance data. In this line, an effort is made to unite provenance with a typical security mechanism. This uses secret sharing concept for accessing a controlled genesis data.

Throughout specific cases both the principles visual encryption and the data provenance are demonstrated. Literature review is an illustration of the need to coordinate and connect these two verticals. Adequate formulations are given for the mathematical model for secret sharing mechanism. Appropriate findings are provided in developing the problem statement with the concepts of corresponding application.

References

1. Gesbert, David J., et al. "Mode selection for data transmission in wireless communication channels based on statistical parameters." U.S. Patent No. 6,760,882. 6 Jul. 2004.
2. Walfish, Sheldon Israel. "Facilitating secure web browsing on untrusted networks." U.S. Patent No. 9,826,018. 21 Nov. 2017.
3. Jamil, Fuzel, et al. "Secure provenance using an authenticated data structure approach." Computers & Security 73 (2018): 34-56.
4. Chacko, Anu, and SD Madhu Kumar. "Big data provenance research directions." TENCON 2017-2017 IEEE Region 10 Conference. IEEE, 2017.
5. Tan, Yu Shyang. Reconstructing Data Provenance from Log Files. Diss. The University of Waikato, 2017.
6. Naor, Moni, and Adi Shamir. "Visual cryptography." Workshop on the Theory and Application of of Cryptographic Techniques. Springer, Berlin, Heidelberg, 1994.
7. Kaur, K., & Khemchandani, V. (2013, February). Securing Visual Cryptographic shares using Public Key Encryption. In Advance Computing Conference (IACC), 2013 IEEE 3rd International (pp. 1108-1113). IEEE.
8. Kong, J.H. ; Seng, K.P. ; Yeong, L.S. ; Ang, L.M. “Image compression with short-term visual encryption using the burrow wheeler transform and keyed transpose” Wireless Communications and Applications (ICWCA 2012), IET International Conference.
9. Cheney, J.; Lab. for Foundations of Comput. Sci., Univ. of Edinburgh, Edinburgh, UK “A Formal Framework for Provenance Security” Computer Security Foundations Symposium (CSF), 2011 IEEE.

10. Patrick McDaniel, Data Provenance and Security. IEEE Security & Privacy Magazine, 9(3), March/April, 2011.
11. Sultana, S.; Dept. of Electr. & Comput. Eng., Purdue Univ., West Lafayette, IN, USA ; Shehab, M. ; Bertino, E. "Secure Provenance Transmission for Streaming Data" Knowledge and Data Engineering, IEEE Transactions on (Volume:25, Issue: 8) Aug 2013.
12. Xinlei Wang; Dept. of Comput. Sci., Univ. of California, Davis, Davis, CA, USA ; Kai Zeng ; Govindan, K. ; Mohapatra, P. "Chaining for securing data provenance in distributed information networks" MILITARY COMMUNICATIONS CONFERENCE,2012-MILCOM2012.
13. Zhang, O.Q.; Ko, R.K.L. ; Kirchberg, M. ; Chun Hui Suen ; Jagadpramana, P. ; Bu Sung Lee "How to Track Your Data: Rule-Based Data Provenance Tracing Algorithms" Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference.
14. Liu, J. (2011). W7 Model of Provenance and its Use in the Context of Wikipedia.
15. Wu, Xiaotian, and Wei Sun. "Extended capabilities for XOR-based visual cryptography." IEEE Transactions on Information Forensics and Security. 9.10 2014: 1592-1605.
16. Lin, K. T. (2012, July). Based on Binary Encoding Methods and Visual Cryptography Schemes to Hide Data. In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on (pp. 59-62). IEEE.
17. Lee, Kai-Hui, and Pei-Ling Chiu. "Image size invariant visual cryptography for general access structures subject to display quality constraints." IEEE transactions on image processing. 22.10 2013: 3830-3841.
18. Daisy, V. Annie, C. Vijesh Joe, and S. Shinly Swarna Sugi. "An image based au-thentication technique using visual cryptography scheme." Inventive Systems and Control (ICISC), 2017 International Conference on. IEEE, 2017.
19. Huang, Y. J., & Chang, J. D. (2013, February). Non-expanded visual cryptography scheme with authentication. In Next-Generation Electronics (ISNE), 2013 IEEE International Symposium on (pp. 165-168). IEEE.
20. Lee, J. S., & Hsieh, M. H. (2013). Preserving user-participation for insecure network communications with CAPTCHA and visual secret sharing technique. IET networks, 2(2), 81-91
21. Alex, Nitty Sarah, and L. Jani Anbarasi. "Enhanced image secret sharing via er-ror diffusion in halftone visual cryptography." Electronics Computer Technology (ICECT), 2011 3rd International Conference on. Vol. 2. IEEE, 2011.
22. Alex, Nitty Sarah, and L. Jani Anbarasi. "Enhanced image secret sharing via er-ror diffusion in halftone visual cryptography." Electronics Computer Technology (ICECT), 2011 3rd International Conference on. Vol. 2. IEEE, 2011.
23. Daniel, Joshua, and Gery Ducatel. "Data acess authenication." U.S. Patent Application No. 16/292,683.
24. Dimitriadis, Athanasios, et al. "D4I-Digital forensics framework for reviewing and investigating cyber attacks." Array 5 (2020): 100015.
25. Dimitriadis, Athanasios, et al. "D4I-Digital Forensics Framework for Investigating Cyber Attacks in Industrie 4.0." (2019).
26. Wagner III, William E. Using IBM® SPSS® statistics for research methods and social science statistics. Sage Publications, 2019.