



Enhancement Of Security Of Unstructured Data Using Picture Key Cryptographic Mechanism: The Implementation

¹Meena Kumari, ²Mr. Abhishek Bhatnagar

¹Research Scholar, Deptt. Of Computer Engineering, IJET Kinana (Jind)

²Assistant Professor & HOD, Deptt. Of Computer Engineering, IJET Kinana (Jind)

Peacefulboy8060@gmail.com, Ap.abhi.ijet@gmail.com

ABSTRACT: Unstructured data is data comes from both machines & human generated & it is broadly classified into two types: Textual and Non Textual. The main purpose of this research is Investigation of new challenges in data ware house management in unstructured data. Then we would convert unstructured data in form of matrix. In this research we have used MATLAB as simulation tool. Here we would make comparative analysis of traditional unstructured data security & proposed unstructured data.

Keyword: Data Warehouse, Data mining, MATLAB, Unstructured data, Data security, XOR, Cryptography

[1] Unstructured Data

Unstructured data is data comes from both machines & human generated & it is broadly classified into two types: Textual and Non Textual. In Non-Textual unstructured data is like still images, videos, & MP3 audio files. Textual unstructured data examples are like email messages, instant messages, memos, PowerPoint presentations etc. & different standards for unstructured data are open XML, SMTP, SMS, CSV & Information & content exchange.

The goals of an Unstructured Data Management System (UDMS) are difficult to outline because everyday researchers & other industry players find new problems & new applications.

However, based in previous works on domain, there are some common goals that these systems would achieve:

- _ Data categorization
- _ Search over data
- _ Enhance user experience
- _ Reveal hidden information
- _ Join sparse knowledge

The most well known example of search over unstructured data is Google [3], that provide to users a search tool about Internet (probably bigger source of unstructured information). Live plasma [4] is a good example of an application that shares these goals, showing relations among films or among musicians, in a innovative user interface. An UDMS should provide a rime work that support building applications within one or more of these goals.

ISSN : 2278-6848



9 772278 684800

© International Journal for Research Publication and Seminar



One of major differences between structured & unstructured environments is that in structured environment data is updated on a regular basis. Whenever your ATM is used, whenever a deposit is made, whenever a check is cashed, your bank account is updated. & it is at these times that structured data is commonly measured, or some record is created.

[2] MATLAB

MATLAB (**matrix laboratory**) is a multi-paradigm numerical computing environment & fourth-generation programming language. Developed by Math Works, MATLAB allows matrix manipulations, plotting of functions & data, implementation of algorithms, creation of user interfaces, & interfacing within programs written in other languages, including C, C++, Java, Fortran & Python. The MATLAB application is built around MATLAB language, & most use of MATLAB involves typing MATLAB code into Command Window (as an interactive mathematical shell), or executing text files containing MATLAB code, including scripts and/or functions.

[3] DATA ENCRYPTION IN UNSTRUCTURED DATA

Encryption changes data or information that is normally plaintext through usage of an algorithm so that someone must possess certain knowledge to access it. This special knowledge is normally called a key. For example, something is encrypted if someone must enter a password to

access it. Encrypted unstructured data codes are unstructured data codes that not everyone could scould & access. They are not very common, since most unstructured data codes are used in marketing, & developers of those codes want them to be accessible by everyone. Secure unstructured data codes could be made that make shoulder enter a password to be able to access content. This is a good idea to make for employees of a company. Company could make secure unstructured data codes that employee had been to enter company password to view. QR codes are normally encoded in plaintext.

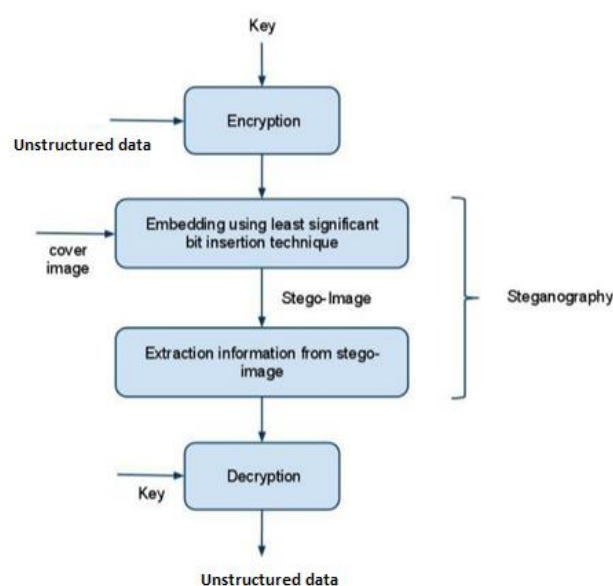


Fig 1. Data Encryption in Unstructured Data

[4] RESULT AND DISCUSSION

Encryption process of unstructured data code:

In order to decrypt unstructured data code we would open encryption.fig in matlab using guide command:

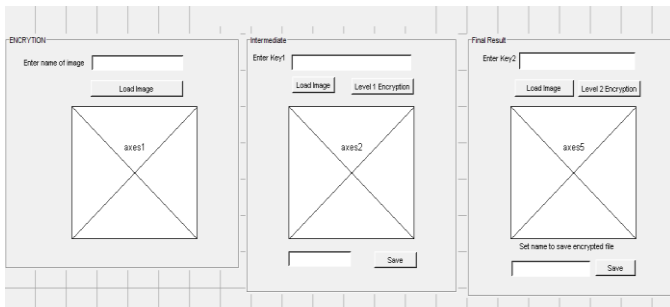


Fig 2. Unstructured data code encrypt

When we load qr3.jpg image for encryption this window is generated:



Fig 3 Load qr3.jpg image for encryption this window is generated

When we click on load image button after typing fkey1.jpg in text box then Fkey 1.jpg image is displayed for level 1 encryption.

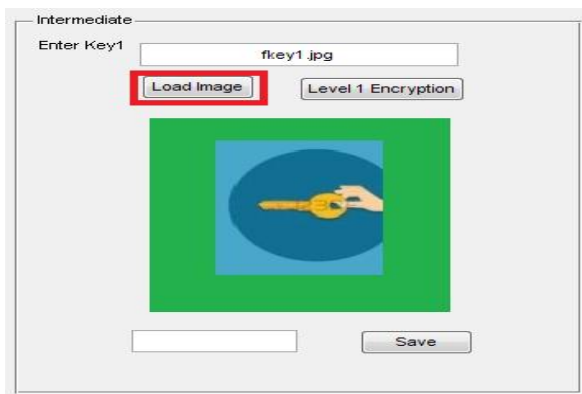


Fig 4 Fkey 1.jpg image is displayed for level 1 encryption.

When we click on level1 encryption button then Intermediate Level 1 Encrypted image is generated.

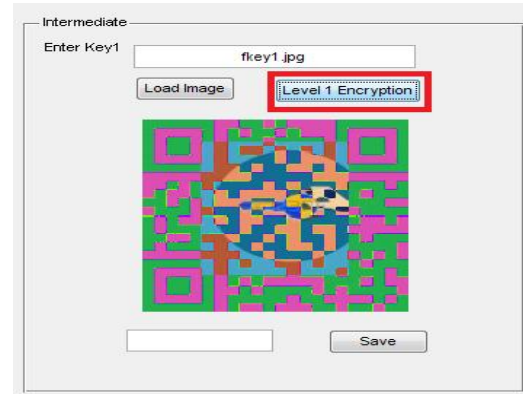


Fig 5 Intermediate Level 1 Encrypted Image generated fkey2.jpg is loaded for level 2 encryption.

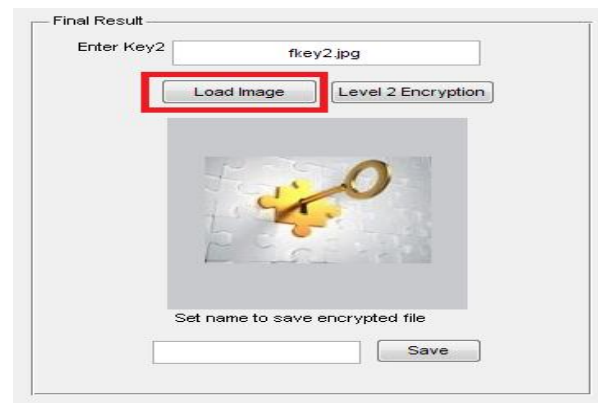


Fig 6 Fkey2.jpg is loaded for level 2 encryption



Fig 7 Level 2 Encrypted image



Decryption process

Encfinal.jpg image is loaded for decryption



Fig 8 Encrypted Image

Fkey2.jpg image is loaded for level 1 Decryption.

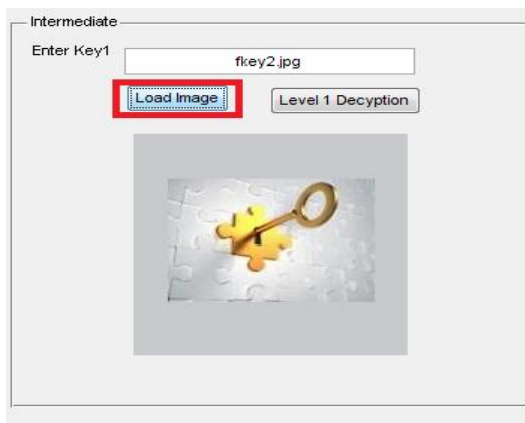


Fig 9 Fkey2.jpg image is loaded

Level 1 Decryption image is shown here.



Fig 10 Level 1 Decryption image is shown
Fkey1.jpg image is loaded for level 2 Decryption.

