Scientific Journal of Impact Factor(SJIF): 3.134
e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 1, January -2015

A Unified Framework for Security and Storage of Information

Debajit Sensarma [1], Samar Sen Sarma [2]

[1] *Department of Computer Science & Engineering, University of Calcutta, Kolkata, West Bengal, INDIA, debajit.sensarma2008@gmail.com*

[2] *Department of Computer Science & Engineering, University of Calcutta, Kolkata, West Bengal, INDIA, sssarma2001@yahoo.com*

Abstract —*Information security and Information storage are two important research aspects now-a-days. While information security plays an important role in protecting the data and assets of an organization, there is a need for efficient storage of information to manage and process large graphs like web, social network, RDF graph, protein network and many more. In this paper, we investigate some algorithms to cope with the above two important problems in brief using a unified framework called graph because graph algorithm provides unified solution approach to many classical and modern application areas by taking graph as an omnipotent mathematical tool. Along with this some future scope have been discussed, which is inferred after studying the existing algorithms.*

Keywords- *Graph; Automorphism; Cryptography; Software Watermarking; Anonymization*

I. INTRODUCTION

Information (meaningful data) is a sequence of symbols that can be uniquely recognized where the symbol themselves have no possible meanings. While dealing with information two situations occur namely transmission of information from here to there and storage of information [1]. Information has different meanings in different contexts (e.g. Mathematical Information, Semantic Information, Physical Information, Biological Information, Economic Information etc.) [2].To protect the information from unauthorized access security of information is needed and besides the security of information there is a need for storing them efficiently for retrieving them without any corruption. In this paper information security and storage is defined briefly and some existing works related to above field using graph algorithms is given in a nutshell. We are interested in graph algorithms because it can be treated as unified solution approach in many classical and modern application areas and many real world problems can be modeled using graph and solution to them can be achieved by taking graph as a mathematical tool such that solving the later problem can give a suitable solution to the former one.

The paper is organized as follows: Section II of this paper describes information Security goals and techniques with cryptography, software watermarking, data annonymization along with some existing survey works. Section III gives the brief study about storage of information and some brief survey of existing works. Concluding remarks and future works are given in Section IV.

II. INFORMATION SECURITY

Information security is the practice of defending information from unauthorized access, use, disclosure, disruption, modification etc. From the characteristic of information, the value of the information comes in. When the characteristics of information changes, the value of the information either increases or decreases depending on the circumstances, e.g. timeliness of information can be a critical factor, because information loses much or it loses its value when it is delivered too late. According to [3] there are some characteristics of information, they are Availability, Accuracy, Authenticity, Confidentiality, Integrity, Utility, and Possession.

Security of information can be achieved using cryptography, software watermarking, Annoymization techniques etc.

A. Cryptography

Cryptography is the art of codifying messages, so that they become unreadable to provide information security [4]. The word cryptography comes from the Greek word "Kryptos", that means hidden, and "graphikos" which means writing. It is the transformation of readable and understandable data into a form which cannot be understood in order to secure the data [5].

a) Cryptography Goals.

By using cryptography many goals can be achieved, these goals can be either all achieved at the same time in one application, or only one of them, these goals are:

- **Confidentiality:** The principle of confidentiality specifies that nobody can understand the received message except the sender and one who has the decipher key.
- **Authentication:** It is the mechanism that is used to prove the identity of the communicating entity. With this process user or the system can prove their own identities to other parties who don't have personal knowledge of their identities.
- **Data Integrity:** This mechanism ensures that the received message has not been altered in any way from its original form. This can be achieved by using hashing at both sides the sender and the recipient in order to create a unique message digest and compare it with the one that received.
- **Non-Repudiation:** It is mechanism which is used to prove that the sender really sent this message and the message was received by the specified party, so the recipient cannot claim that the message was not sent.
- **Access Control:** The principle of access control specifies that who should be able to access resources. The goal is that, if one can access the resource, under which restrictions and conditions the access can be occurred, and what is the permission level of a given access.

