

Available online at <http://www.mecspress.net/ijem>

Security Analysis and Proposal of Kill Model using Modern Quantum Key Distribution in Big Data Mining

Sayantan Gupta ^a

^a *First affiliation, Amtala, Inside "Behind Khela Ghar", Garia, Kolkata 700084, India*

Abstract

Data now days is of the utmost importance in our network system maybe be in Raw form or be it processed Information. This research paper represents the Security Analysis of modern Big Data technology with the various upcoming challenges and threats which has a vast effect on any organization working in this field and thus the various solutions using Quantum and other cryptographic methods with encryption and numerous other pitfalls and the applications of such methods in real life.

Index Terms: QKD-Quantum key Distribution, Big Data, Cryptography, Security Challenges, Big Data Mining, Data Analysis.

© 2017 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Main Text

The use of modern day Big data technology is vast and insecure. Data nowadays is used as a major reference to various decision making, discovery issues of technologies and implementation of such technology in real world with support. In our practical world, we store raw data in something called Database where the data can be reused, restored, changes or modified and accessed accordingly. Now Data Mining is a term or technology which involves retrieving of data in its raw form which is to be used for processing of the available sequence. So, what is Big Data? Big Data is a special term used for unspecified interactions between people and people or people and sensors or sensors and machines. The security of such huge Data is a major concern as encryption and decryption of such character is turning out to be dangerous day by day. The paper at the first defines the various problems in the present day and then proposes a computational KM Model which can be implemented so as to make the complexity reduced and efficient to use.

* Corresponding author. Tel.: +91 9804176347
E-mail address: sayantangupta999@gmail.com

1.1. The Survey

Why use Big Data?

At the very first level, the entire Management and summation of disrupted data is a unique skill which is very much needed in Big Data analysis. The numerous data can be of various types ranging from organization specific data or general purpose data or public. Big data are basically described by the implementation of the 3 V's: Variety, Volume, and Velocity.

Data Volume: In recent times, we have seen a huge increase in data space availability. Today we have 4G. B RAMs in our computers which was a dream few decades back. This huge and large variety of available data makes it complex of implementation of any technologies or mining of data for any specific purpose.

Data Velocity: The basic inner meaning of Big data mining is to search for a specified data or groups of varied data and the entire process should be completed within a specific pre-determined time interval or time range. While the implementation of the primitive or old procedures of mining can take huge time to search for a specific required raw data, which is a major advantage of modern big data analysis.

Data Variety: The term Big data comes from a range of sources which has both defined and undefined data, revealed or hidden data etc. Previously the primitive management technology was made and arranged to implement lesser modules of undefined and irrelevant data, where in the other hand Big Data is geospatial data, processing data, including special location files and various interactions from users. The inheritance of such data makes varied disruptions for analysis, managing, mining and utilizing the data.

Thus, Big Data mining refers to the process undergrown to search for valid information for the benefits of the user. Big data information is valid and used in space research, developing current climatic conditions, social networking sites in which are involved mainly, in the entire medical industry, various government funded agencies looking for data like Survey Department, and Big Data has major implementations in Communications systems. Now we have Two specific targets in data mining so as to develop new efficient technologies and methods which may efficiently presume the nature of the development of the raw data available and give ways to implement future prospective. Making use of such huge raw data space is a huge task as day by day man is evolving so as the society. We need to implement newest technologies to make perfect use of such fancy data.

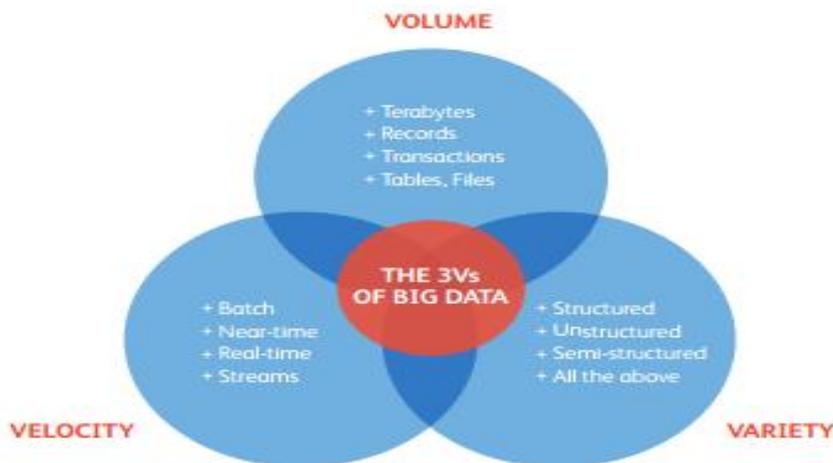


Fig.1.