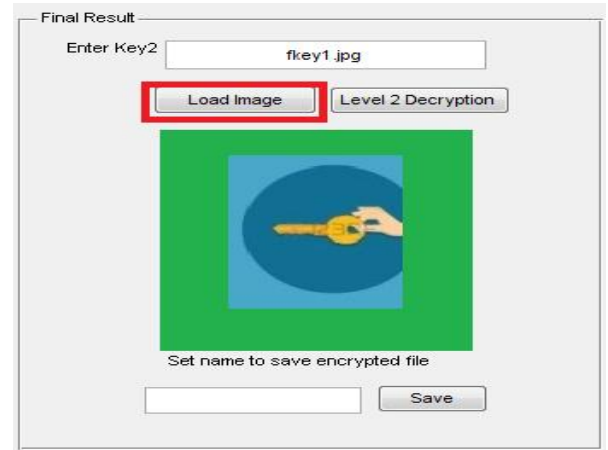


Fig 11 Fkey1.jpg image is loaded for level 2 Decryption

Final level 2 Decrypted image.



Fig 12 Final level 2 Decrypted image

[5] NEED OF SECURITY IN UNSTRUCTURED DATA CODES

One of first applications that come to mind is to use encrypted unstructured data codes on passports, driver license & other identification or even loyalty cards. Assume that every citizen



would have a hidden Id residing in a secured governmental database. This hidden Id points in database to an overt Id printed on passport together with a name & other details. Unstructured data code encrypts a URL & hidden Id. All inspectors had been to do is compare received data to that on passport.

[9] CONCLUSION

Unstructured textual data is being constantly generated via call centre logs, emails & documents on web, blogs, tweets, customer comments, customer reviews, & so on. While amount of textual data is increasing swift, ability to make summary, understand, & make sense of such data for making better firm decisions pending challenging. Investigation of new challenges in data ware house management in unstructured data.

REFERENCE

[1] E. Agichtein, L. Gravano, J. Pavel, V. Sokolova, A. Voskoboynik. Snowball: A prototype system for extracting relations from large text collections. In SIGMOD, 2001.

[2] S. Brin. Extracting patterns relations from world wide web. In WebDB, 1998.

[3] Y. Cai, X. Dong, A. Y. Halevy, J. Liu, J. Madhavan. Personal information management with semex. In SIGMOD, 2005.

[4] X. Chai, B. Vuong, A. Doan, J. F. Naughton. Efficiently incorporating user interaction into extraction integration programs. Technical

Report UW-CSE-2008, University of Wisconsin-Madison, 2008.

[5] F. Chen, A. Doan, J. Yang, R. Ramakrishnan. Efficient information extraction over evolving text data. In ICDE, 2008

[6] E. Chu, A. Baid, T. Chen, A. Doan, J. F. Naughton. A relational approach to incrementally extracting querying structure in unstructured data. In VLDB, 2007.

[7] P. DeRose, X. Chai, B. Gao, W. Shen, A. Doan, P. Bohannon, X. Zhu. Building community wikis: A machine-human partnership approach. In ICDE, 2008.

[8] P. DeRose, W. Shen, F. Chen, A. Doan, R. Ramakrishnan. Building structured web community portals: A top-down, compositional, incremental approach. In VLDB, 2007.

[9] P. DeRose, W. Shen, F. Chen, Y. Lee, D. Burdick, A. Doan, R. Ramakrishnan. Dblife: A community information management platform for database research community (demo). In CIDR, 2007.

[10] A. Doan. Data integration research challenges in community information management systems, 2008. Keynote talk, Workshop on Information Integration Methods, Architectures, Systems (IIMAS) at ICDE-08.

[11] A. Doan, P. Bohannon, R. Ramakrishnan, X. Chai, P. DeRose, B. Gao, W. Shen. User-centric research challenges in community



information management systems. IEEE Data Engineering Bulletin, 30(2):32–40, 2007.

[12] A. Doan, J. F. Naughton, A. Baid, X. Chai, F. Chen, T. Chen, E. Chu, P. DeRose, B. Gao, C. Gokhale, J. Huang, W. Shen, B. Vuong. case for a structured approach to managing unstructured data. In CIDR, 2009.

[13] A. Doan, R. Ramakrishnan, F. Chen, P. DeRose, Y. Lee, R. McCann, M. Sayyadian, W. Shen. Community information management. IEEE Data Engineering Bulletin, 29(1):64–72, 2006.

[14] A. Y. Halevy, M. J. Franklin, D. Maier. Principles of dataspace systems. In PODS, 2006.

[15] J. Huang, T. Chen, A. Doan, J. F. Naughton. On provenance of non-answers to queries over extracted data. PVLDB, 1(1):736–747, 2008.

[16] R. Krishnamurthy, Y. Li, S. Raghavan, F. Reiss, S. Vaithyanathan, H. Zhu. Systemt: A system for declarative information extraction, 2008. SIGMOD Record, Special Issue on Managing Information Extraction.