Chapter 66
Bio-Inspired Private Information Retrieval System Over Cloud Service Using the Social Bees' Lifestyle With a 3D Visualisation

Hadj Ahmed Bouarara
Tahar Moulay University of Saida, Algeria

Reda Mohamed Hamou
Tahar Moulay University of Saida, Algeria

Amine Abdelmalek
Tahar Moulay University of Saida, Algeria

ABSTRACT

In the last decade, a new paradigm had seen the light named Cloud Computing, which allows the delocalization of data and applications on a dematerialized infrastructure accessible from Internet. Unfortunately, the cloud services are facing many drawbacks especially in terms of security and data confidentiality. However, in a world where digital information is everywhere, finding the desired information has become a crucial problem. For the purpose to preserve the user privacy life new approaches and ideas had been published. The content of this chapter is a new system of bio-inspired private information retrieval (BI-PIR) using the lifestyle of social bees, which allows both to find and hid, the sensitive desired information. It is based on a multi-filters cryptosystem used by the server for the encryption of stored document and the retrieval model using a combination of filters by 3 types of workers bees (Purveyor, guardian and cleaner), the queen bee represents the query, and the hive represents the class of relevant documents. We have tested this system on the benchmark MEDLINE dataset with panoply of validation tools (recall, precision, f-measure, entropy, silence, noise, and accuracy) and a comparative study had been realized with other systems existed in literature. Finally, a 3D visualization tool had been developed in order to make the results in graphical format understandable by humans. Our objectives is to improve the services quality of cloud computing.

DOI: 10.4018/978-1-5225-5191-1.ch066

INTRODUCTION AND BACKGROUND

Big data, Cloud Computing, and Data-Driven Decision-making are new concepts that had seen the light recently within the same area, the role of web in peoples' life. With more than 2.5 quintillion bytes of data created every day over the web; those concepts become a highly active domain of research in order to improve their services in many viewpoints such as security aspect (Aljawarneh, 2011).

The homomorphic encryption (HE), is the most useful security concept that is used in privacy preserving domain. With his ability of executing operations on ciphered data without need to decipher it. It gives opportunities to different domains starting from code obfuscation in the 90's until the PIR protocols in the recent few years. The HE uses complex mathematical concepts, such as bilinear applications and prime numbers theory, but this did not prevent the attackers and cryptanalysts to find weaknesses. In fact, some researches proved the inefficiency of this kind of schemes against hackers like in (Hervé, 2005), the authors presented a linear attack against TSZ cryptosystem called Anonymous Pirate Decoder (APD).

The concept of cloud computing offers a set of advantages, such as low cost payment for hard capacities of resources, and no need for cracks, patches or special software to access services…etc. In fact, cloud services present several problems and disadvantages, reside in the fact that it did not provide, neither information about data localisation during treatment (computing or storage), nor physical access to those data which pose problems in term of data confidentiality. As consequences of that, the evolution of cloud computing knows some criticism from users, especially the commercial ones concerning the security of computing and storage treatments.

However, the cloud services know a continuous growing in term of number of users and amount of data available online/offline that is counted by petabytes. The researches proved that 90% of the quantity of data stored on the web was created in last few years, precisely after the introduction of cloud computing. As result of that, the presence of system retrieval information SRI has becomes paramount, defined as a response to the need of user (query), by searching a sub set of documents that are closest to the query, from a wide range of documents that is useful to a claimant. The SRI is the basis of several application domain such as web, forums, and business. But the construction of an effective SRI, is a big challenge in the middle of the computer science and most of the classical methods present a several problems:

- **Diversity of Data:** The non-stable performance of the classical techniques caused by the multiplicity of data as multimedia, biological, medical, marketing, character recognition ... etc.
- **In Terms of Quality of Performance:** The classical algorithms are based only on simple functioning, for this reason they don't find the relevant documents.
- **The Difficulty to Select Parameters:** (Texts representation method, distance measurement metric), caused by the multitude of methods existed in the literature
- **Execution Time:** View the number of texts available on the web, the classical methods based on a single agent cannot analyse a large number of texts within a reasonable time.

Nature has a hundred million of experiments but humans have some years of industry, and consequently, draw inspiration from what nature has found (bio-mimicry technique), it's something interesting. The genius of nature is all around us, often in animals or plants that we encounter every day. There are three levels of bio-mimicry, we can inspire from forms, manufacturing process or ecosystems.

The first part of this chapter, is the development of a new bio-inspired private information retrieval system (BI-PRI), based on the social life of bees and their behaviour, using the distribution of tasks to