ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

OPTIMIZED HASH BASED SECURITY SYSTEM FOR BIG DATA

Dr.K.Thamodaran*, Dr K.Kuppusamy
Professor, Dept. of Computer Science, Marudupandiyar College Thanjavur, Tamilnadu, India-613 403.
Professor and Chair, Dept. of Computer Science and Engineering, Alagappa University, Karaikudi, Tamilnadu, India-630 003.

DOI: 10.5281/zenodo.495153

ABSTRACT

Big data is a collection of very large and complex set of data which is remote away from the capacity of the existing database management tools or traditional data processing applications. The data includes private and sensitive information that need to be kept secure and safe. Numerous security information and event management tools were designed to protect and manage unstructured data. In this paper hash and PSO based security system is proposed for security of Big Data. The hash features are extracted in the spatial domain of the image through PSO to generate the image hash. The extracted hash is encrypted to maintain the secrecy.

KEYWORDS: Big data, Encrption, Image Hash function, PSO, Security.

INTRODUCTION

Information security is becoming one of the essential requirements for the protection of civil rights and liberties. The social opportunities and economic potential presented by digitisation must not be endangered. Additionally, the subjects of information security and the defence against industrial intelligence should play an exceptional role. In the period of information communication technology, the organizations are having more responsibility to take necessary steps to meet the fulfilment and secure the data from external and internal terrorization. The big data is playing very significant role and offer service for innovation, differentiation and growth of our society such as government, finance, security and so on. Big data security analytics is an accepted term for the potential organizations to collect and analyze enormous amounts of security data to detect complete susceptibility and intrusions [13**].**

Access controls are security characteristics that direct how the users and systems communicate and interact with other systems and resources. Access controls can be employed at various layers of a network and individual systems. Access control is the collection of mechanisms that permits managers of a system to exercise a directing or restraining influence over the behaviour, use, and content of a system. The access controls can be categorized into three layers namely administrative controls, physical controls, technical or logical controls. Each category of access control having different access control mechanism that can be carried out manually or

automatically. Every category of access control has several components such as cryptographic keys, uses private keys and digital signatures offer a higher level of security than passwords. The software tools are known as technical controls or logical controls which are used to control subject's access to objects. The software tools are core OS components, add-on security packages, applications, network hardware devices, protocols, encryption mechanisms, and access control metrics. They shield the integrity and availability of resources by restraining the number of subjects that can access them and guard the confidentiality of resources by preventing exposé to unauthorized subjects [6], [20], [21].

Security of digital contents is resolved by facilitating the recipients to verify the authentication of the received contents and prevent forgery using hashing techniques. The content-based digital signature of the data is known

as data hash. The hashing technique extracts a set of features from the large data to create a compressed sign that can be used for security. An image hash is a content-based digital signature of the image data. A secret key is used to extract certain features from the image data to generate image hash. Chai Wah Wu has offered an authentication scheme using DFT to generate hash features. An innovative idea is employed to use the hashing technique in CBIR. The hash value generated and is utilized for extracting the images similar to query image from the large image database. Performance of the work is measured by means of hamming distance [2]. Ee-Chien Chang, Mohan S. Kankanhalli, Xin Guan, Zhiyong Huang, YinghuiWu have proposed Content-based spatial domain image authentication scheme. In this scheme the signature is generated from the result of an extremely low-bit-rate content based compression which is guided by a space-variant weighting function whose values are higher in the more important and sensitive region [3]. Takeyuki Uehara et al. have explained an image authentication scheme, which constructs a Message Authentication Code through integrating a number of feature codes. This code shelters the region of interest in the image and secure from JPEG compression [4].

Ashwin Swaminathan, Yinian and Min Wu have presented the scheme for generating an image hash based on Fourier-Mellin transform features which are invariant to two-dimensional affine transformations and incorporates key-dependent outputs to form a secure and robust image hash [5]. V.Monga et al. have intended an image hashing scheme using non-negative matrix factorization for authentication. An image is considered as matrix and hashing is considered as a randomized dimensionality reduction that retains the essence of the original image matrix while preventing intentional attacks of guessing and forgery [7]. H.B.Kekre et al. have projected an image hashing scheme using CBIR and hamming distance. The hash value is created by means of dig out the identical images to query image among the large image database [9]. Akash Gupta, Alok Shukla, S. Venkatesan have offered a security scheme to make modification in secure remote password protocol to provide secure authentication and access control in big data environment and its benefits over some traditional methods of security implementation being used in current big data environments [18].

In this paper, an optimized image hash security system is proposed using Particle Swarm Optimization (PSO) for user authentication and security of big data. In this system content based image features are first extracted using PSO with secret key from the spatial domain then the image hash encrypted and is utilized to verify the authenticity of users. The subsequent sections of this paper are providing information about the big data strategy, particle swarm optimization(PSO) technique, proposed optimized hash based security system for big data through PSO, experimental results and the conclusion.