1.2. Management vs Big Data vs Organization

Nowadays we have a lot and lots of data incoming in big organizations from various sources like daily interactions, machine values, records management etc., but getting the raw data is very easy, but if we want to store the data and implementation, analysis of the data in real-life applications is very difficult task which has to be done. Being specific a company or brand needs huge and he amount of data t know about its customers and potential views from them to get their feedback and requirements, so more the organizations receives the data the more beneficial it is, but at the same time it is to be noted that data in the Raw form is just no better than Garbage. Since we are talking a lot about data analysis and data mining let us take a deep ride about what is the advantages of the word data mining or data analysis in our today’s world. `

The Various Advantages of Big Data Analysis:

Now with the advancement of technology the different faults in the company are known in no time. This time specific capability of the technology helps the organization to build better strategies and cope with the problem at much better situation. This in return helps the company to stand out among others to deliver good products and services and helps the brand to get feasibly in the market and not to get vanished or completely disrupted.

Now the point is with the Real day insight you can get the information about your competitors very easily and almost instantly. Thus, with such Analysis you can always stay a step ahead of other brands doing business and you can clear in-depth notification about what the company is trying to do to make its market and act accordingly.

The review of marketing or marketing results will be much better ad easier, that in the long run will help to generate additional monetary benefits and will increase the turn over. Therefore, this analysis may determine how the company sales are increasing or decreasing and could help to make an overview of the net sales of any brand.

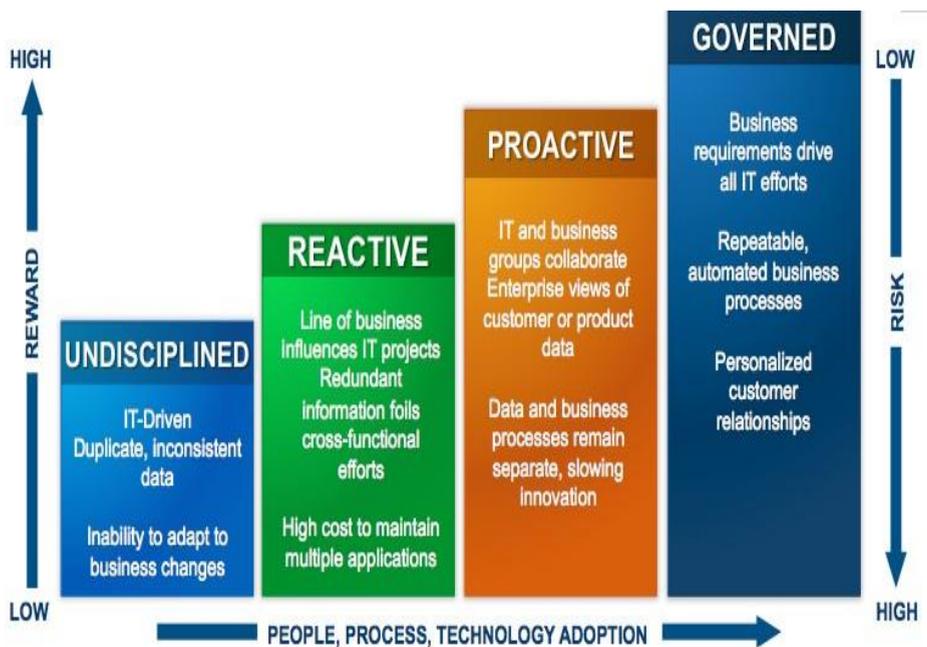


Fig.2.

