ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

A Survey on Robust security mechanism for NoSQL databases

Amreen, Dadapeer

M. Tech Student, Dept. of Computer Science and Engineering , Ballari Institute of Technology and Management(BITM) College, Bellary, Karnataka, India.

Assistant Professor, Dept. of Computer Science and Engineering , Ballari Institute of Technology and Management(BITM) College, Bellary, Karnataka, India.

ABSTRACT: NoSQL databases are a very important for IT application to store the large amount of data. NoSQL wonder has taken the database and IT application world by tempest. Growth and penetration of NoSQL application has created an unprecedented database revolution. Many companies are migrating their IT databases to NoSQL databases. There are many security solutions available for relational databases. As many companies are migrating their data to NoSQL data there should be security mechanism for NoSQL databases. The paper attempts to provide the reversible watermarking algorithm to provide security mechanisms for NoSQL databases.

As reversible watermarking algorithm is already been discussed for relational databases but here we are using reversible watermarking technique to provide security for NoSQL databases in terms of ownership protection and tamer proofing.

KEYWORDS: NoSQL database; Security; Reversible watermarking; relational databases; Ownership protection; tamper proofing.

I. Introduction

NoSQL databases are created for big data processing such as web application platform which consist of the web pages and development of social media websites. Big data are related to all aspects of human activity design, digital services, recording events to research etc. Some of the NoSQL databases are MongoDB, Hadoop-Hbases and CouchDB. It is given in [3] that hadoop is an open source framework through implementation of mapreduce and HDFS for handling big data.

Some of the attributes of NoSql application to which the security is important are a)social websites stores sensitive and personal data which demands high level of security b)NoSQL databases like MongoDB had a previous instances of cyber attacks c)more usage of a distributed data like Hadoop. There for providing the security for NoSQL databases will help the readers in assessing security maturity of NoSQL databases. Here reversible watermarking technique is utilized to guarantee the security for enormous information which is put away in the NoSQL databases. The security is given as far as ownership protection and sealing for wide tamper proofing data formats. This type of watermarking tries to identify the source of the data leakage by tracing a guilty agent. Watermarking has the property that it can provide ownership protection over the digital content by making the data with watermark unique to the owner that is watermark is added to the user data if there is any violation then we can easily identify that our data is corrupted.

II. Related work

Irreversible watermarking technique was first used to secure relational databases which is proposed by Agrawal et al. in [1].The existing system is vulnerable to Data modification attack, Data injection attack**,** Authentication attack, these are some of the holes in the system. The first reversible watermarking scheme for

Copyright to IJIRCCE DOI: 10.15680/IJIRCCE.2016. 0404265

relational databases was proposed in which histogram expansion is used for reversible watermarking for relational databases. However these techniques are not robust against heavy attacks.

In the existing system before embedding watermark information and after extracting the resultant quality of watermark information, is not acceptable for knowledge extraction process and there may be significant loss of watermarked information if there is no appropriate watermark bandwidth that ensures maximum watermark robustness. In the above technique authentication of data quality is done to keep track of the overhead of information. However, this technique is not robust against heavy attacks.

Prediction error expansion watermarking techniques [2], [3] are used for malicious attacks where water mark is embedded only in fractional parts of numerical features only. Here intension of the assailant is to protect the helpfulness of the information; otherwise he can easily compromise the data.

Difference expansion watermarking technique uses a genetic algorithm is used in a proposed robust and reversible solution for relational databases [11]. The above algorithm drawbacks mentioned can be overcome minimizing distortions in the data, increasing watermark capacity and lowering the false positive rate. There for this watermark technique is used to increase the watermark capacity and minimize the distortion. Difference expansion watermarking techniques (DEW), [14], [16] perform transformation and exploit the methods of arithmetic transformation on numeric features. The watermark information is normally in the LSB of features of NoSQL databases which minimizes distortions. But here in reversible watermarking technique value is embedded in the selected features of the dataset by minimizing the distortion to preserve the data quality. Another technique [15] based on support vector regression and difference expansion to protect the databases from being tampered. These all techniques are used to provide the ownership protection. These techniques fail to detect any changes in the data and are vulnerable to modification attacks.

III. PROPOSED ALGORITHM

Reversible watermarking techniques is used to ensure security in terms of ownership protection and tamper proofing for wide range of data formats. This reversible watermarking technique is already used to provide the security for relational databases. But here we are using this reversible watermarking technique to provide the security for NoSQL databases.

The embedded watermark can subsequently be used for claiming and proving ownership. Unique watermark is used to mark the data by using digital content to provide the ownership protection. Here we need to preserve the data quality in watermarked data so that to get data sufficiently high quality and can used for future planning processes in different application domains. The data quality degradation can be overcome by using reversible watermarking technique by embedded watermark data along with recovery of original data. The drawback this technique is that loss of the data quality due to modifying the data to a very large extent.

