significant to be acceptable. This watermarking technique is applied to binned medical data in a hierarchical manner [Bertino et al., 2005].

4.3 Watermarking Based on Non-Numeric Multi-Word Attributes

Ali and Ashraf [Al-Haj and Odeh, 2008] propose a watermarking scheme which is based on hiding binary image in spaces of non-numeric multi-word attributes of subsets of tuples, instead of numeric attribute at bit-level. The watermark is divided into *m* string each containing *n* bits. On the other hand, the database is also divided into non-intersecting subsets each containing *m* tuples. The *m* short strings of the watermark image are embedded into each *m*-tuple subset.

The embedding is done as follows: suppose the integer representation of the *ith*, $i \in [1 \ldots m]$, short string is *di*. A double space is created after *di* words of the pre-selected nonnumeric, multi-word attribute of *ith* tuple in the subset. The *Halder R., Pal S., Cortesi A.: Watermarking Techniques* ... 3177 extraction phase counts the number single spaces appearing before double space which indicates the decimal equivalent of the embedded short binary string. Since the proposed algorithm embeds the same watermark for all non-intersecting subsets of the database, it is robust against subset deletion, subset addition, subset alteration and subset selection attacks. Another advantage for space based watermarking is that large bit-capacity available for hiding the watermark which may also facilitate embedding of multiple small watermarks. However, it may suffer from watermark removal attack if Mallory replaces all double spaces between two words (if exist) by single space for all tuples in the relation.

4.4 Watermarking Based on Tuple or Attribute Insertion

(a) Fake tuples as watermark information.

The approach in [Pournaghshband, 2008] aims to generate fake tuples and insert them erroneously into the database. The fake tuple creation algorithm takes care of candidate key attributes and sensitivity level of non candidate attributes. He uses Bernoulli sampling probability *pi* for the *ith* non-candidate attribute *Ai* to decide its fake value which may be chosen uniformly or as the value with higher occurrence frequency in the existing set of values of *Ai* in the relation. Unlike other algorithms, the detection algorithm is not an inverse algorithm to the watermark generating algorithm and insertion algorithm is probabilistic in nature. Detection algorithm checks to see whether the fake tuples inserted during watermark insertion phase, exist or has been changed. It checks it via primary key.

As soon as it finds one match (i.e. identical or similar tuples), detection is done. The detection will fail for the watermarked database when all of the fake tuples are deleted by benign deletions. The number of fake tuples to be inserted is decided by the database owner. However, the watermark insertion phase must take into account the fact that the values of the fake tuples marks should not by any means degrade the quality of the data in the database and should not impact the query results. One advantage of this scheme is that the ownership can be publicly verified more than once until all the fake tuples are revealed and the scheme does not suffer from incremental updatability.

(b) Virtual attribute as watermark information.

Rather than inserting fake tuples, the author in [Prasannakumari, 2009] proposes another watermarking technique by inserting a virtual attribute in the relation which will serve as watermark containing parity checksum of all other attributes and an aggregate value obtained from any one of the numeric attribute of all tuples. The process of virtual attribute insertion is performed independently for each non-overlapping partitions obtained from the original relation.

This scheme is designed to authenticate the tamper-proof receipt of the database over an insecure communication channel. Although this approach is fragile and 3178 *Halder R., Pal S., Cortesi A.: Watermarking Techniques* ... can easily detect any of the deletion or insertion or alter attacks, it suffers from the watermark removal attack.

5. Conclusion

This paper surveys the current state-of-the-art of different watermarking techniques for relational databases and classifies all the techniques based on (*i*) whether the technique introduces the distortion to underlying data, (*ii*) the type of the cover where mark is embedded, and (*iii*) the type of the Watermark information. Most of the distortion-based watermarking techniques mainly aim at protecting the ownership, whereas distortion-free watermarking techniques mostly are fragile and aim at maintaining integrity of the database information. Although we classify the schemes based on different watermark information, most of the numerical distortion-based schemes follow almost similar steps to identify the candidate bit positions for the watermark. Finally, it has been observed that the usability of the watermarked database and queries still remains an open issue for future research.

References

[Abdel-Hamid et al., 2004] Abdel-Hamid, A. T., Tahar, S., and Aboulhamid, E. M. (2004). A survey on ip watermarking techniques. *Design Automation for Embedded Systems*, 9(3):211–227.

[Agrawal et al., 2003a] Agrawal, R., Haas, P. J., and Kiernan, J. (2003a). A system for watermarking relational databases. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data (SIGMOD '03)*, pages 674–674, San Diego, California. ACM Press.

[Agrawal et al., 2003b] Agrawal, R., Haas, P. J., and Kiernan, J. (2003b). Watermarking relational data: framework, algorithms and analysis. *The VLDB Journal*, 12:157–169.

[Agrawal and Kiernan, 2002] Agrawal, R. and Kiernan, J. (2002). Watermarking relational databases. In *Proceedings of the 28th international conference on Very Large Data Bases (VLDB '02)*, pages 155–166, Hong Kong, China. VLDB Endowment.

