The channel coding theorem and the security of binary randomization

Poorvi L. Vora
Hewlett-Packard Co., 1000 NE Circle Blvd., Corvallis, OR 97330, USA
e-mail: poorvi@ieee.org

Randomization for the purposes of privacy protection refers to the probabilistic perturbation of individual data points to introduce protection through information-theoretic uncertainty. For example, in a telephonic public health survey, a data collector asks the respondent to roll a fair die, and, based on whether the rolled die shows a number divisible by 3 or not, to provide a false or true answer to the question, "Do you have HIV?". The data collector obtains individual answers which are true with probability $\frac{2}{3}$, and can use the answers to calculate the statistics of the prevalence of HIV in the population queried. The individual queried obtains some privacy because whether she has HIV or not is not completely determined by her response. More formally, Alice uses the randomization protocol to provide value $Y \in \mathcal{Y}$ when asked for the value $X \in \mathcal{X}$, given *a posteriori* probability distribution function (pdf) $P(Y|X)$. Randomization has been in use for about twenty years in public surveys and in statistical database security [1], and has recently been proposed as a means of personal privacy protection [3, 2].

In general, the secrecy of randomization is neither information-theoretically [6] nor computationally [7] perfect, as information leakage is the purpose of the protocol. This leaked information can be used by a dishonest data collector to attack the protocol. None of the existing work on randomization addresses in a satisfactory manner its security with respect to attacks that use the leaked information. We address this problem by proposing that one think of the protocol as a channel for the information to be protected. In the example of the telephonic public health survey, if the protocol were replaced by a memoryless binary symmetric channel with probability of error $= \frac{1}{3}$, the outputs of the channel and the protocol would be indistinguishable. The channel view of the protocol is non-typical, because traditional protocols are intended to have zero (information-theoretic or computational) capacity. For this problem, however, the channel view gives access to a number of major theorems in communication and coding theory that have implications for attacks and the security of randomization. Unlike in communication theory, though, the user of this channel, Alice, uses it to *limit* communication given certain constraints.

We consider the randomization of binary-valued $X \in \Sigma = \{0, 1\}$ to obtain $Y \in \Sigma$ given probability of truth p. We denote the randomized response to a request for bit X by $\phi(X)$, the channel $(\mathcal{X}, P(Y|X), \mathcal{Y})$ by Φ, the capacity of Φ by $C(\Phi)$, and a sequence of requested bits by $\mathbf{q}$. In this paper we present results on a specific class of attacks on the protocol, which we call Deterministically-Related Query Sequence (DRQS) Attacks. Roughly speaking, this is the class of attacks where the bits requested by the data collector are completely determined by a (smaller) set of target bits that interest him. The deterministic relationship among the requests may not be obvious to the respondent, as, for example, "losing Calcium" is not always detectable as being related to "female AND over 40".

Definition 1: *A (k, n) DRQS attack on binary protocol ϕ is a one-to-one DRQS map $\Lambda : \Sigma^k \rightarrow \Sigma^n$ and an estimation map $\Psi : \Sigma^n \rightarrow \Sigma^k$ for estimating $\mathbf{q}$ from $\phi(\Lambda\mathbf{q})$. Its rate is $\frac{k}{n}$.*

It is easy to see that, in the communication channel framework, DRQS attacks correspond to codes - the set of requested bits forms a codeword, and the target bits the message.

Theorem 1: *A one-to-one correspondence exists between the set of all one-to-one (k, n) binary channel codes and the set of all (k, n) DRQS attacks.*

Clearly, the estimation error decreases as the data collector asks more questions. We demonstrate tight bounds on the efficiency of *reliable* DRQS attacks which are DRQS attacks that take error to zero while maintaining rate. Our definition draws from the definition of an "achievable" rate in information theory [4, pg. 194].

Definition 2: *A reliable DRQS attack of rate r is said to exist on a binary protocol ϕ if $\exists$ a sequence of (rn, n) DRQS attacks in which the maximum probability of estimation error $\rightarrow 0$ as $n \rightarrow \infty$.*

From the correspondence with channel codes, the channel coding theorem and its converse [5, 4] apply to reliable DRQS attacks.

Theorem 2: *Given a protocol ϕ, reliable DRQS attacks of rate r exist $\forall r < C(\phi)$.*

Theorem 3: *Given a protocol ϕ, reliable DRQS attacks of rate r do not exist $\forall r > C(\phi)$.*

Corollary: *For a symmetric binary protocol with small bias β (flipping a bit with probability $0.5 \pm \beta$), the tight upper bound on the rate of a reliable DRQS attack is $O(\frac{1}{\beta^2})$.*

Acknowledgments

The author would like to thank Umesh Vazirani for his encouragement and for an observation leading to Theorem 1.

References

[1] Nabil R. Adam and John C. Worthmann, "Security-control methods for statistical databases: a comparative study", *ACM Computing Surveys*, Vol. 21, No. 4, pp. 515-556, December 1989.

[2] D. Agrawal and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms", *Proceedings of the Twenteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, Santa Barbara, California, USA, May 21-23 2001.

[3] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining", *Proc. of the ACM SIGMOD Conference on Management of Data*, Dallas, May 2000.

[4] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[5] C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948.

[6] C. E. Shannon, "Communication theory of secrecy systems", *Bell Syst. Tech. J.*, vol. 28, pp. 657–715, 1949.

[7] A. C. Yao, "Theory and Application of Trapdoor Functions", *23rd IEEE Symposium on Foundations of Computer Science*, pp. 80-91, Chicago, Illinois, 3-5 November 1982.

0-7803-7728-1/03/$17.00 ©2003 IEEE.