Understanding information leakage in searchable encryption

1. ABSTRACT

Information leakage in searchable encryption is poorly understood and it is often left to the users of systems to evaluate it. This travel grant will enable Prof. Warinschi to undertake collaborative work together with Prof. Alexandra Boldyreva (Georgia Tech) and Prof. Geoffrey Smith (Florida International University) to develop a principled approach to dealing with this type of leaks.

Specifically, the award will support three visits of Prof. Warinschi to meetings of the EAGER project awarded to Profs. Boldyreva and Smith by the NSF (NSF award #1749069).

2. PREVIOUS TRACK RECORD

Prof. Bogdan Warinschi joined the University of Bristol as a Lecturer in the Computer Science Department in February 2007. He has a Ph.D. in Computer Science from the University of California at San Diego, and was previously affiliated as a postdoctoral researcher with University of California at Santa Cruz, Stanford University, and the French National Research Institute in Informatics (INRIA). He conducts research into security and cryptography, with a particular emphasis on proof methods for the security of protocols.

Dr. Warinschi served on more than thirty program committees of top conferences, most recently on Security and Privacy '16, Eurocrypt'15 and Eurocrypt'14. He is internationally known for his roly in establishing the research area generically known as *computational soundness* [A8, A10]. The basic idea is to combine two fundamentally different paradigms for proving security, one based on symbolic (formal methods-like) techniques and one based on complexity and computation theory. The main benefit of the approach is that it enables simpler analysis which uses higher level abstractions yet which offers computational security guarantees.

In addition to his expertise in the area of computational soundness, this proposal relies on the expertise of Prof. Warinschi in the design of security models and his interest in analysing systems used in practice. In particular, he has maintained a constant interest in designing cryptographic security models, the first step towards the rigorous analysis of any cryptographic system. His research addresses the security of group signatures [A1], proxy signatures [A3], defence mechanisms for DOS resistance (cryptographic puzzles) [A7], key-exchange [A5], voting schemes [A2], and cryptographic APIs [A9]. This work is complemented by security analysis on protocols and primitives that are deployed. These include Public Key Infrastructures (PKI) [A4], the Trusted Platform Module (TPM) [A6], the Helios Internet voting schemes [A2], and the ubiquitous Transport Layer Security protocol (TLS) [A11].

His recent research interests are in understanding and enabling the use of cryptography in practical scenarios. This travel grant will support his participation in a project that falls within this broad direction.

Prof. Alexandra Boldyreva Prof. Boldyreva is an Associate Professor at Georgia Tech Institute for Information Security & Privacy. She is a world leading expert security of encryption schemes with additional functionality properties. In particular she has helped establish foundations for deterministic, order-preserving and searchable encryption [].

Prof. Geoffrey Smith Prof. Smith is a Professor at the School of Computing and Information Sciences, Florida International University.

References

[A1] Mihir Bellare, Daniele Micciancio, and Bogdan Warinschi. "Foundations of group signatures: formal definition, simplified requirements and a construction based on trapdoor permutations". In: *EUROCRYPT*. Vol. 2656. LNCS. Warsaw, Poland: Springer-Verlag, May 2003, pp. 614–629.

[A2] David Bernhard et al. "Adapting Helios for Provable Ballot Privacy". In: *ESORICS*. 2011, pp. 335–354.

[A3] Alexandra Boldyreva, Adriana Palacio, and Bogdan Warinschi. "Secure Proxy Signature Schemes for Delegation of Signing Rights". In: *J. Cryptology* 25.1 (2012), pp. 57–115.

[A4] Alexandra Boldyreva et al. "A Closer Look at PKI: Security and Efficiency". In: *Public Key Cryptography – PKC 2007*. June 2007, pp. 458–475. ISBN: 978-3-540-71676-1.

[A5] Christina Brzuska et al. "Composability of Bellare-Rogaway key exchange protocols". In: *CCS*. 2011, pp. 51–62.

[A6] Liqun Chen, Ming-Feng Lee, and Bogdan Warinschi. "Security of the Enhanced TCG Privacy-CA Solution". In: *TGC*. 2011, pp. 121–141.

[A7] Liqun Chen et al. "Security Notions and Generic Constructions for Client Puzzles". In: *ASIACRYPT*. 2009, pp. 505–523.

[A8] Véronique Cortier and Bogdan Warinschi. "Computationally Sound, Automated Proofs for Security Protocols". In: *ESOP*. 2005, pp. 157–171.

[A9] Steve Kremer, Graham Steel, and Bogdan Warinschi. "Security for Key Management Interfaces". In: *CSF*. 2011, pp. 266–280.

