A Thesis

on

A Novel Approach to Secure Big Data Using Attribute Based Honey Encryption

by

GAYATRI KAPIL

Department of Information Technology

Submitted in fulfillment of requirement of degree of

Doctor of Philosophy
to the

Babasaheb Bhimrao Ambedkar University
(A Central University)
Lucknow, Uttar Pradesh, India

January-2019

ABSTRACT

This era is called as digital era since we all are interconnected through digital devices and marketing services available round the clock. It is the world where Internet has made it possible for us to connect with anybody any time irrespective of the distance between them. Also, numerous companies related to various fields like banking, social networking, marketing, healthcare, defence etc. are relying to create huge amount of data for providing quality services. The amount of data is growing day by day because data is created by everyone and for everything from mobile devices, digital camera, digital game, multimedia sensor, video lectures and social networking sites, etc. Different studies have suggested that data will grow up to 44ZB in the world by 2020 wherein approximately 2.9ZB increase is predicted only in India. This abrupt growth of data may boom the market and provide abundant opportunities for midsize and small companies to enter into the big data market and also, explore this space.

Though increased of big data and its associated technologies offer numerous advantages over traditional technologies but one of the major issue is the security of big data. Big data includes personal and private information, so neglecting its security may lead to terrible consequences including bad reputation, financial problem and sometimes it compromises national security also. Therefore, it is obvious that security should be considered above all while storing and processing this large amount of sensitive information. Security doesn't mean only protecting the data in our network, but also monitoring the data continuously in a controlled way.

In this continuation, one of the most common platforms used to store and process large amount of data is Hadoop. Initially, when Hadoop was designed, the designers were mainly focusing only on the processing and management issues of the huge data. Security of data was not considered as important issue therefore, it was ignored almost completely. Initially users and services in Hadoop were not authenticated and Hadoop was designed to process and store large amount of data in a distributed cluster manner without considering the security of an individual machines. The major drawback in cluster distribution technology is that all users or programs have the same level of access to the clusters and they could read, write and delete any data set. To solve the issues, some traditional security mechanisms have been introduced to secure a small scale static data through authentication, authorization, editing and encryption within a cluster. Moreover, security professionals have started to think of more robust security system for Hadoop. Accordingly, security system of Hadoop has been improving since it was designed and Hadoop has become a more popular platform to store and process large amount of data. Various projects have been started to evolve the security of Hadoop.

Researchers & practitioners have identified and explored different level of security enhancements. But, these studies have been proven to be insufficient because of the growing issues repeated day by day. To date, the open source community has not addressed these security gaps, and remains focused on creating improved Hadoop technologies. The evolution of big data comes with the challenge to secure massive, streaming and increasingly private data. Also, big data exposes the industries as well as

individual to numerous data security threats. In addition, when public cloud is used as big data storage for analytics, the risk of security breach may also increases. That is, our major concern is to protect our private data from unauthorized access. To achieve this, various sophisticated techniques are available across the industry and the statistical result gives additional confidence to the user to migrate data and computations to the big data storage.

There are various cryptography techniques available in the markets which are based on basic encryption process. In the proposed research, the researcher has explored existing researches on cryptographic techniques to securing big data given by various researchers and practitioners. The available security techniques and approaches have some limitations like high storage and computation cost as well as complexity whereas some encryption techniques are suffering from brute force attack and crib matching. In this thesis, researcher tries to overcome these issues and proposed a new encryption approach for securing big data in HDFS.

The proposed encryption technique is based on attributes based Honey encryption which is an expansion of public key encryption that would allow the users to encrypt and decrypt messages based on user attributes. This will help to put attackers in the blacklist. The users who are white listed and are able to match with the given attributes will only be able to decrypt the encrypted messages. The proposed encryption approach provides dual layers security for stored HDFS data as well as the data in transmission. In this approach, initially we apply the attributes based encryption (based on cipher text policy based attribute encryption) on a particular file and further provide more security to that file through password

protection (which is based on Honey encryption). The selections of attributes in the proposed algorithm are quite flexible. There is no restriction on the number of attributes in the proposed algorithm and authority can add any number of attributes into the proposed algorithm as long as they have a uniform format. The access policy in our proposed algorithm is a propositional formula which is quite simple and easy to be computed. A policy is generated from the attributes using operation performed in random order to make it more secure.

Implementation of the proposed algorithm (ABHE) is done in two different platforms which comprise with Hadoop and without Hadoop environment. Firstly, the performance of the proposed encryption algorithm has been evaluated and its performance has been compared with the existing encryption algorithms namely AES, DES, and Blowfish with different size of text files vary from KB to MB. The performance parameters include encryption time, decryption time, encryption time throughput, decryption time throughput and power consumption. The simulation results concluded that the proposed algorithm (ABHE) has better performance than the existing algorithms.

After that, the proposed algorithm has been integrated with Hadoop environment. Various performance parameters have been compared with AES & AES with OTP algorithms with different sizes of text file vary from MB to GB on Hadoop cluster. In the experimentation, performance is evaluated based on input data size (increasing in gigabytes). The same has been compared against existing proposed algorithms namely AES and AES-OTP with the proposed one.

Validation of proposed algorithm has been done with the two different data sets i.e. (integration with and without Hadoop environment) the proposed approach justifies the robustness of conceptual approach. For validation of the proposed algorithm, Student t- test (t-Test: Paired Two Samples for Means) is carried out. Statistical observation t-Statistic value (for both data sets) is larger than absolute t critical value. Since the p value (for both data sets) is less than value of alpha, 0.05. Hence, alternate hypotheses at a very good level of significance for both data sets are accepted and hence the approach/framework.

A Thesis

on

A Novel Approach to Secure Big Data Using Attribute Based Honey Encryption

by

GAYATRI KAPIL

Department of Information Technology

Submitted in fulfillment of requirement of degree of

Doctor of Philosophy
to the

Babasaheb Bhimrao Ambedkar University
(A Central University)
Lucknow, Uttar Pradesh, India

January-2019