There are two basic types of cryptography: Symmetric Key and Asymmetric Key. Symmetric key algorithms are the quickest and most commonly used type of encryption. In Symmetric key cryptography, a single key is used for both encryption and decryption. There are few well-known symmetric key algorithms i.e. DES, RC2, RC4, IDEA etc. Public or asymmetric key cryptography involves the use of key pairs: one private key and one public key. Both are required to encrypt and decrypt a message or transmission. Example- RSA, Elliptic curve cryptography (ECC), ElGamal encryption system etc.

Graph theory provides problems with large computational complexity, e.g. the search for shortest (weighted) Hamiltonian cycle is an intractable problem. Besides this, various structures may be created on a given graph like the number of paths, number of cut-sets or number of spanning trees grows much faster than exponentially. This can be used in the design of efficient and secure ciphers which may be more secure than the ciphers in the past [6].

b) Existing Works

According to [6] graphs may be used for the design of stream ciphers, block ciphers or public-key ciphers. This method used paths between a pair of graph vertices for designing effective poly-alphabetic substitution ciphers, where the period of alphabet changeovers is depends on the number of paths between a pair of selected vertices on the graph, the number of spanning trees or is a multiple of these values. The authors of [7] also patented an encryption method based on charting path on a graph by using the graph as the encryption key. The plain text expressed that path by the sequence of vertices and cipher text expresses the same using sequence of edges between the vertices. In this method many plain text of choices are mapped to single cipher text, so it is very difficult to match a given cipher text with a plain text of choice, thus insuring that the cipher points to the true plain text. In [5] the idea to use families of simple graphs of large girth in Cryptography had been explored. The general idea in is to consider the set of vertices as the plain space and a path in the graph as an encryption tool such as password. Next, the absence of short cycles ensures that different passwords convert chosen plaintext to different cipher texts. The secret sharing allows splitting a secret into different pieces, called shares, which are given to the participants. Only certain group (the authorized set of participants) can recover the secret. There exits various graph based secret sharing schemes. The author of [8] proposed a visual cryptography scheme that allows the sharing of multiple secret images on graphs. In this method every node and edges of an arbitrary graph are assigned an arbitrary image. Images on the vertices are "public" and images on the edges are "secret". The problem is to make such a construction that when encoded images of two adjacent vertices are overlapped the secret image of the corresponding edge is unfolded. The authors of [9] have designed a secure cryptographic protocol taking advantages of two intractable problems namely Hamiltonian Path Problem [10] and Graph Automorphism problem [11]. The algorithm is constrained based partial symmetric key algorithm. Classical Data Encryption Standard (DES) algorithm coupled with two above mentioned intractable problems are used to design the secure algorithm (GMDES). We have worked with a 4-cube graph and each vertex of that graph is encoded with 4 bit binary number. An arbitrary Hamiltonian Path of that graph represented by a sequence of binary symbols is used as the secret key and its 16 non-automorphic Hamiltonian Paths chosen randomly are used as the sub keys. The sub keys are stored in a secure mapping table of sender in an encrypted form. Here a secure key exchange protocol [12] which is based on two keys (partial key and secret key) and a secret table only known to sender and receiver are used. The algorithm provides integrity, authenticity and non repudiation when transferring the message and public key. The key feature of the algorithm is until the receiver knows the sub keys provided he/she knows the secret key, the probability of decrypting text in a finite amount of time is very less. The efficiency of proposed algorithm surpasses the classical Data Encryption Standard (DES) algorithm in general. Different methods of graph based cryptography protocols are listed in [13-22].

B. Software Watermarking

In the world of internet, multimedia communication becomes very easy, efficient and cost effective. Digital multimedia can be easily tampered and manipulated. Digital watermarks have been proposed as a means for copyright protection for multimedia data.
Generally, watermarking systems for digital media involve two distinct stages: (1) watermark embedding to indicate copyright and (2) watermark detection to identify the owner [23].
There are three main requirements of digital watermarking. They are transparency, robustness, and capacity [24].