The watermark pre processing stage processes diverse parameters for computation of an optimal watermark. These parameters are utilized for watermark encoding and translating. The fundamental centre of watermark encoding stage is to embed watermark data so as to not influence the information quality. During watermark embedding, data gets balanced by available information transmission or capacity of the watermark information. To ensure the robustness the bandwidth of the watermark should be sufficiently large but not so large that that it affects the data quality. The data owner chooses the measure of information adjustment such that the quality is not compromised for a specific database application before hand and subsequently characterizes ease of use limitations to bring decent twisting into the information. Numerical components can be taken under consideration from any dataset and a reasonable element is resolved to embed watermark on the premise of common data. After watermarking, the information is discharged to the intended recipients over a correspondence channel that is thought to be insecure and termed as the "attacker channel" in this research domain. The information may experience a few malicious attacks in the attacker channel. The efficiency and effectiveness of reversible watermarking is depicted through robustness investigation controlled by its reaction to subset insertion, change and cancellation attacks.

In this project we develop a Reversible watermarking solution for No SQL databases. No SQL data is watermarked using reversible watermarking scheme, so that we can identify the data theft, data alternation and ensure ownership right.

IV. OVERVIEW OF MODULES

The below figure shows in which first data in the form of key, value pair is stored in the table which is present in NoSQL database then watermark is calculated using the HMAC-SHA1 algorithm and that watermark is embed with the data and stored in the different table in NoSQL database. Then again new watermark is calculated using the same algorithm. The old watermark is extracted from the data while retrieving is compared with the new watermark to check security violation. If both the watermarks are same then data is not corrupted if both the watermark mismatches then there is a security violation.

The figure shown below contains four modules.

- NoSQL Data Insertion: The data is inserted in a NoSQL database. NoSQL HBS database were primarily created to full fill the big data processing demands of Web application platforms that involved dynamic web pages and development of social media websites like Facebook and Yahoo. Data is taking as input and watermarked data into DB is taken as output. The first main process is Watermarked Data Insertion.
- NoSQL Data Extraction: We extract the data information which was stored in NoSQL database.
- Feature Extraction: The features of each data from NoSQL are extracted here.
- Watermark Insertion: We can insert watermark concept on data's in the NoSQL database.

