Strong Code Obfuscation Algorithms and Their Applications – Final Report –

Project leader: Dr. Levente Buttyán
NKFI-id.: 116675

January 30, 2020

1 Background of the Project

1.1 Motivations

Obfuscating programs or with other words hiding information in algorithms with precise security guarantees is a major open problem in software engineering that cryptography has not been able to solve yet. The landmark result of Garg et al. [GGH+13] from 2013 meant the first step towards the solution. Their first general-purpose obfuscator candidate turned out to be an extremely useful tool that helped to answer several open questions in cryptography. At the same time, the practical applicability of these new results is limited by two critical factors. First, the extreme complexity and inefficiency of both the first and the subsequent general-purpose obfuscator candidates (see [B1]); second, the uncertainty around the security of obfuscation originating from the use of untested assumptions, that still needs to be eliminated from the constructions (if this is possible at all).

The sketched situation motivated many other researchers and us to investigate the practical applicability of the immense number of new ideas that emerged from the birth of cryptographic obfuscation.

1.2 Minor Deviation from the Work Plan

One of our initial goals was to use obfuscation to solve a problem related to user revocation in attribute-based encryption (ABE), which is a type of functional encryption (FE). We found that in order to use obfuscation for this specific goal, it would be crucial to utilize so-called probabilistic indistinguishability obfuscation [CLTV15] that relies on the obfuscation of general programs. It turned out that we cannot eliminate or relax this requirement, and we would need the heavy machinery of general-purpose obfuscation instead of being able to rely on obfuscation of some restricted functionality. As the resulting inefficiency would have contradicted the aimed practicality of the desired solution, we decided[1] to turn our attention towards

[1] Our report on the modification was accepted on 19. 10. 2017.

another subclass of FE, called searchable encryption (for the summary of our related results see §2.3).

2 Summary of Results

In the time frame of the project we managed to accomplish two goals that we described in our research proposal. Our results were presented

- at two conferences and a workshop, while the corresponding works appeared among the revised and selected papers in conference proceedings [C1, C2, C3] (as planned),
- in one journal paper [J1] (instead of the planned two journal papers),
- and in one book [B1] (instead of one of the planned journal papers).

Next, a brief overview of the results follows.

2.1 Systematization of Knowledge

In order to gain a deeper understanding of the state of the art of cryptographic obfuscation, we started the project with an extensive literature survey. This work turned out to be harder and more important than we thought initially. After the publication of the first candidate obfuscator of [GGH$^+$13], obfuscation became a central hub of cryptographic research for the next couple of years. By 2017, over 190 related papers dealt with the topic, while previously, this number was below 30. After a crystallization, new results again started to appear less often. The sudden improvement caused that it became very difficult to follow the developments and to gather the key thoughts from the vast number of papers, which themselves were looking for the right definitions, methods, and formulations. In [B1], we systematized the different results on obfuscator constructions, providing a unified view and discussing the different methods that use tools from almost all areas of cryptography.

Related to this work, a popular science article [M1] was also prepared (in Hungarian) in the frame of the project, which was dedicated to a wider audience outside the research community.

2.2 Data Market Security

One of the insights we gained from the preparation of [B1] was that the protection of simpler computations used in practice is hard to achieve by scaling down the complex techniques of general-purpose obfuscation. Indeed, currently there are only a few functionalities that are known to be obfuscatable without the heavy machinery of general-purpose techniques, and these are limited to such simple functions as point functions or pattern matching with wildcards (a.k.a. conjunctions) [BW19]. Because of this, our approach was to utilize FE, a key building block of obfuscators (see [B1, §4.3, §6]) to protect practical functionalities.

As an application domain, we considered data markets that are designed for helping the monetization of data generated by smart devices and the internet of

things (IoT). We analysed data markets in [C2] based on the possible security requirements of the involved participants. In this work, we identified more than 30 possible scenarios and connected them to the relevant areas of cryptography. Our analysis highlighted several open problems motivating further research on cryptographic primitives related to function protection.

In a subsequent work [C1], we focused on one of these open problems and asked the question whether it is possible to achieve a reasonable level of input and function privacy simultaneously in two-party private function evaluation protocols (PFE). PFE enables two parties to jointly execute a computation such that one of them provides the input while the other chooses the function to compute. According to the traditional security requirements, a PFE protocol should leak no more information, neither about the function nor the input, than what is revealed by the output of the computation. Existing PFE solutions inherently restrict the scope of computable functions to a certain function class with given output size, thus ruling out the direct evaluation of such problematic functions as the identity map, which would entirely undermine the input privacy requirement. We observed that when not only the input x is confidential but certain partial information $g(x)$ of it as well, standard PFE fails to provide meaningful input privacy if g and the function f to be computed fall into the same function class.

In [C1], we initiated the study of partial information protection in the context of PFE. As a first step, we put forward a definitional framework to assess security and proposed the notions of Controlled PFE (CPFE), and a relaxed version of it (rCPFE) that guarantees weaker (but still reasonable) k-anonymity style function privacy allowing a trade-off between security and efficiency. We showed conceptually simple, generic realizations of both CPFE and rCPFE. In the latter case, we utilized the relaxed function privacy guarantee (through using FE) to enable the reusability of the protocol messages in case of multiple function evaluations. Concretely, when evaluating the same function(s) on multiple, say d inputs (which scenario is typical in a data market), the communication and online computation overhead of our rCPFE protocol only increases with an additive factor proportional to d instead of a multiplicative factor as in ordinary PFE.

