ABSTRACT

The term of big, data is an important concept that represents the exponential growth of volume and variety of data collection, one of the advantages of this technology is the ability of treating heterogeneous data, such as textual documents, also as its name indicates ‘big data’ refers to the huge volume of data counted by petabytes which implies an information retrieval extension to help users to find their need this extension must incorporate other protection against existent threats. In fact, big data services such as cloud computing do keep traces about user activities and queries, which compromise the user privacy and can offer useful information to network hackers that attack users or even cloud server to adapt or personalize their platforms without the user’s agreement, and maybe the most known attack can be man-in-the-middle during a storage or extracting data session between a user and a cloud server for this cause, the need to a secure protocol represent a high necessity which gives birth to the concept of Private Information Retrieval (PIR), as the authors mention before, one of the checks is the vast mass of data that hinders the correct handling of data and increases the error rate for retrieving a relevant information, for that the use of new techniques and approaches that allow the improvement of retrieval models over this kind of services is an important case to be processed. In this purpose, the authors introduce a new proposition called Meta-heuristic Privet Information Retrieval (M-PIR) in order to benefit from the success of meta-heuristics methods and improve the efficiency of PIR protocols in term of returned information; to better meet the needs of users, they use a bag of word for the text representation, TFIDF as weighting for the digitalization, the benchmarking MEDLINE corpus for the experimentation and panoply of validation tools (Recall, Precision, F-measure and Entropy) for the evaluation of our results. So that the paper is over the application of a meta-heuristics algorithms on a set of PIR protocols using a multitude of cryptographic schemes in order to study the influence of these schemes on quality of results.

Keywords: Big Data, Cloud Servers, Cryptographic Schemes, Meta-Heuristics, Private Information Retrieval Protocols, Validation Tools

DOI: 10.4018/ijcace.2014070101
INTRODUCTION AND PROBLEMATIC

The concept of data-driven decision-making is now known broadly over the world, especially with the coming of big data and its promises, in those days there is a big enthusiasm to this technology because of its real goals even if there is currently a wide gap between its theoretic concepts and goals and its realization, people upload more than 2.5 quintillion bytes of data per year.

In order to get much better understanding our work, we like to give you a complete grasp about the big data and its concepts, especially the cloud computing, simply the cloud computing is a concept of outsourcing the data from local machines to a set of virtual servers situated at distance by presenting the processing of those data as web services which guarantee an economical useless in term of capacity of material required to the process and also in term of financial concepts by basing on rule of pay what you use. Otherwise, the cloud computing is a model of sharing resources over the web using web services in order to enable convenient on demand, we talk about developing applications without the need to install a recommended platform and storing data without need to big spaces in the computer using storage as a service(Greer, 2012; Li, 2012).

Concerning the concept of big data according to authors in (Jinbao, 2013), it is as a simple of an abstract layer used in order to give a visual way to manage stored data of multiple structures and formats over global storage devices in a fundamental architecture like what Figure 1 indicates.

Privacy, timeless, scalability of data is the most important problems that big data recognize starting from the first step of data acquisition; in fact search information over big data is like searching for a needle in haystack because of the huge volume of data, nowadays, relevance information has become a crucial problem that’s why the construction of an efficient private information retrieval model (PIR) is a challenge in the middle of computer sciences and most of classical methods present several problems such as:

Figure 1. General architecture of big data (Jinbao, 2013)
A Service-Oriented Approach for the Optimal Product/Service Design Business Process
www.igi-global.com/article/service-oriented-approach-optimal-product/75560?camid=4v1a