- **Transparency:** Transparency or fidelity is perceptual similarity between the original and the watermarked versions of the cover work. The digital watermark should not affect the quality of the original software after it is watermarked.
- **Robustness:** It ensures that the watermark is difficult for an attacker to remove.

- **Capacity:** Capacity or data payload is the number of bits a watermark encodes within a unit of time or work. This property describes how much data should be embedded as a watermark to successfully detect during extraction.

Figure.1 gives the classification of software watermarking pictorially [24].

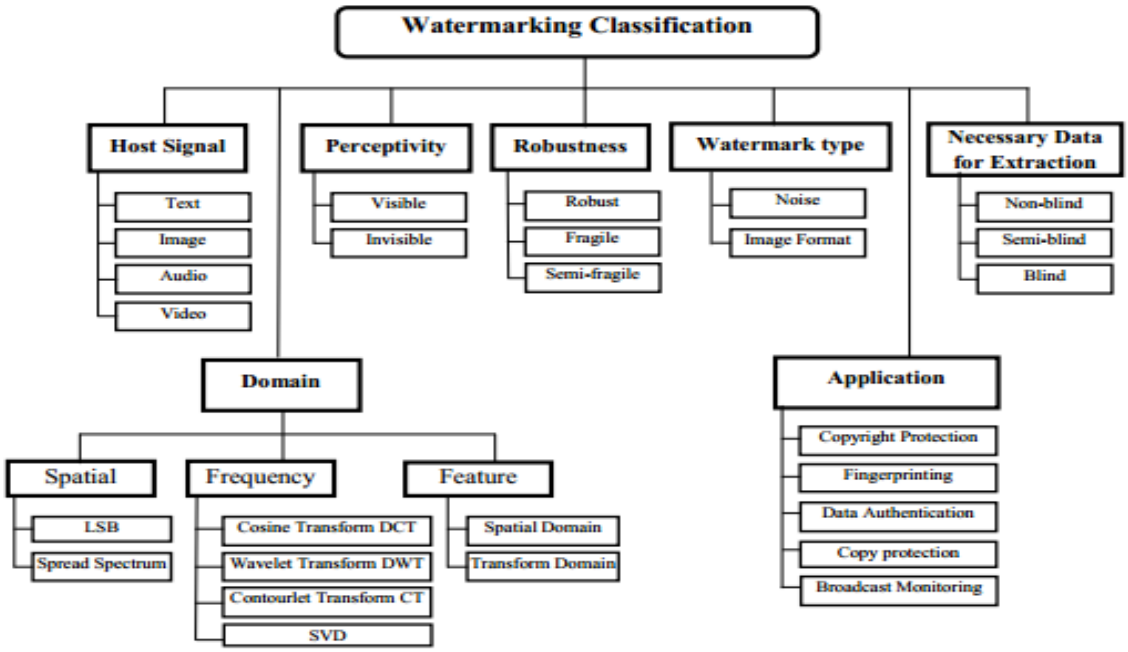

Figure.1. Classification of Software Watermarking

a) Existing Works

Graph theoretic approach can be used to design software watermark in robust fashion. The method in [25] works with control/data flow graphs and uses a random walk method to embed the watermark. The algorithm in [26] namely QP algorithm, is originally developed to embed a signature into a solution of a graph coloring problem. But it is available for watermarking by register allocation. Unfortunately, QP algorithm has a drawback that it may fail to recognize the embedded messages. Authors proposed QPS algorithm [27] that overcomes the weak point of QP algorithm. QPS algorithm ensures the complete recovery of the embedded messages. However, in QPS algorithm, the lengths of the embeddable messages are extremely shortened. Next in [28] the authors proposed two new software watermarking algorithms, Color Change (CC) and Color Permutation (CP) that are based on register allocation. The performance is measured with benchmark applications where CC and CP showed better data-rate than the previous algorithms QP and QPS. In [29] the authors propose a novel Internet based IP protection scheme. The input to the proposed scheme is a generic graph corresponding to a digital system design. Watermarking of the graph and its encryption are achieved using a new linear feedback shift register (LFSR)-based locking scheme. The proposed scheme makes unauthorized disclosure of valuable designs almost infeasible, and can easily detect any alteration of the design file during transmission. It ensures authentication of the original designer as well as non-repudiation between the seller and the buyer. Besides this, more algorithms are listed in [30-33].

