R210 Essay: Encrypted data systems

Chongyang Shi (*cs940*)

January 18, 2018

1 Summaries of research

In 2014, Popa et al. [1] described the design of a searchable encryption framework called *Mylar*, which allows application developers to integrate encrypted storage of data without significantly refactoring their code. Theoretically, stored data can only be read by the client, but remains searchable by the server on client request. The designed operational model also provides client-side code verification to protect the server from malicious clients. Mylar was proposed as a solution to some usability problems faced by encrypted data systems such as permission control, searchability and client verification. However, Mylar's design suffers from some significant drawbacks, such as the lack of access revocation, as well as stringent and potentially overbearing security requirements on application developers (e.g. difficulties in integrating complex access graphs) and users (e.g. user must verify HTTPS access).

Two years after Mylar was proposed, Grubbs et al. [2] published a paper to illustrate various flaws they discovered in Mylar through passive and aggressive attacks, based on compromised access of varying severity, under which Mylar should have provided security assurance by definition. They were able to infer some user information based on passively observed data in some experiments; and through the introduction of a malicious server, perform brute-force dictionary attacks to decipher a majority of encrypted data when given sufficiently large number of past user queries. These attacks compromised several security guarantees supposedly provided by Mylar's underlying model, prompting discussions on whether property-preserving encryption (PPE) schemes are fundamentally insecure.

Durak et al. [3] studied information leakage in order-revealing encryption (ORE), which is a similar type of encryption to PPE, but rather reveals ordering information of data to enable more efficient database operations at the cost of security. Through various attacks performed on two datasets encrypted with ORE, they were able to demonstrate that correlations between columns of data as well as known contextual information can reveal more information than previously modelled by various leakage profiles. Performing attacks developed in this research on more datasets also fitting the paper's description can be useful in further illustrating their effectiveness.

2 Key themes of research

2.1 Usability considerations sometimes hinder security assurance

As all encrypted data systems involve users that may or may not be technically proficient, the security of the system can be influenced by usability considerations. To improve user experience, a lot of applications do very little to validate the identity of the user beyond their login credentials, weakening the protection against active attacks offered by an integrated Mylar framework (since a trusted identity service provider will be unavailable in most cases) [1, 3.1]. As observed by Grubbs et al. [2, 3.3], it is also difficult to model updates to documents in Mylar. This is a feature required by users in most practical applications. They further observed [2, 8.1] that asking the user to confirm every shared document can be very damaging to usability, but is required by Mylar to prevent malicious server attacks. In the research by Durak et al. [3, Sec. 3], the ability for the user to enter correlated data over multiple columns (e.g. geographical coordinates) is often required in practice, but was found to reveal further information in inter-column correlation attacks on ORE systems.

2.2 Security degradation from specifications to their implementations

The implementation of a encrypted data system often deviate from its specification in practice, which may have a negative impact on security. In the design of Mylar [1, 4.1], transitive access permissions combined with chained encryption can cause problems when implemented improperly, or when revocations are required (as clients can cache the private keys they gained access to). Grubbs et al. [2, 5.1] observed that it was difficult for application developers to reach correct security decisions when integrating an implementation. Durak et al. [3, Sec. 4] also found that existing ORE leakage profiles may have underestimated information leakage in implementations, as ORE specifications were not designed to obfuscate unexpected orderings or contextual information often found in data.

2.3 Additional knowledge of encrypted data can increase leakage

Mylar seeks to prevent the server from learning the content of the user's search requests with a token-based encryption system [1, 5.1]. Unfortunately, Grubbs et al. [2] discovered that a malicious server can gain additional knowledge of the content by forcefully assigning users to shared documents, enabled by the flawed implementation of sharing acceptance in Mylar. The additional knowledge allowed the server to eventually deduce a majority of the text content being searched. Several of ORE attacks devised by Durak et al. [3, Sec. 4] also utilised additional knowledge in geographical and time-series representations to accurately approximate encrypted ORE data.

3 Ideas in current context

None of the designers of Mylar participated in the follow-up research by Grubbs et al. [2], which revealed security vulnerabilities in Mylar's design under passive and active attacks. There however have been correspondence between the two groups of researchers in reference to the vulnerabilities raised [4] [5]. Disagreements focused on three aspects: whether the underlying security model of Mylar was already proved to be insufficient by Curtmola et al. [6] in 2006; whether Mylar's implementation of active attack protection was flawed; and whether Mylar's statements of security assurances had been altered. Merits of these arguments are subjects of an ongoing debate.

Mylar utilised searchable symmetric encryption (SSE), a type of property-preserving encryption. An partial solution to the active attack vulnerability raised by Grubbs et al. [2] has been proposed by Etemad et al. [7] very recently, which implemented forward secrecy similar to that used in TLS for SEE, to ensure that newly added files cannot be linked to searches done on previous files, without impacting performance. A security proof has been supplied, subject to verification by other researchers.