MATERIALS AND METHODS

Big Data Strategy

In 1998, very first time Mr.John has illuminated the term “Big Data” in a Silicon Graphics Mashey with the title of "Big Data” and the Next Wave of Infra Stress [11]. Every day the huge amount of data is derived. This data is known as Big Data [12]. In 2012, Rodriguez has defined that for years, statisticians have been working with large volumes of data in fields as diverse as astronomy, bioinformatics, and data mining. Big Data is different because it is generated on a massive scale by countless online interactions among people, transactions between people and systems, and sensor-enabled machinery. In 2013, Horrigan has defined that the big data as non sampled data, characterized by the creation of databases from electronic sources whose primary purpose is something other than statistical inference. Big Data is a cross-disciplinary concept which contains more data than making sense out of. Big Data is a call for us computer scientists to once again provide even better

methods to crunch even more diverse, even more complex, even more dynamic, even more fine-grained, even larger data. Big Data brings new opportunities for institutions of higher education, as institutions continue to face unprecedented challenges in their environment. The value of big data to an organization falls into two categories: analytical use, and enabling new products. Big data analytics can reveal insights hidden previously by data too costly to process, such as peer influence among customers, revealed by analysing shoppers transactions, social and geographical data. Big data is increasingly becoming a factor in production, market competitiveness and, therefore, growth. Cutting-edge analysis technologies are making inroads into all areas of life and changing our day-to-day existence. Sensor technology, biometric identification and the general trend towards a convergence of information and communication technologies are driving the big data movement [17], [14], [19], [15], [16].

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Big Data includes five 5 major characteristics. They are enlightened as follows:

- *Volume:* It represents the size of the big data set. It is the most noteworthy characteristic of big data.
- *Variety:* Plentiful resources (internal or external) have supplied various data to the companies. This data are represented in the form of either structured or unstructured.
- *Velocity:* The production rate of big data is called its velocity and it is very high. The heavy increase in data means that the data should be analyzed more rapidly. The faster the data increases, the quicker the need for the data increases; therefore the process shows increase as well.
- *Veracity:* It indicates the correctness of the data. The data should be acquired from legal resources and it should be secure. Only authorized people should have the access authorization.
- *Value:* An outcome should be generated after all of the processes and the result should supplement the process.

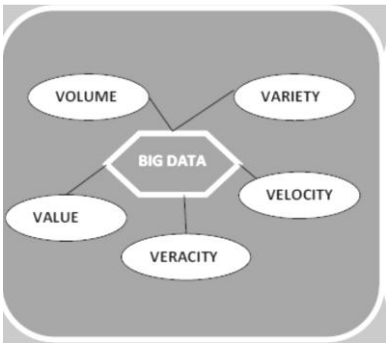

Figure 1: Dimensions of Big Data

PARTICLE SWARM OPTIMIZATION

Kennedy and Eberhart have developed an evolutionary computation system called PSO [8]. A particle swarm optimization technique is formed from the imitation of the social behavior of bird flocks. Swarm Intelligence is an ingenious distributed intelligent conception to solve optimization problems that initially regard as its motivation from herding phenomena in vertebrates [1]. In PSO scheme, the particle is regarded as a bird in the search space to generate a solution. Every particle establishes the direction and distance of the next move with respect to velocity and optimized function which decides a fitness function. The scheme tracks the optimal particle at present in the solution space. In modified particle swarm optimization algorithm, every particle fix on its inertial factor concurrent to the nearer degree between the fitness of itself and the optimal particle [10]. Each particle makes effort to revise its position with regard to consequent information:

- the space between the present position and position of particle best
- the space between the present position and position of global best

This development can be represented by the concept of velocity. The modification of velocity of each agent is performed with help of equation (1) in inertia weight approach (IWA).

$$V_{k+1} = W * V_k + C_1 * r_1 * (p_k - x_k) + C_2 * r_2 * (g_k - x_k) \quad (1)$$

where, W – non negative inertia factor, V_k - velocity of particle, x_k - present position of particle,

C_1-determine the relative influence of the cognitive component, C_2 - determine the relative influence of the social component, p_k - *pbest* of particle , g_k - *gbest* of the group, r_1, r_2 – the population is getting diversity with help of random numbers and are consistently distributed in the interval [0,1].