Helps to keep up with the customer needs: Better view of various brands promotions, upcoming launches and marketing strategies and your user requirement details gives the much-needed data of the recent and upcoming user trends and thresholds. Thus, this will help to make the decisions regarding the promotions or marketing much faster and easier and thus in return will attract the potential customers in the long run.

The business with you do or the various services which you provide improves in quality which in turn helps you to get better benefit from the turn and helps to get a positive feedback from the potential customers. Likewise, if a brand continually takes good sound of the feedback from its customers then it can actively make changes in its products or services to give better benefits. Let us take an example, Real -Time Fire Alarm system can give you notification if something goes wrong at that point and therefore it could help you to get protection from any harm.

Now thing about Fraud cases. These frauds can be easily detected with the help of the analysis and this will help the organization in the long term from disasters and will prevent any kind of illegal proceedings. As we all know money attracts people, and the people can be bad not better and may be worse. So this real time analysis helps your company from any kind of hackers and prevents your private information or money from getting stolen or damaged. Let us consider the diagram:

1.3. Quantum Key Distribution in Data Mining

Quantum Key Distribution has a special ability to detect the presence of any third party or Eve in the system algorithm. But with the primitive technology division, the various issues regarding the Security and Protection based data use bit by bit-wise implementation and these implementations are not of much advantageous for analysis of big data analytics. Now we see that, using quantum bits (qubits) implementations one can get better security and protection validity from the primitive counter-parts. As we see when many user and their interactions get involved there is a high chance or passive attacks in the system algorithm. These kind of security provisions and threats which effect the data analysis can be solved by implementation of the Quantum Architecture which makes the algorithm more efficient and user friendly by reducing the time complexity which we will see later in this paper. Therefore, in this paper, we have implemented three specific ways which can give the much need raw data security and protection issues can be solved for further data centers/keepers-which store the data in huge amounts in bulk. Now the First way we found out by looking the problems is we must understand the security and privacy issue of the data analysis of the Data Keepers which can be located in different Environments each having specific security hazards ad protection issues. The Second thing is that, the primitive analysis of data found that most of the users wanted the implementation of hard keys which would virtually provide more security and privacy in key mismanagement and data verification between the customer system and the verification server. At this stage, we have decided to implement and introduce Quantum Key Distribution which will help to generate and distribute strong and specific keys with complexity much less than the existing ways and we implement the Double Hand algorithm which gives the necessary privacy based issues with hands on implementations over security which is applicable over much harsh environmental conditions. The Third issue is that privacy protection is the need of the hour which means that the data which we get from various sure must be protected from various people like even from the organization employees. Although the issues related to privacy are solved with utmost importance in most of the private companies, the ethical issues of data analysis in government sectors and government organizations are very much still at risk and are unsafe. Therefore, in this paper, we searched for the various solutions of data protection & privacy as in big organizations like healthcare sector and education sector industries. We will hope that these researches and finding will make good improvement in these major sectors regarding the various issue of data privacy and data protection to the maximum extent. All these issues can be solved using Quantum Key Distribution which is also efficient apart from solving all these issues and problems. The various systems available to us may be easy to design and efficient and thus depend upon the various interactions of the user and the data keeper located in the environments.

1.4. Disadvantages of Current Big Data Analysis

- Coping up with the Speed requirements:

In today's super competitive society, the organizations do not have any time to analyze and secure the data, whatever they have to do, it should be completed in no-time otherwise the world is very harsh. Either you do it or someone else will take your place. Realistic and Real time approaches helps the companies to market analysis and helps them to make useful decisions much quickly and at a rapid pace, but the main threat is accessing and implementation of such huge volumes of and then giving the required speed to the market. This challenge only grows bigger in size and becomes harder to implement. Now the quickest solution is hardware complexity. Some brands are using high speed processors and huge data centers to cope up with the incoming data that is available to them and thus solves the problem in some cases. Another method is putting the information in process wherein many system architecture and programmable machines are frequently used to solve the issues. Both of the ways allows the company to implement data analysis into the next level and thus defends the organization from the various issues related to speed and time.