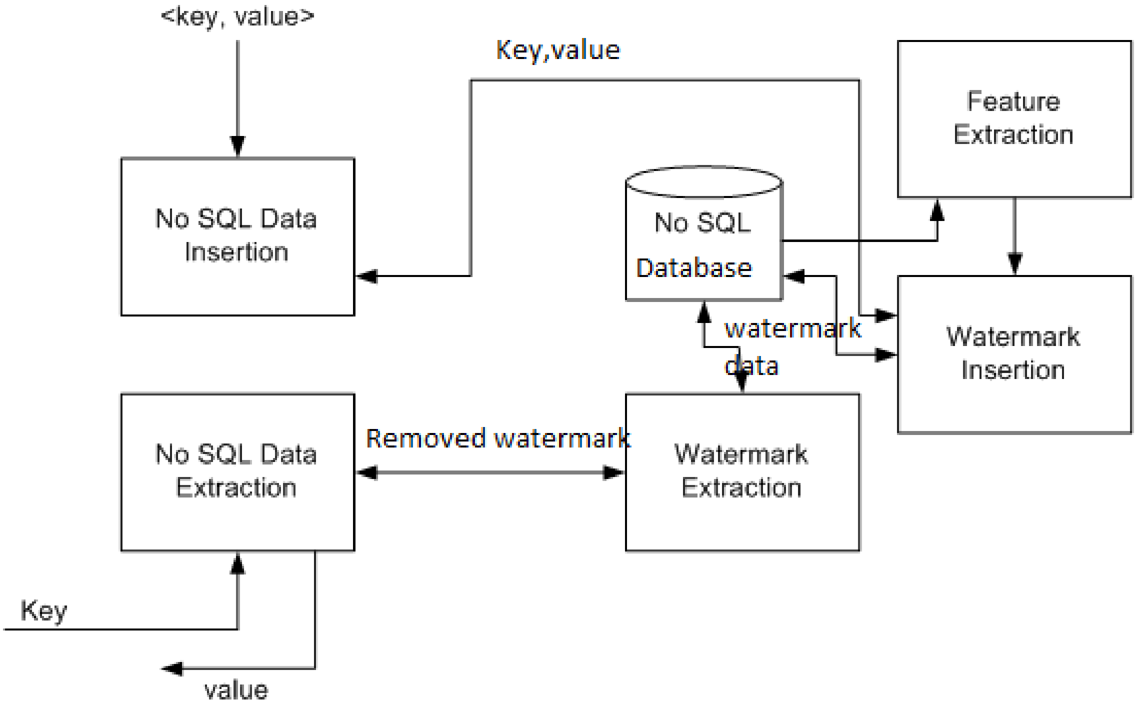

Fig 1: System architecture

DOI: 10.15680/IJIRCCE.2016. 0404265

V. Pseudo code

```
Step 1: Generate the key, value pair.
Step 1: Store the key, value pair data in the different tables of NoSQL databases.
Step 2: Calculate the watermark using HMAC-SHA1 algorithm (wm1).
Step 3: Add the watermark to which is stored the step 2.
Step 4: Store the data which is embedded with the watermark.
Step 5: Calculate the new watermark using the HMAC-SHA1 algorithm that is wm2.
Step 6: Extract the watermarked data from NoSQL databases.
Step 7: Extract the watermark from the data.
Step 8: Compare the new watermark calculated in the step 5 with watermark which is extracted in the step 7.
        If (wm1=wm2)
              Data is secure
        else
              Data is changed
Step 9: The repeat the same process for different tables.
Step 10: Make use of data which is secured.
Step 11: End
```

VI. Conclusion

In this paper we present and examine the new security problem for big data in NoSQL databases which aims to provide the security in terms of tamper proofing for wide range of data formats and ownership protection. We used a reversible watermarking algorithm to provide the security for NoSQL by using a unique watermark to mark the data and by using reversible watermarking technique which allows recovery of original data along with the embedded watermark information. Here Due to reversible watermarking technique data quality can be maintained, to reduce the distortions, and also able the recover the original data from the watermarked data. However watermarking technique is already been discussed for relational databases there for main contribution of this project is to provide the security for NoSQL databases.

References

1. R. Agrawal and J. Kiernan, “Watermarking relational databases, “in Proceedings of the 28th international conference on Very Large DataBases. VLDB Endowment, 2002, pp. 155–166.
2. M. E. Farfoura and S.-J. Horng, “A novel blind reversible method for watermarking relational databases,” in Parallel and Distributed Processing with Applications (ISPA), 2010 International Symposium on. IEEE, 2010, pp. 563–569.
3. Masoumeh RezaeiJam, Leili Mohammad Khanli, Mohammad Kazem Akbari, “A Survey on Security of Hadoop”, *IEEE*, 2014.
4. 5 Security Requirements for Big Data Hadoop Implementation.
5. D. M. Thodi and J. J. Rodriguez, “Prediction-error based reversible watermarking,” in Image Processing, 2004. ICIP’04. 2004 International Conference on, vol. 3. IEEE, 2004, pp. 1549–1552.
6. CSA, “Expanded Top Ten Big Data Security and Privacy Challenges”, CSA Report, 16 June2013https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf.
7. Saman Iftikhar, M. Kamran and Zahid Anwar, “RRW - A Robust and Reversible Watermarking Technique for Relational Data”, IEEE Transactions on Knowledge and Data Engineering val X, No : XX , 2015.
8. Asadulla Khan Zaki, “NoSQL DATABASES: NEW MILLENNIUM DATABASE FOR BIG DATA”, International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.
9. Yuri Demchenko, Canh Ngo, Cees de Laat, Peter Membrey and Daniil Gordijenko, “Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure”, Springer International Publishing Switzerland 2014, pp. 76–94.
10. Elisa Bertino & Ravi Sandhu.”Database Security- Concepts, Approaches and Challenges”, IEEE, 2005.
11. E. Sonnleitner, “A robust watermarking approach for large databases,” in Satellite Telecommunications (ESTEL), 2012 IEEE First AESS European Conference on. IEEE, 2012, pp. 1–6.
12. S. Subramanya and B. K. Yi, “Digital rights management,” Potentials, IEEE, vol. 25, no. 2, pp. 31–34, 2006.

13. K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases," Journal of Systems and Software, 2013.
14. A. M. Alattar, "Reversible watermark using difference expansion of triplets," in Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, vol. 1. IEEE, 2003, pp. I–501.
15. J.-N. Chang and H.-C. Wu, "Reversible fragile database watermarking technology using difference expansion based on svr prediction," in Computer, Consumer and Control (IS3C), 2012 International Symposium on. IEEE, 2012, pp. 690–693
16. G. Gupta and J. Pieprzyk, "Database relation watermarking resilient against secondary watermarking attacks," in Information Systems security. Springer, 2009, pp. 222–236.
17. Sethuraman Srinivas, Archana Nair," Security Maturity in NoSQL Databases – Are they Secure Enough to Haul the Modern IT Applications?" International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015.
18. Saman Iftikhar, M. Kamran and Zahid Anwar," RRW - A Robust and Reversible Watermarking Technique for Relational Data", IEEE Transactions on Knowledge and Data Engineering val X, No : XX , 2015.

BIOGRAPHY

AMREEN is a student in the Computer science department from Ballari institute of technology and management, Ballari. Pursuing Master's in computer networks and engineering. Her research interests are computer networks, web technology etc.

DADAPEER is an Assistant Professor in the Dept. of Computer Science and Engineering at Ballari Institute of Technology and Management, Ballari, Karnataka, India. He is working on the smart security system for various databases.

DOI: 10.15680/IJIRCCE.2016. 0404265