C. Data Anonymization

Data anonymization is the process of anonymizing data which describe people, by encrypting or removing personally identifiable information from data sets. The Privacy Technology Focus Group defines it as "technology that converts clear text data into a nonhuman readable and irreversible form and encryption techniques in which the decryption key has been discarded." Data anonymization enables the transfer of information across a boundary, such as between two departments within an agency or between two agencies, while reducing the risk of unintended disclosure [34].

a) Existing Works

The structure of the graph itself and the degree of the nodes can reveal the identities of individuals. To address this issue, authors in [35] study a specific graph-anonymization problem. They called a graph k-degree anonymous if for every node "n", there exist at least k-1 other nodes in the graph with the same degree as "n". This definition of anonymity prevents the re-identification of individuals by adversaries with a priori knowledge of the degree of certain nodes. Next, the growing popularity of social networks has generated interesting data management and data mining problems. An important concern in the release of these data for study is their privacy, since social networks usually contain personal information. Simply removing all identifiable personal information (such as names and social security number) before releasing the data is insufficient. In [36] authors propose k-automorphism to protect against multiple structural attacks and develop an algorithm (called KM) that ensures k-automorphism. In [37] k-isomorphism based social network privacy preserving scheme is proposed by anonymization of network and forming "k" pair wise isomorphic sub-graphs to protect structural attack. They have shown that the problem is NP-hard and authors devise a number of techniques to enhance the anonymization efficiency while retaining the data utility. This gives a satisfactory performance on a number of real datasets. Protecting sensitive relationships among the individuals in the anonymized social networks is considered in [38]. This is closely related to the link prediction problem that has been widely studied in the link mining community. The work in [39] considers the identity disclosure problem in weighted graph and discusses a class of important background knowledge attacks with weight-related properties. A general model for weight anonymization is provided with a concrete case, called k-volume anonymization as a proof of concept. Experimental results shows the proposed approaches give good performances on both privacy and utility for synthetic and real-world weighted graphs. More works can be found in [40-43].

III. STORAGE OF INFORMATION

The information must be stored in an efficient way. To achieve the goal here a brief study on efficient database storage and database replication policy is considered.

- Maintaining large interconnected data with the popularity of web and more recently the widespread use of social networks, the need to process and find information in very large graphs impose several challenges; how to process such large graphs efficiently, since they probably do not fit into main memory. So, performing any type of computations or storing graphs for archival is a challenging task. One of the solutions to manage large interconnected information is to represent the information with a unified framework called graphs and use graph compression techniques that supports basic navigation queries directly over the compressed structure, without the need of decompression. This allows simulating any graph algorithm in main memory using much less space than a plain representation.

- Graphs are the unified framework for constructing both natural and human made structures. Many practical problems and real world applications such as physical, biological and social systems can be modeled using graphs. With the appearance of these applications, developments of graph databases are very useful to store graph data. Due to the existence of noise (e.g., duplicated graphs) in the graph database, we investigate the problem of storing the same graphs in the single graph database. Therefore, detecting and eliminating of automorphic graphs (i.e. duplicate graphs) in a graph database become an important research area.

- Data replication is very necessary for achieving scalability, fault tolerance, reliability and high availability of data in the database. Data objects are copied in different servers so that whenever one of the database servers fails, if it is too overloaded or geographically too far away from the requesting user, a data copy can be retrieved from one of the other servers.

  The data replication problem [44] can be formulated as the special case of Bin Packing Problem with Conflict (BPCC) in the following sense:
  - n servers correspond to n bins.
  - bins have maximum capacity C.

- k data fragments correspond to k objects
- each object has a weight (a capacity consumption) $c_i \leq C$.