- Realization of the data:

For the organizations, it takes a lot of effort and man power to understand the data and thus real time visualization comes in the much later part which is the main cause of the faults in the companies. For example, for the analysis of data from social networking sites, you must know the potential user for such the interactions may be different – and you have to know what you want to analyze from the customer. Now without a specific reason or value, implementation of tools is mostly to be of no use to the customer.

The main solution in this place is to have a specialized reason and set to fulfill the requirement needs of the user from the data analysis. We have to make sure the people implementing the needs have a deeper and knowledgeable understanding of the information, and will therefore help to sort out the challenges. If we implement this architecture, we can ensure communication privacy between the potential users and the validation server which will help to increase the overall Security.

- Data Review Quality (DRQ)

Now the point is even if we find the sequential data and we analyze the data accordingly for the user who will be using the information or product, the value of data will be effected in a bigger manner if the time constrain is much higher than its counter parts. This problem was and will be a major drawback with any data mining procedure, and when we consider the bulk volumes of incoming data the system becomes more difficult to handle. Thus, DRQ plays a very important role in this factor. Again, data analysis will only be a valuable way if the data quantity is made sure. To take care of this issue the organization needs to have advanced technology and specific people in their department to handle this situation.

- Displaying the Desired Output:

Plotting the graphs and pointing the situations can be a very difficult task when the data incoming is very high and the data volume become significantly large. For example, if you are having 100 million columns of sales data that you're trying to analyze. The manager who is wanting to see 100 million data values on the go should have a very hard time. The solution to solve this problem is to make specific clusters of the incoming data into multi-level orientation as if the less high groups of information become easily visible.

1.5. Results

In this paper, which we made, quantum key distribution provides the highest level of processes and protection with less time as well as space complexity that improves the storage quality and space keeping all the security concerns of data in mind. Here, we need to visualize the effectiveness of symmetric data-key pair with a Cipher text block which is useful to handle the security threats of the data since the making of the design, of the block is made very simple yet efficient. We know that the Complexity grows to a higher level when we use huge blocks but if we want we can drastically reduce the inbuilt required steps depending upon our needs. The cipher block which we are using implements the use of Germanium which made the efficient search of the keys for the user and in turn is one of the best quantum key distribution logic in data security methods. In this Q.K.D procedures and mechanisms, ciphers models which are specifically designed with germanium are very handy to make a useful key management system algorithm for potential data locators. Let us take any cipher model to be of key size 'k'. So, we come to the equation:

In (I) we see that it minimizes the procedures and thus time complexity when the Quantum key is made in the mechanism:

$$\text{Complexity } O(2^{k^{1/2}}) \quad (1)$$

However, in the primitive key management system, this following formula (I) had to use for many possible mechanism steps, which in turn required more time and space complexity.

$$\text{Complexity } O(2^k)$$

1.6. Quantum Model or the Kill. M Layer

The Valid key model which is very much required to set-up a secured and protected link between the customer and the nodal server can be implemented using this Kill Model. In this paper, we have analyzed and gave the Quantum Kill Model algorithm which is very useful to disarranged layers of data, the main reason as the process takes only two complete mechanism procedures. This proposed Kill model makes sure that the data is secured and privacy is maintained well between the layers among itself which will be beneficiary for the sole purpose and the model verifies that the efficiency of such algorithms and procedures are much useful in terms of complexity in space and time which is of the utmost importance.

Layers in This Kill Model

Proposed Layer 1: Location of the Data Locator (front end)

Let us consider the fact that data transmission i.e. exchange of data takes place between the DL-Data Locator and proposed/theoretical model that provides the authentication for the big data security protocols. Now when the active potential customers send and transmits big data contained with TB (1024 Gigabytes) to ZB (1024 Terabytes) to any data locator, this proposed model will be able to detect as well analyze the incoming data properly and in no time. This proposed model will detect the conditions and identify the moving user and the incoming data using QKD and authenticated security algorithms. The analyzed data such as private data or litigate information are then processed and initialized according to the 3V's of the data.

Proposed Layer 2: Data reading and receiving layer

In this following layer, the channel or the medium between the data locator and the moving users must be

transparent or wireless functioning wherein many conditions like security threats are taken in consideration. When the raw data in huge Volume enters into this model, the various processes control the ongoing operation within this layer during the Kill layer processing takes place.