[A10] Daniele Micciancio and Bogdan Warinschi. "Soundness of Formal Encryption in the Presence of Active Adversaries". In: *TCC*. 2004, pp. 133–151.

[A11] Paul Morrissey, Nigel P. Smart, and Bogdan Warinschi. "The TLS Handshake Protocol: A Modular Analysis". In: *J. Cryptology* 23.2 (2010), pp. 187–223.

3. BACKGROUND

Cloud storage is currently experiencing explosive growth. Users regularly rely on the cloud to store business-critical information, and the average organization uploads 13.9 TB of data to the cloud each month [3]. With all of this data stored remotely, security of data is becoming a core requirement for any organization looking to leverage the public cloud.

In this setting, even if one is willing to trust the cloud, it is crucial to also guarantee security against intrusion attacks or dishonest employee actions. Ensuring this level of security without compromising functionality and efficiency is one of the major barriers in the face of wider cloud storage adoption, especially when compliance with legal requirements such as (the US regulation comprised by) PCI DSS and HIPAA, as well as privacy requirements under laws such as the EU Data Protection Regulation.

It is not hard to observe that simply using off-the-shelf encryption schemes does not "work" for cloud storage application. This is because standard encryption is so strong as to hide *all* poartial information about the plaintexts, so ciphertexts cannot be used even to search on the underlying plaintexts.

Computing on encrypted data The theoretical appeal, but especially the potential for practical impact, has attracted a lot of research in the area of computing on encrypted data, where a cloud is empowered to operate on data that is encrypted without getting access to said data. There are a few theoretical solutions to this problem, e.g. Oblivious RAM (ORAM) [33, 23, 35] or fully homomorphic encryption (FHE) [22]. These solutions are decades away from being sufficiently efficient.

A more practical approach proposes to design encryption schemes which leak *some* information, sufficient to enable running algorithms on top of encrypted data, while revealing as little as possible about what exactly is encrypted. The practical implications for secure solutions for this problem are staggering and have attracted a significant amount of research. Consider for example the rather restricted functionality offered by symmetric searchable encryption. There is by now a plethora of academic research on the topic []. The solutions proposed differ in the level of functionality, security and efficiency they provide: these properties are almost always at odds with each other. Most of the schemes address the exact-match keyword search, but there are also solutions for enabling sorting the encrypted data, performing conjunctive, Boolean, range, similarity, substring queries, etc. Some protocols require a user or a proxy to create an encrypted index of the data and send it to the server along with encrypted data. Some are in the public- and some are in the symmetric-key settings. Some solutions are transparent in that the server can handle ciphertexts in the exact same way it searched unencrypted data, so no server-side modifications are required. The existing solutions also vary greatly security-wise. This is not surprising given that one often must compromise security in order to achieve better efficiency or functionality.

The problem All of the above searchable encryption schemes and, more generally, schemes where information is leaked in exchange of improved functionality share a major shortcomming: the security implications of the information that is disclosed are only rarely clear. Very roughly, security definitions for such schemes formalize the information leaked as a function which maps the secret information that should be protected to a bitstring. A scheme is deamed secure if any attack against the scheme can be simulated by an adversary who only has access to the leaked information.

The leakage function is therefore a parameter of the definition but there are no restrictions imposed on it. Specifying the leakage function associated to a scheme is left to the scheme designer, and determining what are acceptable functions is left to the users of the scheme. For some functions, e.g. a function that returns the first field of a database record, it may be clear and relatively simple to determine if the risc is acceptable. However, only informal arguments where attacks that exhibit obviously insecure behaviour have appeared in the literature. What is missing is a principled approach to the problem of understanding and quantifying information leakage in this type of schemes.

4. RESEARCH PROJECT

The research hypothesis

A possible solution to this problem is rooted in work by the PI in the context of electronic voting [B4]. There, it is observed that no matter how well individual votes are proected, the result of the voting process necessarily leaks information about the votes. The simplest way to see this is in the case where the winner is by unanimity: here it is clear how everyone voted. The problem mirrors the one outline above for searchable encryption: the

information revealed by the result of the system is not accounted for by security definitions for privacy in voting systems. For voting systems this problem has been addressed by the PI [B4]. The main result shows that it is possible to use various notions of *computational entropy* to measure the information that is revealed to the adversary. Moreover, it is possible to exhibit a mapping between different notions of entropy (Hartley, min, average, conditional) and attacker models which clarify precisely what is the information that the attacker aims to obtain.

