Summary of A LifeCycle Model for Privacy Preserving Record Linkage

Brett Walenz

September 27, 2016

Bradley Malin presented an overview of work done towards linking records across hospitals and clinics in the Chicago, IL area. The basic task is to increase the accuracy of summary statistics regarding medical conditions, health events, and drug reactions throughout the area. Unfortunately, the United States has consistently rejected a universal id for healthcare tracking, leading to a fragmentation of medical records, which can lead to overcounting. In addition, there is a constraint placed on the project that requires privacy is preserved between institutions. Each institution is not allowed to discover information about specific patients, or aggregate information about groups of people.

Their solution was a fairly standard cryptographic approach[1]. Suppose we have two institutions, A and B. They introduce two third-party components, a key server and an honest broker, that provide privacy. The process works as follows. The key server generates keys that A and B then use to hash records using a SHA (unspecified which version) hash function. SHA hashes are one-way encryption functions: a key and a string generates a k-bit hash where it should not be possible to find two inputs giving the same output unless the inputs are identical.

The honest broker receives blocks of hashed values, where the number of blocks is slightly fuzzed (it was a little unclear on what guarantees they were providing). This fuzzing prevents an adversary from deducing accurate estimates of each block, which might lead to reverse engineering the blocking technique or other information. Lastly, the honest broker runs whatever matching algorithm is implemented[3], which appeared to be all exact matches (and based on the way they described the cryptography process, is likely the only approach feasible).

At the end, he described the implementation of the system itself[2] and how they extended OpenEMPI (Open Enterprise Master Patient Index) for privacy preservation.

References

[1] A. N. Kho, J. P. Cashy, K. L. Jackson, A. R. Pah, S. Goel, J. Boehnke, J. E. Humphries, S. D. Kominers, B. N. Hota, S. A. Sims, et al. Design and implementation of a privacy preserving electronic health record linkage tool in chicago. *Journal of the American Medical Informatics Association*, page ocv038, 2015.

[2] M. Kuzu, M. Kantarcioglu, E. A. Durham, C. Toth, and B. Malin. A practical approach to achieve private medical record linkage in light of public resources. *Journal of the American Medical Informatics Association*, 20(2):285–292, 2013.

[3] M. Kuzu, M. Kantarcioglu, A. Inan, E. Bertino, E. Durham, and B. Malin. Efficient privacy-aware record integration. In *Proceedings of the 16th International Conference on Extending Database Technology*, pages 167–178. ACM, 2013.