Proposed Layer 3: Quantum key processing layer

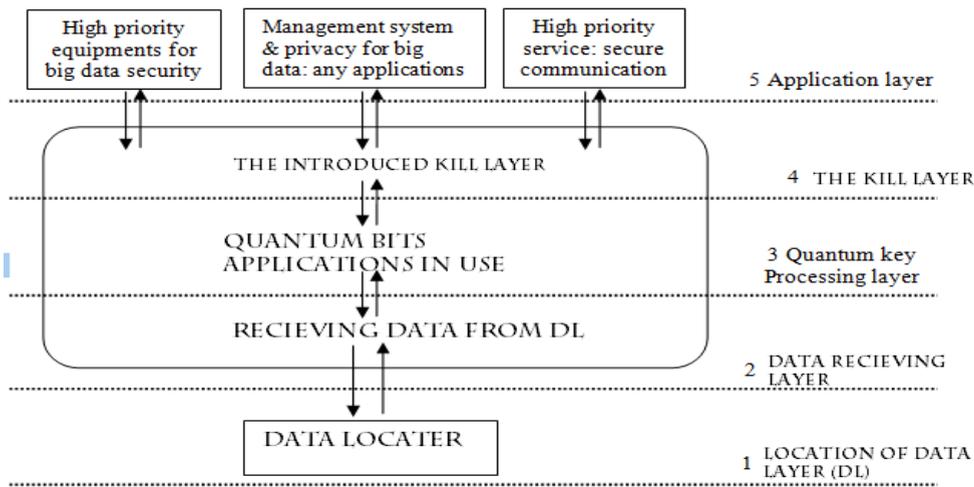
The Quantum key processing or QKD layer distribution surely depends upon a wide number of factors like the type of algorithm used, volume size of the data and the anticipated level of the security protocols which depends on the various types of the data available. But a condition is required which is that the Data Locator must be able to take multiple numbers of data through various forms and through the various ports which improve the quantum key layer.

Proposed Layer 4: The KM layer

Depending upon the quality and quantity of the incoming raw data this Kill layer verifies and secures the communication layer thereby creating or generating specific security data keys which are then to be transferred to the next layer for further processing of the data to be used. The various algorithms based on quantum key distribution are applied in this process to ensure secure data transfer.

Proposed Layer 5: The Application Layer

The various security concerns of data which depends upon the implementations of the applications in which data should be made secure with the highest priority satisfying the company rules and regulations which thereby takes care of the management details and handle with the various levels of the security and protection.



2. Why Implement the Kill Model?

1. Kill Layer encoding ensures Security of the Data

Typically, the raw data is most dangerous when it is transferred or moved from one location to another or through one layer to another layer via different channel and mediums. The Kill layer works when the data is

transported or at relocated from one place to another, thereby making it the best solution which one can have. This transfer doesn't depend upon how the data is stored or where the data is located which moreover reduces time and space complexity.

2. Maintains Integrity in all forms

The question of integrity comes when the valuable processed information gets changed or altered. Nowadays hackers not only steal data but also alter and modify it to get benefit from fraud cases. Now taking the point of view, if it is possible for skilled individuals to extract or modify the data any at point of time, it is also possible for the user or customer to detect any fraud or modified information from the organization which in turn will drastically effect the economic condition of the brand. The Kill layer ensures in maintaining the integrity of the data in such cases which of prime importance.

3. Encoding of the Kill layer protects Privacy

Encoding of this Kill layer can have utilized to protect and secure private information which are sensitive to many cases like for Example: Military Information. This layer in the long run therefore helps to ensure protection and privacy, thus reducing the changes of attraction of both criminals and competitors.

4. Encryption of the Kill Layer Protects Data across Devices and Introduction of IAVL

Multiple (and moving) devices make up a big part of life and transferring data or modifying valid data is a big risky task. Encryption of such technology can help to protect and secure data across various kinds of devices, even during the time of transfer and storage. There can be extra additional actions like Improved Authentication and Verification Layer (IAVL) which will help to determine valid and potential users.