The particle make a decision to move to next position through equation (1) and regarding its own experience, which is the memory of its best earlier position, and the practice of its most successful particle in the swarm. The particle explores the solution in the problem space in the range of [−s, s]. The particle is updating its position by means of equation (2).

$$\mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{V}_{k+1} \quad (2)$$

where, V_{k+1}-Modified Velocity, X_k-Current Position , X_{k+1}-Modified Position.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

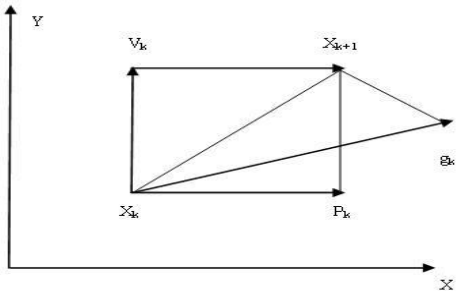

Figure 2: Updating PSO Searching Point

PROPOSED SECURITY SYSTEM FOR BIG DATA

The proposed hash based security system is developed for secure communication of big data using Particle Swarm Optimization (PSO). In this system, PSO is used to generate the hash vectors, and SHA-512 algorithm is employed to create secure and robust 512 bit image hash code. All coefficients of the secret image are used for hash generation and PSO technique with 256 bit secret key is used to select high energy coefficients to generate hash code from 8x8 non-overlapped blocks of secret image.

In this system, an innovative method is used to improve the inertial factor correctly as mentioned below. Considering a maximization problem, the inertial factors of the particles are updated according to equation (3).

$$w_m = \frac{rn}{pm}\left|\frac{f_{cp} - f_{opc}}{f_{opc}}\right|, \quad if \quad w_m > w_0 \quad \text{then} \quad w = w_m, \qquad (3)$$

$$if \quad w_0 > w_m \quad \text{then} \quad w = w_0.$$

where, rn -random number , pm -Parameter, f_{cp} -fitness of current particle , f_{opc} -optimal particle currently. Fitness function f(x) for PSO training intended for authentication scheme is given in equation (4).

$$\text{Fitness Function} \quad f(x) = \frac{1}{n}\sum_{i=1}^{n} COS_i \qquad (4)$$

A. PSO Algorithm

The PSO algorithm proposed in section III is used to modify the velocity and position of particles when generating and extracting hash features.

Step 1. Generate randomly the initial position and velocity of the particles within predefined ranges.
Step 2. The velocities of all particles are updated in every iteration according to equation(1), where w will be gained according to equation (3).

Step 3. The positions of all particles are updated according to equation (2) after updating, should be checked and limited to the allowed range.
Step 4. When condition is satisfied, the pbest and gbest values are revised.

B. Proposed Hash Generation Algorithm

Step 1. Generate the secret key S_k.
Step 2. Form hash features using all coefficients of from 8x8 non-overlapped blocks of secret image and select high energy coefficients through particle swarm optimization.
Step 3. Compress the feature vector in required length using SHA-512.
Step 4. Concatenate the hash bits generated to form the final hash H.
Step 5. Perform encryption on final hash H to obtain H^* through the secret key.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

RESULTS AND DISCUSSION

The authenticity of user to access the big data can be improved by means of adapting optimized image hash and secret key. The encrypted image hash also offers more security for big data. The PSO plays a vital role in selecting high energy vectors for security purpose. The 256 bit secret key is used to form the image hash vectors and encryption of final hash for providing high level of security and user authentication. The key space should be effectively large enough to make the security system meaningful. Key space means the total number of different keys can be exercised to perform encryption and decryption processes. A 256 bits key is employed in this scheme. If any intruder has attempted to break the key, as a result the attacker has to try out 2^{256} ($2^{256} \approx 1.1579 \text{ X } 10^{77}$) combinations of the secret key. An image hash encryption via such a large key space is more than enough for better security.

Hash-Lena-PSO

Embedded Hashing on High Energy -Lena:
192,105,27,205,66,215,159,94,125,180,167,19,86,30,189,240,183,83,227,238,48,170,132,252,80,248,154,100,108,136,98,76,98,218,79,169,63,108,155,6,45,105,61,98,192,14,19,49,61,86,196,184,63,98,242,69,214,237,195,5,229,188,222,143.

Encrypted Hash -Lena :
252,240,18,98,174,242,170,19,40,169,9,50,15,21,15,158,143,134,220,207,67,18,87,32,110,22,214,96,18,242,211,141,137,81,202,185,1,87,124,188,76,82,12,217,156,146,37,16,112,159,53,242,125,159,136,106,13,105,70,110,47,224,154,77.

Decrypted Hash-Lena :
192,105,27,205,66,215,159,94,125,180,167,19,86,30,189,240,183,83,227,238,48,170,132,252,80,248,154,100,108,136,98,76,98,218,79,169,63,108,155,6,45,105,61,98,192,14,19,49,61,86,196,184,63,98,242,69,214,237,195,5,229,188,222,143.