An alternative approach to the issue of information leakage as raised by Grubbs et al. [2] and Durak et al. [3] has been developed by Pouliot et al. [8], known as *weakly randomized encryption.* In this method, information leakage can be eliminated if a low-entropy distribution can be inferred from all possible values of a column. This is however only effective against passive attacks, which assume that the attacker cannot manipulate users to gain further information.

4 Literature review

Hahn and Kerschbaum [9] developed a method to allow leakage-free updates to files encrypted under searchable encryption, resolving a usability issue in Mylar. While capable of integrity checking itself, Mylar was also suggested as a confidentiality protection complement to integrity protection tools [10]. Another complementary use of Mylar was suggested by Egorov and Wilkison [11] to facilitate data sharing between end-to-end encryption clients.

More recently, Grubbs et al. [12] also published an enhancement to order-revealing encryption attacks by Durat et al. [3, Sec. 4], in which they exploited auxiliary information of the data to achieve near-perfect level of information recovery from ORE. Lacharité et al. [13] also devised similar attacks on ORE to achieve a high level of information recovery. A generalised method of attack on searchable encryption to exploit flawed security abstractions in encrypted database management systems was also developed [14] by Grubbs et al.

Attacks on Mylar developed by Grubbs et al. [2] were also considered by Herley and Oorschot [15] when exploring problems associated with inappropriate assumptions and threat models in security research. Grubbs et al. also made a distinction between snapshot and persistent passive attacks on encrypted data systems, both of which were checked by Fuller et al. [16] against various protected search systems as part of a wider evaluation.

(1178 words according to texcount.)

References

[1] R. A. Popa, E. Stark, J. Helfer, S. Valdez, N. Zeldovich, M. F. Kaashoek, and H. Balakrishnan, "Building web applications on top of encrypted data using mylar." in *NSDI*, 2014, pp. 157–172.

[2] P. Grubbs, R. McPherson, M. Naveed, T. Ristenpart, and V. Shmatikov, "Breaking web applications built on top of encrypted data," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 2016, pp. 1353–1364.

[3] F. B. Durak, T. M. DuBuisson, and D. Cash, "What else is revealed by order-revealing encryption?" in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 2016, pp. 1155–1166.

[4] R. A. Popa *et al.*, "Response to "Breaking web applications built on top of encrypted data" (CCS 2016) by P. Grubbs, R. McPherson, M. Naveed, T. Ristenpart and V. Shmatikov," November 2016. [Online]. Available: http://css.csail.mit.edu/mylar/security.html

[5] M. Naveed, "Mylar: The guide for the perplexed," September 2016. [Online]. Available: https://docs.google.com/document/u/1/d/1NJHodio0UHs4fLhbLrbpoQJYriOEuUPNmB1e8cPRXfY/pub

[6] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer and communications security.* ACM, 2006, pp. 79–88.

[7] M. Etemad, A. Küpçü, C. Papamanthou, and D. Evans, "Efficient dynamic searchable encryption with forward privacy," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 1, pp. 5–20, 2018.

[8] D. Pouliot, S. Griffy, and C. V. Wright, "The strength of weak randomization: Efficiently searchable encryption with minimal leakage," 2017.

[9] F. Hahn and F. Kerschbaum, "Searchable encryption with secure and efficient updates," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 2014, pp. 310–320.

[10] N. Karapanos, A. Filios, R. A. Popa, and S. Capkun, "Verena: End-to-end integrity protection for web applications," in *Security and Privacy (SP), 2016 IEEE Symposium on.* IEEE, 2016, pp. 895–913.

[11] M. Egorov and M. Wilkison, "Zerodb white paper," *arXiv preprint arXiv:1602.07168*, 2016.

[12] P. Grubbs, K. Sekniqi, V. Bindschaedler, M. Naveed, and T. Ristenpart, "Leakage-abuse attacks against order-revealing encryption," in *Security and Privacy (SP), 2017 IEEE Symposium on.* IEEE, 2017, pp. 655–672.

[13] M.-S. Lacharité, B. Minaud, and K. G. Paterson, "Improved reconstruction attacks on encrypted data using range query leakage," IACR Cryptology ePrint Archive, Tech. Rep. 701, Tech. Rep., 2017.

[14] P. Grubbs, T. Ristenpart, and V. Shmatikov, "Why your encrypted database is not secure," in *Proceedings of the 16th Workshop on Hot Topics in Operating Systems.* ACM, 2017, pp. 162–168.

[15] C. Herley and P. C. van Oorschot, "Sok: Science, security and the elusive goal of security as a scientific pursuit," in *Security and Privacy (SP), 2017 IEEE Symposium on.* IEEE, 2017, pp. 99–120.

[16] B. Fuller, M. Varia, A. Yerukhimovich, E. Shen, A. Hamlin, V. Gadepally, R. Shay, J. D. Mitchell, and R. K. Cunningham, "Sok: Cryptographically protected database search," *arXiv preprint arXiv:1703.02014*, 2017.