The goal is to place the objects in the minimum number of bins without exceeding the maximum capacity with no two replicas placed in the same bin.

A. Existing Works

In [45] authors exploit symmetry information by making use of the overlap in neighbors and analyzing how information is reduced by shrinking the network. This method used a greedy algorithm to determine the orbit of symmetry identifications, to achieve compression. The authors of the paper [46] proposed a novel DAGC algorithm to identify and removal of automorphic graph storing into the graph database using AdE index structure. AdE index structure incorporates graph structural information of each graph in the database. In [47] authors try to find redundant graphs in chemical graph databases by finding symmetric graphs which have more than one vertex with the same label. The authors [48] have proposed a one dimensional Bin Packing Problem with a compatibility graph where objects are represented with vertices and there is an edge between two vertices iff the objects can be placed in the same bin. This algorithm can be used to solve data replication problem. More related works can be found in [49, 50].

IV. CONCLUSION & FUTURE WORKS

In this paper, we present a brief survey of the current research related to the information security techniques and information storage techniques where graph is used as a unified framework. From the above discussion regarding the existing and recent works, we can conclude that although, many techniques have been proposed, still, there are weaknesses in the above mentioned schemes, e.g. there are many works that focuses on private key cryptography and secret sharing schemes but few works exists on graph based public key cryptography which may be more secure than above two methods. In software watermarking schemes for Intellectual Property (IP) protection, some graph based algorithms produce same watermarked graph with the different messages or signatures. This violates the uniqueness of the watermark. Again some of the work depends on private key watermark having a disadvantage that, if adversary can somehow know the key, he can extract the watermark. These types of shortcomings can be tackled by unique representation of the watermark graph and using public key concept. Next, graph anonymization techniques can be applied in the field of cryptography, steganography and privacy or security of networks as from a given anonymized graph it is very difficult to generate the original graph and last but not the least, the size of the network so as the graph is growing day by day. So, there is a need to design an efficient graph compression algorithm to store and manage the large graphs. Furthermore, future depends on more detailed research on these topics.

ACKNOWLEDGMENT

The authors would like to thank University Of Calcutta, West Bengal, India, Department of Science & Technology (DST), New Delhi, for financial support and the reviewers for their constructive and helpful comments and specially the Computer without which no work was possible.

REFERENCES

[1] Hamming, R. W. (1986), Coding and information theory. Prentice-Hall, Inc.
[2] Floridi, L. (2010). Information: A very short introduction. Oxford University Press.
[3] Whitman, M., & Mattord, H. (2011). Principles of information security. Cengage Learning.
[4] Kahate, A. (2013). Cryptography and network security. Tata McGraw-Hill Education.
[5] Ustimenko, V. (2001). CRYPTIM: Graphs as tools for symmetric encryption. In Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (pp. 278-286). Springer Berlin Heidelberg.
[6] Paszkiewicz, A., Górska, A., Górski, K., Kotulski, Z., Kulesza, K., & Szczepański, J. (2001). Proposals of Graph Based Ciphers, Theory and Implementations. In NATO Regional Conference and Information System 2001, Proc. Vol (Vol. 2, pp. 168-177).
[7] S. Gideon, Denial Cryptography based on Graph Theory, US patent 6823068-2004 http://www.patentstorm.us/patents/6823068.html.
[8] Lu, S., Manchala, D., & Ostrovsky, R. (2008). Visual cryptography on graphs. In Computing and Combinatorics (pp. 225-234). Springer Berlin Heidelberg.
[9] Sensarma, D., and Sen Sarma, S., (2014). GMDES: A GRAPH BASED MODIFIED DATA ENCRYPTION STANDARD ALGORITHM WITH ENHANCED SECURITY. International Journal of Research in Engineering and Technology, eISSN: 2319-1163, Vol. 3, Issue. 3, PP. 653-660.