Hash- Mandrill –PSO

Embedded Hashing on High Energy –Mandrill :
198,189,98,13,203,234,211,140,209,89,6,132,167,72,126,248,114,58,110,38,43,35,205,118,115,180,96,246,214,64,169,248,39,239,236,179,127,68,107,88,49,27,142,49,160,193,19,55,113,34,126,225,120,136,162,81,124,107,133,135,65,179,24,174

Encrypted Hash –Mandrill :
250,36,107,163,38,207,230,193,133,68,168,165,254,67,204,150,74,239,81,7,88,155,30,170,77,90,44,242,168,58,24,57,204,100,105,163,65,127,140,226,80,32,191,138,252,93,37,22,60,235,143,171,58,117,216,126,167,239,0,236,139,239,92,108.

Decrypted Hash–Mandrill
:198,189,98,12,202,234,211,140,208,89,6,132,167,72,126,248,114,58,110,38,43,35,205,118,115,180,96,246,214,64,169,248,39,239,236,179,127,68,107,88,49,27,142,49,160,193,19,55,113,34,126,225,120,136,162,81,124,107,133,135,65,179,24,174.

CONCLUSION

The optimized hash based security system is developed for secure communication of big data using PSO. In this system, the secret image is considered to construct the hash code through PSO and secret key. The PSO system prefers the high energy coefficients to create the hash code. The proposed system offers authentication for the client and big data. The optimized hash based security scheme is providing more security for big data retrieval by users. The PSO system prefers the high energy coefficients to create the hash code or digest. The observed outcomes are proving effectiveness of the proposed system.

REFERENCES

[1] Clerc M and Kennedy J, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space", IEEE Transactions on Evolutionary computation, vol.6(1), 2002, pp 58-73.

[2] Chai Wah Wu, "On The Design Of Content-Based Multimedia Authentication Systems", IEEE Transactions on Multimedia, Vol. 4, No. 3, 2002, pp 385-393.
[3] Ee-Chien Chang, Mohan S. Kankanhalli, Xin Guan, Zhiyong Huang, Yinghui Wu, "Robust Image Authentication Using Content Based Compression. Multimedia Systems", Springer-Verlag., 2003, pp 1-10.
[4] Takeyuki Uehara, Reihaneh Safavi-Naini and Philip Ogunbona, "A Secure and Flexible Authentication System for Digital Images", Multimedia Systems, Springer Verlag, Vol. 9, 2004, pp. 441-456.
[5] Ashwin Swaminathan, Yinian and Min Wu, " Robust and secure image hashing", IEEE Transactions on Information Forensics and Security, Vol. 1, No. 2, 2005, pp 215-229.
[6] Shon Harris and Allen Harper, " Gray Hat Hacking: The Ethial Hacker's Handbook (All-In-One)", January 2005.
[7] Monga.V, M.K.Mihcak, "Robust and Secure Image Hashing Via Non- Negative Matrix Factorizations", IEEE Transactions Information Forensics and Security vol 2(3), 2007, pp376–390.
[8] Maurice Clerc, Particle Swarm Optimization, ISTE publishers, First South Asian Edition.,2007
[9] H. B. Kekre, Dhirendra Mishra,"Image Retrieval Using Image Hashing", Techno-Path: Journal Of Science, Engineering & Technology Management, Vol. 1 No.3, 2009
[10] Jinrong Zhu , "A Modified Particle Swarm Optimization Algorithm" Journal of Computers, Vol. 4, No. 12, 2009, pp 1231-1236.
[11] F. Diebold. On the Origin(s) and Development of the Term "Big Data". Pier working paper archive, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, 2012.
[12] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: challenges, issues, and opportunities," in Database Systems for Advanced Applications, Springer, Berlin, Germany, 2013. pp. 1–15,
[13] J. Oltsik. Defining big data security analytics. Network world, 1 April 2013.
[14] Alvero A. Cardenas, Pratyusa K. Manadhata, Sreeranga P.Rajan, "Big Data Analytics for Security" IEEE Security&Privacy, vol.11 no.6, Nov.-Dec. 2013, pp. 74-76.
[15] Michael Minelli, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses,"Wiley, 2013, ISBN:'111814760X.
[16] J. Hurwitz, "Big Data for Dummies," Wiley, 2013, ISBN:978-1-118-50422-2.
[17] Thomas F. Dapp , "Big data : The untamed force" , May 5, 2014 ,
[18] Akash Gupta, Alok Shukla, S.Venkatesan, "Big Data: Cryptographically Enforced Access Control And Secure Communication", Proceedings of 6th IRF International Conference, Chennai, India, 10th May. 2014, ISBN: 978-93-84209-16-2.
[19] Richard Zuech, Taghi M Khoshgoftaar and Randall Wald, " Intrusion detection and Big Heterogeneous Data: a Surve, *Journal of Big Data* (2015) 2:3.
[20] Reiner Kappenberger, Protecting Your Data against Cyber Attacks in Big Data Environments ,14 – ISSA Journal, February 2016.
[21] http://www:en.wikibooks.org/wiki/Fundamentals_of_Information_Systems_Security/Access_Control_Systems.