The idea behind this project is to

An important ingredient of this research project is a generalization of the traditional notion of entropy. In a series of papers co-authored by Prof. Smith [B3, B1, B2, B9] a generalization of the notion of entropy *g-leakage*, an approach to quantifying information leakage aking to entropy, but in a way that allow to identify a matching adversarial model. The US National Security Agency awarded in 2015 the Best Cybersecurity Research Paper

From this perspective g-leakage generalizes the use of entropy in measuring vote privacy [B4].

The suggestion of using the g-leakage approach to measure information leakage in encryption schemes was suggested by the PI to Prof. Boldyreva in 2014, who has later organized a preliminary 3-day meeting with the PI and Prof. Smith in April 2015. In the meeting several interesting research directions were identified; one of them was to use the setting of searchable encryption as a test case to

This line of research has recently been granted support by the US National Science Foundation who has recently awarded a grant for collaborative work (*EAGER: Collaborative: Quantifying Information Leakage in Searchable Encryption* – Award #1749069) for developing a principled understanding of information leakage in searchable encryption.

The broad goal is to extend the approach in [B4]

This grant will run between January 2018 and June 2019.

5. PAST AND ONGOING COLLABORATIONS

The the Prof. Warinschi and Prof. Boldyreva have been collaborating for more than a decade, and their collaboration resulted in several research papers [B6, B7, B5, B8]; some of these are highly cited. The PI and Prof. Boldyreva are currently undertaking joint work in the area of encryption for machine learning with one research paper under submission and another one in preparation.

References

[B1] Mário S. Alvim et al. “Additive and Multiplicative Notions of Leakage, and Their Capacities”. In: *IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19-22 July, 2014*. IEEE Computer Society, 2014, pp. 308–322. ISBN: 978-1-4799-4290-9. DOI: 10.1109/CSF.2014.29. URL: https://doi.org/10.1109/CSF.2014.29.

[B2] Mário S. Alvim et al. “Axioms for Information Leakage”. In: *IEEE 29th Computer Security Foundations Symposium, CSF 2016, Lisbon, Portugal, June 27 - July 1, 2016*. IEEE Computer Society, 2016, pp. 77–92. ISBN: 978-1-5090-2607-4. DOI: 10.1109/CSF.2016.13. URL: https://doi.org/10.1109/CSF.2016.13.

[B3] Mário S. Alvim et al. “Measuring Information Leakage Using Generalized Gain Functions”. In: *25th IEEE Computer Security Foundations Symposium, CSF 2012, Cambridge, MA, USA, June 25-27, 2012*. Ed. by Stephen Chong. IEEE Computer Society, 2012, pp. 265–279. ISBN: 978-1-4673-1918-8. DOI: 10.1109/CSF.2012.26. URL: https://doi.org/10.1109/CSF.2012.26.

[B4] David Bernhard et al. “Measuring vote privacy, revisited”. In: *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM. 2012, pp. 941–952.

[B5] Alexandra Boldyreva, Adriana Palacio, and Bogdan Warinschi. “Secure proxy signature schemes for delegation of signing rights”. In: *Journal of Cryptology* 25.1 (2012), pp. 57–115.

[B6] Alexandra Boldyreva et al. “A closer look at PKI: Security and efficiency”. In: *Public Key Cryptography–PKC 2007* (2007), pp. 458–475.

[B7] Alexandra Boldyreva et al. “Foundations of Non-malleable Hash and One-Way Functions.” In: *ASIACRYPT*. Vol. 5912. Springer, pp. 524–541.

[B8] Richard J Lipton, Rafail Ostrovsky, and Vassilis Zikas. “Provably secure virus detection: Using the observer effect against malware”. In: *LIPIcs-Leibniz International Proceedings in Informatics*. Vol. 55. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2016.

[B9] David M. Smith and Geoffrey Smith. “Tight Bounds on Information Leakage from Repeated Independent Runs”. In: *30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21-25, 2017*. IEEE Computer Society, 2017, pp. 318–327. ISBN: 978-1-5386-3217-8. DOI: 10.1109/CSF.2017.18. URL: https://doi.org/10.1109/CSF.2017.18.

A. JUSTIFICATION OF RESOURCES

The grant will support participation of the PI to three of the meetings of the EAGER projects. The estimated cost of one visit is 2700GBP:

- international travel: 1200 GBP
- hotel costs: 10 x 100 GBP = 1000 GBP
- subsistence: 10 x 50 GBP = 500 GBP

The total cost of three visits is 8100GBP.

B. PATHWAYS TO IMPACT

C. WORKPLAN