ARC'18

مؤتمر مؤسسة قطر
السنوي للبحوث

QATAR FOUNDATION
ANNUAL RESEARCH CONFERENCE

البحث والتطوير: التركيز على
الأولويات، وإحداث الأثر

R&D: FOCUSING ON PRIORITIES,
DELIVERING IMPACT

19-20 مـــارس
19-20 MARCH

مؤسسة قطر
Qatar Foundation
لإطلاق قدرات الإنسان.
Unlocking human potential.

Computing & Information Technology - Paper Presentation

http://doi.org/10.5339/qfarc.2018.ICTPP735

Substring search over encrypted data

Abdullatif Shikfa*

KINDI center for computing research, College of Engineering, Qatar University
* ashikfa@qu.edu.qa

Our data, be it personal or professional, is increasingly outsourced. This results from the development of cloud computing in the past ten years, a paradigm that shifts computing to a utility. Even without realizing it, cloud computing has entered our lives inexorably: every owner of a smartphone, every user of a social network is using cloud computing, as most IT companies and tech giants in particular are using infrastructure as a service to offer services in the model of software as a service. These services (dropbox, google, facebook, twitter...) are simple to use, flexible...and free! Users just send their data and they get all services without paying. Actually, these companies are making most of their revenues by profiling the users thanks to the data that the users willingly provide. The data is the indirect payment to benefit from these services. This raises privacy concerns at the personal level, as well as confidentiality issues for sensitive documents in a professional environment. The classical way of dealing with confidentiality is to conceal the data through encryption. However, cloud providers need access to data in order to provide useful services, not only to profile users. Take a cloud email service as example, where the emails are stored and archived in the cloud and only downloaded to the user's phone or computer when the user wants to read them. If the emails are encrypted in the cloud, the cloud cannot access them and confidentiality is enforced. However, the cloud can also not provide any useful service to the user such as a search functionality over emails. To meet these conflicting requirements (hiding the data and accessing the data) a solution is to develop mechanisms that allow computation on encrypted data. While generic protocols for computation on encrypted data have been researched developed, such as Gentry's breakthrough fully homomorphic encryption, their performance remains unsatisfactory. On the contrary, tailoring solutions to specific needs result in more practical and efficient solution. In the case of searching over encrypted data,

دار جامعة حمد بن خليفة للنشر
HAMAD BIN KHALIFA UNIVERSITY PRESS

Cite this article as: Shikfa A. (2018). Substring search over encrypted data. Qatar Foundation Annual Research Conference Proceedings 2018: ICTPP735
http://doi.org/10.5339/qfarc.2018.ICTPP735.

searchable encryptions algorithms have been developed for over decade and achieve now satisfactory performance (linear in the size of the dictionary). Most of the work in this field focus on single keyword search in the symmetric setting. To overcome this limitation, we first proposed a scheme based on letter orthogonalization that allows testing of string membership by performing efficient inner products (AsiaCCS 2013). Going further, we now propose a general solution to the problem of efficient substring search over encrypted data. The solution enhances existing “keyword” searchable encryption schemes by allowing searching for any part of encrypted keywords without requiring one to store all possible combinations of substrings from a given dictionary. The proposed technique is based on the previous idea of letter orthogonalization. We first propose SED-1, the base protocol for substring search. We then identify some attacks on SED-1 that demonstrate the complexity of the substring search problem under different threat scenarios. This leads us to propose our second and main protocol SED-2. The protocol is also efficient in that the search complexity is linear in the size of the keyword dictionary. We run several experiments on a sizeable real world dataset to evaluate the performance of our protocol. This final work has been accepted for publication in the IOS journal of computer security https://content.iospress.com/articles/journal-of-computer-security/jcs14652.