All about that Data: Towards a Practical Assessment of Attacks on Encrypted Search*

Seny Kamara
Brown University

Abdelkarim Kati
Mohammed-VI Polytechnic University

Tarik Moataz
Aroki Systems

Thomas Schneider
TU Darmstadt

Amos Treiber
TU Darmstadt

Michael Yonli
TU Darmstadt

1 Motivation

Motivated by calls for privacy and data breaches of cloud services, efforts to broadly deploy Encrypted Search Algorithms (ESAs) are moving forward. ESAs allow search on encrypted data and can be found in research as well as industry, e.g., Microsoft's Always Encrypted. They are built using various techniques, which represent complex tradeoffs between efficiency, expressiveness, and security.

Concerning the security aspect, ESAs leak different information, e.g., whether and when a query repeats. To assess how exploitable specific leakage is, researchers designed *leakage attacks* that try to uncover the encrypted queries or the encrypted database. They represent significant advances for understanding leakage and achieve increasingly lower errors with increasingly less information. While their evaluations showed that they can be very successful for certain classes of user queries (query distributions), they did not incorporate *which classes of queries would actually be performed in a deployed, real-world system*. This is because real-world query data is extremely hard to find and, therefore, it is unclear how susceptible real-world systems are. Furthermore, reliance on a restricted set of datasets, e.g., the Enron e-mail dataset for keyword search, and overwhelmingly closed-source implementations leave an uncertainty how reproducible results are across datasets. So while practitioners can usually lay out practical requirements on the tradeoff aspects of efficiency and expressiveness, the practical implications of the aspect of security remain open. Thus, *a practical assessment of attacks on encrypted search* incorporating real-world query data and a breadth of use cases remains a well-known challenge for understanding leakage and, hence, enabling widespread *adequate* deployments of ESAs.

*This is an abstract for the talk given by Amos Treiber (`treiber@encrypto.cs.tu-darmstadt.de`) at the Real World Crypto Symposium 2022. Please cite the conference version of our work [1].

2 Our Work

With our work [1], we address these limitations with new software, more realistic datasets, and broad evaluations as an important step towards the necessary practical assessment. Motivated by the prevalence of closed-source software, we re-implement major leakage attacks for encrypted keyword and range search in a new open-source framework called LEAKER*. With LEAKER, attacks can be easily integrated and evaluated on arbitrary data, allowing researchers and practitioners to focus more on studying the effects of leakage. For more realistic evaluations, we uncover many new datasets in a variety of use cases that—for the first time—include query data. Armed with this, we perform an extensive re-evaluation of leakage attacks and show that in many cases their practical risk is not as expected. For example, some attacks on encrypted keyword search using identifier or individual volume leakage work significantly better than expected with very little adversarial knowledge. In these cases, our evaluations display a disparity with the previous assumption that users would query for infrequent keywords, whereas our real-world data suggests otherwise. It is thus important to hide this leakage in deployed systems.

3 This Talk

In this talk, I will present the results and implications of our work [1]. I will start with a quick overview of ESAs and their applications, leakage attacks and their evaluations, and the resulting cryptanalytic challenges *in real-world conditions*. My talk will then highlight the important takeaways of our more realistic evaluations relevant to practitioners from industry and researchers from academia who want to broadly deploy ESAs. In addition, my talk will focus on bringing practitioners and researchers closer together, as in this domain they can greatly benefit from each other.

For *researchers*, I will use the mentioned challenges to motivate LEAKER as well as our (and hopefully subsequent) datasets as future benchmarks to continually advance our common understanding of leakage. This also further benefits industry because reference tools for research results are necessary for standardization processes like NIST's recent *Privacy-Enhancing Cryptography* project†.

Despite our considerable efforts, the evaluations and conclusions of our work are far from final and still suffer from the fact that for increasingly practical studies of attacks more (especially query) data is desperately needed, which is largely unavailable to researchers. Thus, targeting *practitioners*, I will also use the mentioned challenges to foster and motivate the necessity of collaborations with researchers. With LEAKER, attacks can easily be evaluated locally by practitioners without violating the privacy of subjects. These evaluations would greatly benefit the research community since they would encompass the more realistic data that only practitioners have access to. Such collaborations for researching search systems are already common in the information retrieval community and similar analyses are required here for researchers to conduct evaluations truly applicable to real-world conditions. I therefore hope that with this talk and call for collaborations I can likewise bring industry and researchers working on encrypted search systems closer together towards practically assessing their security.

*Code is available under the MIT open source license at `https://encrypto.de/code/LEAKER`.

†`https://csrc.nist.gov/projects/pec`

References

[1] Seny Kamara, Abdelkarim Kati, Tarik Moataz, Thomas Schneider, Amos Treiber, and Michael Yonli. SoK: Cryptanalysis of Encrypted Search with LEAKER - A framework for LEakage AttacK Evaluation on Real-world data. In *IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2022.