Available online at www.sciencedirect.com
ScienceDirect

Theoretical Computer Science 382 (2007) 157–166

Theoretical Computer Science

www.elsevier.com/locate/tcs

On the (im)possibility of non-interactive correlation distillation

Ke Yang[*,1]

1600 Amphitheatre Parkway, Google Inc., Mountain View, CA 94043, United States

Abstract

We study the problem of non-interactive correlation distillation (NICD). Suppose that Alice and Bob each have a string, denoted by $A = a_0a_1 \cdots a_{n-1}$ and $B = b_0b_1 \cdots b_{n-1}$, respectively. Furthermore, for every $k = 0, 1, \ldots, n-1$, (a_k, b_k) is drawn independently from a distribution $\mathcal{N}$, known as the 'noise model'. Alice and Bob wish to 'distill' the correlation non-interactively, i.e., they wish to each apply a function to their strings, and output one random bit, denoted by X and Y, such that $\Pr[X = Y]$ can be made as close to 1 as possible. The problem is, for what noise models can they succeed? This problem is related to various topics in computer science, including information reconciliation and random beacons. In fact, if NICD is indeed possible for some general class of noise models, then some of these topics would, in some sense, become straightforward corollaries.

We prove two negative results on NICD for various noise models. We prove that, for these models, it is impossible to distill the correlation to be arbitrarily close to 1. We also give an example where Alice and Bob can increase their correlation with one bit of communication (in this case they need to each output two bits). This example, which may be of interest on its own, demonstrates that even the smallest amount of communication is provably more powerful than no communication.

Keywords: Correlation distillation; Communication complexity; Random beacon; Information reconciliation; Fourier analysis

1. Introduction

1.1. Non-interactive correlation distillation

Consider the following scenario. Let $\mathcal{N}$ be a distribution over $\Sigma \times \Sigma$, where Σ is an alphabet. We call $\mathcal{N}$ a 'noise model'. Suppose that Alice and Bob each receive a string $A = a_0a_1 \cdots a_{n-1}$ and $B = b_0b_1 \cdots b_{n-1}$, respectively, as their local inputs. For every $k = 0, 1 \ldots, n-1$, (a_k, b_k) is drawn independently from $\mathcal{N}$. Now Alice and Bob wish to engage in a protocol to 'distill' their correlation. At the end of the protocol, they wish to each output a bit, denoted by X and Y, respectively, such that both X and Y are 'random enough', while $\Pr[X = Y]$ can be made as close to 1 as possible, possibly by increasing n. We call such a protocol a *correlation distillation protocol*. Furthermore, if Alice and Bob wish to do so *non-interactively*, i.e., without communication, we call this 'non-interactive correlation distillation' (NICD). Notice that, in NICD, the most general thing for Alice and Bob to do is to each apply a function

[*] Tel.: +1 650 623 6546; fax: +1 650 623 6710.
E-mail address: yangke@google.com.

[1] This work was done while the author was a student at Carnegie Mellon University, Pittsburgh, PA 15213, USA.

0304-3975/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2007.03.007

to their local inputs and outputs one bit. The problem of NICD is, for what noise model can Alice and Bob achieve this goal?

We note that NICD is indeed possible for some noise models. For example, if a noise model $\mathcal{N}$ is in fact 'noiseless', i.e., $\Pr_{(a,b)\in\mathcal{N}}[a = b] = 1$, then NICD is possible. See Section 1.3 for more discussions. However, we are interested in the 'noisy' noise models, for example the *binary symmetric model*, where Alice and Bob each have an unbiased bit as input, which agree with probability $1 - p$, and the *binary erasure model*, where Alice's input is an unbiased bit x and Bob's input is x with probability $1 - p$, and a special symbol $\perp$ with probability p. These models are studied extensively in the context of error-correcting codes [3,9], where Alice encodes her information before sending it through a 'noisy channel'. It is known that there exist efficient encoding schemes that withstand these noise models and allow Alice and Bob to achieve almost perfect correlation. However, in the case of NICD, the 'raw data' are already noisy. Can the techniques in error-correcting codes be used here, and is NICD possible for these noise models?

1.2. Motivations and related work

Besides the obvious relation to error-correcting codes, the study of NICD is naturally motivated by several other topics. We review these topics and discuss some of the related work.

1.2.1. Information reconciliation

Information reconciliation is an extensively studied topic [4,6–8,15] with applications in quantum cryptography and information-theoretical cryptography. In this setting, Alice and Bob each receives a sequence of random bits drawn from a noise model, while Eve, the eavesdropper, also possesses some information about the bits. Alice and Bob wish to 'reconcile' their information via an 'information reconciliation protocol', where they exchange information in a noiseless, public channel in order to agree on a random string U with very high probability. Therefore, information reconciliation protocols are somewhat like correlation distillation protocols. However, the primary concern for information reconciliation is *privacy*, i.e., that Eve gains almost no information about U. Intuitively, since Eve can see the conversation between Alice and Bob, maximum privacy would be achieved if information reconciliation can be performed without communication.

1.2.2. Random beacons

A random beacon is an entity that broadcasts uncorrelated, unbiased random bits. The concept of random beacons was first introduced in 1983 by Rabin [18], who showed how they can be used to solve various problems in cryptography. From then on, random beacons have found many applications in security and cryptography [2,5,10, 11,14]. There are many proposals for constructing a *publicly verifiable* random beacon; among them are those that use the signals from a cosmic source [16]. In these proposals, Alice (as the beacon owner) and Bob (as the verifier) both point a telescope at an extraterrestrial object, e.g. a pulsar, and then measure the signals from it. Presumably these signals contain a sufficient amount of randomness. Then Alice converts her measurement results into a sequence of random bits, and publishes them as beacon bits. Bob can then verify the bits by performing his own measurement and conversion. However, it is inevitable that there would be discrepancies in the results of Alice and Bob, due to measurement errors (described by a noise model). These discrepancies may cause the beacon bits published by Alice to disagree with those computed by Bob. One of the major concerns in the study on random beacons is to prevent *cheating* in the presence of measurement error. In other words, one needs to design a mechanism to prevent Alice from maliciously modifying her measurement data in order to affect the beacon bits, while pretending that the modification comes from the measurement error. Notice that, in general, there is no communication between Alice and Bob. We note that if NICD is possible, then the cheating problem would be solved, since NICD protocols can be used to distill almost perfectly correlated bits. Then, with very high probability, the bits output by Alice and Bob should agree, and this essentially removes the measurement error.

1.2.2.1. Related work. As we have discussed, the problem of NICD lies, in some sense, at the foundations of both the studies of information reconciliation and random beacons. In fact, researchers from both areas have, to some extent, considered the problem of NICD. In particular, a basic version of the problem concerning a very special type of NICD protocol over the symmetric noise model was discovered and proven independently by several researchers since as early as 1991, including Alon, Maurer,and Wigderson [1] and Mossel and O'Donnell [16]. They proved that NICD is