To demonstrate the practicality of the rCPFE approach, we instantiated our generic protocol for the inner product functionality enabling secure statistical analysis in a controlled manner under the standard Decisional Diffie–Hellman (DDH) assumption. Our proof of concept implementation shows that the reusability property indeed results in a significant performance improvement over the state of the art secure inner product evaluation method of [DSZ15].

We note that the relevance of our application scenario was indirectly confirmed by the establishment of the Hungarian Artificial Intelligence Coalition[2]) in October, 2018 that has initiated the development of a Hungarian data market[3] following other international examples[4].

[2]See `https://digitalisjoletprogram.hu/hu/tartalom/mesterseges-intelligencia-koalicio` (accessed 30-01-2020).

[3]See `https://www.datamarket.hu/` (accessed 30-01-2020).

[4]For example `https://datamarket.at/en/` (accessed 30-01-2020).

2.3 Secure and Searchable Cloude Storage

One of the basic computations that is often required over hidden data is keyword search, which is essentially the comparison of encrypted values. Searchable symmetric encryption (SSE) allows the secure storage of sensitive data on untrusted servers in the cloud without losing all the flexibility that plaintext data would allow. More precisely it supports keyword search over the ciphertexts in the following way: encrypted queries called trapdoors can be sent to the server which can test whether any of the stored ciphertexts matches the keyword underlying the trapdoor. In [C3], we focused on certain scenarios in which search over the whole database is not necessary and showed that the otherwise inefficient sequential scan (in linear time) can be very practical. This is due to the fact that adding new entries to the database comes for free in this case while updating a complex data structure without information leakage is rather complicated. With this approach, we built an SSE scheme based on bilinear pairings and proved its security against adaptive chosen-keyword attacks (IND-CKA2) in the standard model under the widely used symmetric external Diffie-Hellman (SXDH) assumption. Using the FE terminology, our result correspond to a function private, secret key FE scheme for point functions.

This result was generalized in [J1] by proposing an SSE construction (under the SXDH assumption in the standard model) that is built from an arbitrary semantically secure symmetric-key cipher and a message authentication code with EU-CMA security (i.e. that is existentially unforgeable under chosen message attacks). This modular approach allows for flexible choices of the underlying primitives helping the adoption of the scheme to concrete applications. We note that a follow-up work [Vaj19] further extended our results.

Publications Related to the Project

Conference and Workshop Publications

[C1] Máté Horváth, Levente Buttyán, Gábor Székely and Dóra Neubrandt. There Is Always an Exception: Controlling Partial Information Leakage in Secure Computation. *J. H. Seo (Ed.): Information Security and Cryptology – ICISC 2019, LNCS 11975, pp. 1—17, 2020.*

[C2] Máté Horváth, Levente Buttyán. Problem Domain Analysis of IoT-Driven Secure Data Markets. In: E. Gelenbe et al. (eds) *Security in Computer and Information Sciences. Euro-CYBERSEC 2018. Communications in Computer and Information Science*, vol 821. pp. 57-67, Springer, 2018.

[C3] Máté Horváth, István Vajda. Searchable symmetric encryption: Sequential scan can be practical. *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp 1–5, 2017.

Journal Publication

[J1] Máté Horváth, István Vajda. Searchable Symmetric Encryption for Restricted Search. *Journal of Communications Software and Systems*, 14(1):104–111, 2018. Quartile ranking: Q4.

Book

[B1] Máté Horváth, Levente Buttyán. Cryptographic Obfuscation: A Survey. *Springer Briefs in Computer Science, ISBN 978-3-319-98040-9.* Springer, Expected publication: 2020. The manuscript is available here: `http://eprint.iacr.org/2015/412`.

Miscellaneous

[M1] Máté Horváth. Programok titkai – A kriptográfiai obfuszkáció születése (in Hungarian). *In: Élet és Tudomány, 73:(34) pp. 1065–1067, 2018.*

[M2] Máté Horváth, Levente Buttyán. Controlled Private Function Evaluation from Functional Encryption. *Technical report*, 2018.

References

[BW19] Ward Beullens and Hoeteck Wee. Obfuscating simple functionalities from knowledge assumptions. In *Public Key Cryptography (2)*, volume 11443 of *Lecture Notes in Computer Science*, pages 254–283. Springer, 2019.

[CLTV15] Ran Canetti, Huijia Lin, Stefano Tessaro, and Vinod Vaikuntanathan. Obfuscation of probabilistic circuits and applications. In *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part II*, pages 468–497, 2015.

[DSZ15] Daniel Demmler, Thomas Schneider, and Michael Zohner. ABY - A framework for efficient mixed-protocol secure two-party computation. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2015*. The Internet Society, 2015.

[GGH+13] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 40–49. IEEE, 2013.

[Vaj19] István Vajda. Construction for searchable encryption with strong security guarantees. *International Journal of Computer Network and Information Security (IJCNIS)*, 11(5):1–10, 2019.