[Agrawal and Srikant, 2000] Agrawal, R. and Srikant, R. (2000). Privacy-preserving data mining. *ACM SIGMOD Record*, 29(2):439–450.

[Al-Haj and Odeh, 2008] Al-Haj, A. and Odeh, A. (2008). Robust and blind watermarking of relational database systems. *Journal of Computer Science*, 4:1024–1029.

[Bertino et al., 2005] Bertino, E., Ooi, B. C., Yang, Y., and Deng, R. H. (2005). Privacy and ownership preserving of outsourced medical data. In *Proceedings of the 21st International Conference on Data Engineering (ICDE '05)*, pages 521–532, Tokyo, Japan. IEEE Computer Society.

[Cui et al., 2006] Cui, X., Qin, X., Sheng, G., and Zheng, J. (2006). A robust algorithm for watermark numeric relational databases. In *Proceedings of the 2010 International conference on Intelligent computing (ICIC '06)*, pages 810–815, Kunming, China. Springer Lecture Notes in Control and Information Sciences.

[Guo et al., 2006a] Guo, F., Wang, J., and Li, D. (2006a). Fingerprinting relational databases. In *Proceedings of the 2006 ACM symposium on Applied computing (SAC '06)*, pages 487–492, Dijon, France. ACM Press.

[Guo et al., 2005] Guo, F.,Wang, J., Zhang, Z., Ye, X., and Li, D. (2005). An improved algorithm to watermark numeric relational data. In *Proceedings of the 6th International Workshop on Information Security applications (WISA '05)*, pages 138–149, Jeju Island, Korea. Springer LNCS, Volume 3786.

[Guo et al., 2006b] Guo, H., Li, Y., Liua, A., and Jajodia, S. (2006b). A fragile watermarking scheme for detecting malicious modifications of database relations. *Information Sciences*, 176:1350–1378.

[Gupta and Pieprzyk, 2009] Gupta, G. and Pieprzyk, J. (2009). Database relation watermarking resilient against secondary watermarking attacks. In *Proceedings of the 5th International Conference on Information Systems Security (ICISS '09)*, pages 222–236, Kolkata, India. Springer LNCS, Volume 5905.

[Hacigumus et al., 2002] Hacigumus, H., Iyer, B., and Mehrotra, S. (2002). Providing database as a service. In *Proceedings of the 18th International Conference on Data Engineering (ICDE '02)*, pages 29–38, San Jose, California, USA. IEEE Computer Society.

[Halder and Cortesi, 2010a] Halder, R. and Cortesi, A. (2010a). A persistent public watermarking of relational databases. In *Proceedings of the 6th International Conference on Information Systems Security (ICISS '10)*, pages 216–230, Gandhinagar, Gujarat, India. Springer LNCS, Volume 6503.

[Halder and Cortesi, 2010b] Halder, R. and Cortesi, A. (2010b). Persistent watermarking of relational databases. In *Proceedings of the IEEE International Conference on Advances in Communication, Network, and Computing (CNC '10)*, pages 46–52,

Calicut, Kerala, India. IEEE Computer Society. [Halder et al., 2009] Halder, R., Dasgupta, P., Naskar, S., and Sarma, S. S. (2009). An internet-based ip protection scheme for circuit designs using linear feedback shift register (lfsr)-based locking. In *Proceedings of the 22nd Annual Symposium on Integrated Circuits and System Design(SBCCI '09)*, Natal, Brazil. ACM Press.

[Hu and Grefen, 2002] Hu, J. and Grefen, P. (2002). Component based system framework for dynamic b2b interaction. In *Proceedings of the 26th International Computer Software and Applications Conference on Prolonging Software Life: Development and Redevelopment (COMPSAC '02)*, pages 557–562, Oxford, England. IEEE Computer Society.

[Hu et al., 2005] Hu, T., Chen, G., Chen, K., and Dong, J. (2005). Garwm: Towards a generalized and adaptive watermark scheme for relational data. In *Proceedings of the 6th International Conference in Advances in Web-Age Information Management (WAIM '05)*, pages 380–391, Hangzhou, China. Springer LNCS, Volume 3739.

[Hu et al., 2009] Hu, Z., Cao, Z., and Sun, J. (2009). An image based algorithm for watermarking relational databases. In *Proceedings of the 2009 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '09)*, pages 425–428, Zhangjiajie, Hunan, China. IEEE Computer Society. 3188 *Halder R., Pal S., Cortesi A.: Watermarking Techniques ...*

[Huang et al., 2009] Huang, K., Yue, M., Chen, P., He, Y., and Chen, X. (2009). A cluster-based watermarking technique for relational database. In *Proceedings of the 1st International Workshop on Database Technology and Applications (DBTA '09)*, pages 107–110, Wuhan, China. IEEE Press.

[Huang et al., 2004] Huang, M., Cao, J., Peng, Z., and Fang, Y. (2004). A new watermark mechanism for relational data. In *Proceedings of the 4th International Conference on Computer and Information Technology (CIT '04)*, pages 946–950, Wuhan, China. IEEE Computer Society.