[10] Garey, Michael R., and David S. Johnson. (1979), Computers and intractability. Vol. 174. San Francisco: freeman.
[11] Biggs, N. (1971), Finite groups of automorphisms. course given at the University of Southampton, October-December 1969. Vol. 6. CUP Archive.
[12] Sensarma, D., Banerjee, S., and Basuli, K. (2012). A New Scheme for Key Exchange. International Journal of Modern Engineering Research (IJMER), ISSN (Online): 2249-6645, 2(3).
[13] Blundo, C., De Santis, A., Stinson, D. R., & Vaccaro, U. (1995). Graph decompositions and secret sharing schemes. Journal of Cryptology, 8(1), 39-64.
[14] Suneetha, C. H., & Kumar, D. S. (2011). A Block Cipher using Graph Structures and Logical XOR Operation. International Journal of Computer Applications, 33(7).
[15] Beimel, A., Malkin, T., Nissim, K., & Weinreb, E. (2007). How should we solve search problems privately?. In Advances in Cryptology-CRYPTO 2007 (pp. 31-49). Springer Berlin Heidelberg.
[16] Priyadarsini, P. L. K., & Ayyagari, R. (2013, August). Ciphers based on special graphs. In Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on (pp. 460-465). IEEE.
[17] Charles, D. X., Lauter, K. E., & Goren, E. Z. (2009). Cryptographic hash functions from expander graphs. Journal of Cryptology, 22(1), 93-113.
[18] Jin, S., & Peng, J. (2009, January). Key Graphs and Secret Sharing Be Used in Network Multicast Security. In Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on (pp. 1-5). IEEE.
[19] Kotorowicz, J. S., Ustimenko, V., & Romańczuk, U. (2011, September). On the implementation of stream ciphers based on a new family of algebraic graphs. In Proceedings of the Conference CANA, FedSCIS (pp. 485-490).
[20] WenBin, H., & Jenq-Shiou, L. (2011). Use of Hamiltonian Cycles in Cryptograph (No. arXiv: 1107.3172).
[21] Kulesza, K., & Kotulski, Z. (2003). Secret sharing for n-colorable graphs with application to public key cryptography. arXiv preprint cs/0310053.
[22] Sun, H. M., & Shieh, S. P. (1996). An efficient construction of perfect secret sharing schemes for graph-based structures. Computers & Mathematics with Applications, 31(7), 129-135.
[23] herekar, D. S., Thakare, D. V., Jain, D. S., Miss Ashwini, D. B., Tijare, P. A., Deshpande, M. S. A., & Singh, M. P. (2011). Attacks and countermeasures on digital watermarks: classification, implications, benchmarks. International Journal of Computer Science and Applications, 4(2).
[24] Hussein, E., & Belal, M. A. (2012, September). Digital Watermarking Techniques, Applications And Attacks Applied To Digital Media: A Survey. In International Journal of Engineering Research and Technology (Vol. 1, No. 7 (September-2012)). ESRSA Publications.
[25] Venkatesan, R., Vazirani, V., & Sinha, S. (2001, January). A graph theoretic approach to software watermarking. In Information Hiding (pp. 157-168). Springer Berlin Heidelberg.
[26] G. Qu and M. Potkonjak (1999). Hiding signatures in graph coloring solutions, In Information Hiding Workshop'99, pages 348-367.
[27] G. Myles and C. Collberg (2003). Software watermarking through register allocation: Implementation, analysis, and attacks, In Proceedings of the International Conference on Information Security and Cryptology, pages 274-293.
[28] Lee, H., & Kaneko, K. (2009, April). Two new algorithms for software watermarking by register allocation and their empirical evaluation. In Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on (pp. 217-222). IEEE.
[29] Halder, R., Dasgupta,P.,Naskar,S., and Sarma., S. S. (2011), An Internet-based IP Protection Scheme for Circuit Designs using Linear Feedback Shift Register-based Locking. Engineering Letters 19, no. 2: 84.
[30] Qu, G., & Potkonjak, M. (1998, November). Analysis of watermarking techniques for graph coloring problem. In Proceedings of the 1998 IEEE/ACM international conference on Computer-aided design (pp. 190-193). ACM.
[31] Collberg, C., Kobourov, S., Carter, E., & Thomborson, C. (2003, October). Error-correcting graphs for software watermarking. In Proceedings of the 29th Workshop on Graph Theoretic Concepts in Computer Science (pp. 156-167).
[32] Jiang, Z., Zhong, R., & Zheng, B. (2009, October). A software watermarking method based on Public-Key cryptography and graph coloring. In Genetic and Evolutionary Computing, 2009. WGEC'09. 3rd International Conference on (pp. 433-437). IEEE.
[33] Saha, D., Dasgupta, P., Sur-Kolay, S., & Sen-Sarma, S. (2007, March). A novel scheme for encoding and watermark embedding in VLSI physical design for IP protection. In Computing: Theory and Applications, 2007. ICCTA'07. International Conference on (pp. 111-116). IEEE.
[34] Privacy Technology Focus Group Report. United States Department of Justice. 2006. p. 52.
[35] Liu, K., and Terzi, E. (2008), Towards identity anonymization on graphs. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 93-106. ACM.
[36] Zou, L., Chen, L., & Özsu, M. T. (2009), K-automorphism: A general framework for privacy preserving network publication. Proceedings of the VLDB Endowment, 2(1), 946-957.
[37] Cheng, J., Fu, A. W. C., & Liu, J. (2010, June). K-isomorphism: privacy preserving network publication against structural attacks. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 459-470). ACM.