3. Conclusion

Nowadays security of data is a major headache for most of the booming industries and market as Big Data has its advantages and a world full of issues and risks. If the environment or the organization can implement measures to ensure the security and privacy of such huge data, not only the company but the customer will also be benefitted. But as the world is moving on with huge and huge amount of incoming data which is available in the market during various kinds of transactions we need to find ways and measure to prevent any frauds in such cases. The proposed model has the major advantage of forming distinct layers which will protect the data from any kind of disruption and unsecured access and responds quickly which is the most beneficiary part of the whole research.

Acknowledgements

Firstly, I would like to Thank my Parents for supporting me thought the paper and my institution to help me with guidance. I would also like to Thank this Journal, without their help, I think this paper would not be so successful to represent.

References

- [1] ISSUES, CHALLENGES, AND SOLUTIONS: BIG DATA MINING Jaseena K.U.1 and Julie M. David.

- [2] MapReduce: Simplified Data Processing on Large Clusters Jeffrey Dean and Sanjay Ghemawat jeff@google.com, sanjay@google.com Google, Inc.
- [3] Big Data: Uses and Limitations Nathaniel Schenker Associate Director for Research and Methodology National Center for Health Statistics Centers for Disease Control and Prevention Presentation for discussion at the meeting of the NCHS Board of Scientific Counselors.
- [4] SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING Venkata Narasimha Inukollu¹, Sailaja Arsi¹ and Srinivasa Rao Ravuri³ Department of Computer Engineering, Texas Tech University, USA Department of Banking and Financial Services, Cognizant Technology Solutions, India.
- [5] Big Data for Supporting Low-Carbon Road Transport Policies in Europe: Applications, Challenges and Opportunities Michele De Gennaro | Elena Paffumi.
- [6] IOT Technique in Data minig in Analysis Dreaming Dr. Rajat Sen, Dr. M. Mano Department of CSE , UMI, India.
- [7] Quantum Data Mining in Organisations Behavior in Time Model T.N Dutta, S.W. John Cisco Inc.
- [8] Performance Analysis, Controller Selection And Verification Of Electric Motor For Mechatronics Motion Control Applications, Using New MATLAB Built-In Function And Simulink Model Ahmad A. Mahfouz, Farhan A. Salem.
- [9] Design Of A Optimized Parallel Array Multiplier Using Parallel Prefix Adder K.KalaiKaviya, D.P.Balasubramanian, S.Tamilselvan.
- [10] Design Artificial Intelligent Parallel Feedback Linearization of PID Control with Application to Continuum Robot Farzin Piltan, Sara Emamzadeh, Sara Heidari, Samaneh Zahmatkesh, Kamran Heidari.
- [11] Analysis of Resistance Parasitic of Single Wall CNT bundle with Copper for VLSI Interconnect Tarun Parihar, Abhilasha Sharma.
- [12] Research of Product Classifying and Coding Based on the Things Characteristics and Realization of the Relevant Management System Chen Xi, Hao Lijian, Yang Guichun, Zhang Jialia, Xia Guijiana, Ma Yanlia.
- [13] Performance Study on the System of Real-Time VBR Service with Shared Cache Hong-fei Zhang, Sheng-ye Huang.
- [14] Image Denoising by Nonlinear Diffusing on Mixed Curvature Gao Jian, Zhang Feiyan, Qin Qianqing
- [15] Research of Association Rule Mining Algorithm Based on Improved FP-Tree Chen zhuo, Lu nannan, Li shiqi, Han tao.

Authors' Profiles



Sayantan Gupta (Born 14.10.1997), Kolkata, India has keen interest in Data Analysis and Mining and as well as Quantum Key Distribution and has published many papers in Quantum Distribution. With a thoughtful mind, he has intentions to serve the world and make it sustainable for future needs and development.

How to cite this paper: Sayantan Gupta, "Security Analysis and Proposal of Kill Model using Modern Quantum Key Distribution in Big Data Mining", International Journal of Engineering and Manufacturing (IJEM), Vol.7, No.4, pp. 51-59, 2017. DOI: 10.5815/ijem.2017.04.05