Journal of Statistical Planning and Inference 148 (2014) 20–37

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Review

On some connections between statistics and cryptology

Palash Sarkar*

Applied Statistics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 26 February 2013
Accepted 14 May 2013
Available online 23 May 2013

Keywords:
Cryptology
Statistics
Cryptanalysis
Randomised response
Differential privacy
Public key encryption
Linear cryptanalysis
Differential cryptanalysis
Side channel attacks

ABSTRACT

The goal of this work is to describe some connections between cryptology and statistics. Starting from basic frequency analysis, throughout history, statistical ideas have been employed to attack cryptographic systems and continue to be important in modern day cryptanalysis. Brief descriptions of hypothesis testing based distinguishing attacks, differential cryptanalysis and statistical ideas used in side channel attacks are provided. From the designer's point of view, we consider three connections. A brief description is provided of the cryptographic ideas that have been suggested to strengthen the technique of randomised response. In recent times, a new notion of privacy of statistical databases has arisen and has been called differential privacy. We provide a brief description of this idea and mention some of its applications. Lastly, we consider the problem of defining and proving security of public key encryption schemes and provide a simple example of this method. It is hoped that the topics outlined here will motivate researchers to further investigate the connection between these two subjects.

Contents

* Tel.: +91 3325752830.
E-mail addresses: palash.sarkar@gmail.com, palash@isical.ac.in

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.05.008

1. Introduction

Cryptography is an old science. Over the history of mankind, the requirements and methods of communication have continuously grown in sophistication. Along with this growth, has arisen the need for maintaining the confidentiality and integrity of communication. Geographically separated parties need to exchange information that is safe from eavesdropping or modification. In the current world, it is possible to think of any number of scenarios where this is a basic requirement. Cryptographic methods and techniques have arisen in response to such challenges. Since the advent of computers and digital communication devices, research in the design and analysis of cryptographic techniques has become extremely important. Historical accounts of cryptology can be found in Kahn (1996) and Singh (2000) and Stinson (2005) provides a good textbook introduction to the subject.

Modern Statistics plays an important role in almost all branches of human knowledge. Since ancient times, methods for cryptanalysis of encryption systems have employed ideas which in the modern light would be called statistical. Given a sequence of symbols which has been produced from a message using a secret key, it is natural to try and apply statistical techniques to extract information about the key or the message. This suggests that a large part of crypanalysis is inherently statistical in nature.

In contrast, if one considers the cryptographic designer's point of view, then the attempt would be to design a system which resists attacks. Large classes of attacks are modelled as probabilistic algorithms and schemes are constructed and proved to be resistant to such attacks. This helps in gaining confidence in a system to be used. The cryptosystem designer's view point involves more of (discrete) probability theory as opposed to that of a cryptanalyst's view point.

The purpose of the present review is to briefly describe some of the connections between statistics and cryptology that have been developed and continue to be explored. In Section 2, we begin with a description of techniques used in pre-modern cryptography and their cryptanalysis. The described systems are no longer in use, but, the cryptanalytic methods involve statistical ideas which justifies their inclusion in this review.

A brief contextual description of cryptography is provided in Section 3. This is necessary for a reader to appreciate the cryptanalytic techniques mentioned later. We also provide a brief description of public key cryptography (PKC) at the very end (in Section 7). While statistical methods have not been used much in the study of PKC, we feel an overview of modern cryptography will be incomplete without providing some idea of PKC. Also, as mentioned above, defining security of public key encryption schemes and proving concrete schemes to be secure often require non-trivial applications of discrete probability. Since our goal is to explore the connection to statistics, we just make a brief mention of this rich research area.

Statistical ideas play an important role in understanding and analysis attacks. Section 4 is devoted to exploring this connection. A large class of cryptanalytic attacks can be modelled as a test of hypothesis problem. This is described and a concrete example is given. Two other attacks are also described. Differential cryptanalysis is a powerful attack technique that has been developed for analysing ciphers. Another interesting class of attacks is to obtain side-channel information as the encryption algorithm executes. Such information can be measurements of the power consumption or even electro-magnetic radiation. The measurements pertaining to such information constitute the data. Statistical techniques are then applied to glean information about the secret key from such data. We provide a simple description of one such attack.

Randomised response technique (RRT) is a well studied statistical technique to protect the privacy of a responder in a sample survey. A twist on this technique has been proposed. Suppose that a responder maliciously deviates from the RRT protocol to try and bias the outcome of the survey in a particular manner. Cryptographic ideas can be used to model a protocol which protects user privacy *and* insulates against maliciously behaving responders. There has been some work along this line and in Section 5 we provide a brief summary of the relevant ideas. The topic is promising and it is conceivable that more work can be done on this topic.

In the modern age, data collected as part of a sample survey are stored in large databases and remain archived for long durations. Such data often contain sensitive information about an individual. An important issue in providing access to the data is that it should not infringe on an individual's privacy. This is a long studied problem in official statistics. In recent times, a new definition of privacy, called differential privacy (DP), has emerged. It has been argued that DP is the 'right' notion of privacy. Section 6 presents a short overview of DP and mentions some of the works that have been done on this topic. DP is a promising area and is being actively researched.

2. Pre-modern cryptography

In this section, we review some early encryption algorithms and their cryptanalysis. The schemes, though simple, were widely used in their times until the methods of breaking them became clear. Cryptanalysis of these ciphers used ideas which would now be called statistical. We should remark that several important ciphers, such as Enigma and Lorentz, were used and broken during the second world war. These are not discussed here since their descriptions and cryptanalysis would require much more space.