[38] Zheleva, E., & Getoor, L. (2008). Preserving the privacy of sensitive relationships in graph data. In Privacy, security, and trust in KDD (pp. 153-171). Springer Berlin Heidelberg.
[39] Li, Y., & Shen, H. (2010, December). On identity disclosure in weighted graphs. In Parallel and Distributed Computing, Applications and Technologies (PDCAT), 2010 International Conference on (pp. 166-174). IEEE.
[40] Gionis, A., & Tassa, T. (2009). k-Anonymization with minimal loss of information. Knowledge and Data Engineering, IEEE Transactions on, 21(2), 206-219.
[41] Li, Y., & Shen, H. (2010, December). Anonymizing graphs against weight-based attacks. In Data Mining Workshops (ICDMW), 2010 IEEE International Conference on (pp. 491-498). IEEE.
[42] Lu, X., Song, Y., & Bressan, S. (2012, January). Fast identity anonymization on graphs. In Database and Expert Systems Applications (pp. 281-295). Springer Berlin Heidelberg.
[43] Zheleva, E., & Getoor, L. (2011). Privacy in social networks: A survey. In Social network data analytics (pp. 277-306). Springer US.
[44] Hussain, A., & Khan, M. N. A. (2014). Discovering Database Replication Techniques in RDBMS. International Journal of Database Theory & Application, 7(1).
[45] Sun, J., Bollt, E. M., & Ben-Avraham, D. (2008). Graph compression—save information by exploiting redundancy. Journal of Statistical Mechanics: Theory and Experiment, 2008(06), P06001.
[46] Thaing, A. N., & Oo, K. M. (2011). DAGC: Identification and Filtration of Automorphic Graphs in Graph Databases. International Journal of Computer Applications, 30(4).
[47] Vijayalakshmi, R., Nadarajan, R., Nirmala, P., & Thilaga, M. (2010). A Novel Approach for Detection and Elimination of Automorphic Graphs in Graph Databases. Int. J. Open Problems Compt. Math, 3(1).
[48] Sensarma, D., & Sarma, S. S. (2014). A NOVEL GRAPH BASED ALGORITHM FOR ONE DIMENSIONAL BIN PACKING PROBLEM. Journal of Global Research in Computer Science, 5(8), 1-4.
[49] Adler, M., & Mitzenmacher, M. (2001). Towards compressing web graphs. In Data Compression Conference, 2001. Proceedings. DCC 2001. (pp. 203-212). IEEE.
[50] Wang, H. (2010). Managing and mining graph data (Vol. 40). C. C. Aggarwal (Ed.). New York: Springer.

