Multi-Prover Interactive Proofs: Unsound Foundations

Claude Crépeau[1] * and Nan Yang[2] **

[1] McGill University, Montréal, Québec. crepeau@cs.mcgill.ca
[2] Concordia University, Montreal, Canada. na_yan@encs.concordia.ca

Abstract. Several Multi-Prover Interactive Proofs (MIPs) found in the literature contain proofs of soundness that are lacking. This was first observed [1] in which a notion of *Prover isolation* is defined to partly address the issue. Furthermore, some existing Zero-Knowledge MIPs suffer from a catastrophic flaw: *they outright allow the Provers to communicate via the Verifier*. Consequently, their soundness claims are now seriously in doubt, if not plain wrong. This paper outlines the lack of isolation and numerous other issues found in the (ZK)MIP literature. A follow-up paper will resolve most of these issues in detail.

1 Introduction

It has been a long-held intuition that if Alice and Bob share an inconsistent set of beliefs then, if they are questioned individually, one can expose that inconsistency. This is the idea behind the theory of Multi-Prover Interactive Proofs (MIPs): a polynomial-time Verifier who is trying to discern truth from falsity from a set of all-powerful Provers who cannot signal with each other. This theory originated from the work of Ben-Or, Goldwasser, Kilian and Wigderson [2], and we denote the class of languages with such interactive proofs by **MIP** (and its zero-knowledge counterpart **ZKMIP**). In that paper and subsequent work of Babai, Fortnow and Lund [3], it was claimed that **ZKMIP** = **MIP** = **NEXP**.

The proof of security in [2] and many subsequent MIPs reduces the breaking of soundness to signalling. However, in the last decade, two major problems with MIPs/ZKMIPs have emerged. The first is that the Provers do not actually need to signal in order to break some MIPs, as demonstrated in the work of Cleve, Høyer, Toner and Watrous [4]; they can perform *no-signalling tasks* which do not allow communication (for example, using shared entanglement). That is, there is a fundamental and yet subtle difference between what is *local* and what is *no-signalling*. The second by Crépeau, Salvail, Simard and Tapp [1] is that while the Provers are unable to signal between themselves, the Verifier could inadvertently perform a non-local task for them; in the extreme case, the Verifier may plainly signal *for* the Provers.

* Supported in part by Québec's FRQNT and Canada's NSERC.
** Supported in part by Prof. David Ford and by Prof. Jeremy Clark.

By combining the two problems, a Verifier can perform a no-signalling task for the Provers, and thus allow them to break soundness of their protocols. In this case, not only are the Provers perfectly no-signalling, they do not even need any extra no-signalling resources (such as quantum entanglement).

The role that the Verifier must play in these MIPs was studied in [1]. It was defined and shown that a Verifier must be *isolating*, so that it will never (inadvertently or not) perform a *non-local* task (no-signalling or signalling). We show here that many existing MIPs do not satisfy isolation, even in a weak sense.

More recently, the model of Multi-Prover Interactive Proofs was extended to allow entangled Provers and the class of languages accepted under this new setting is called $\mathsf{MIP}^\star$ [5]. It was recently shown that $\mathsf{NEXP} \subseteq \mathsf{MIP}^\star$ [6] but we do not know whether equality holds. Similarly, the model of Multi-Prover Interactive Proofs was extended to allow No-signalling Provers and the class of languages accepted under this new setting is called $\mathsf{MIP}^{\mathsf{ns}}$ [5]. We now know that $\mathsf{MIP}^{\mathsf{ns}} = \mathsf{EXP}$ [7]. We use some of these results to illustrate our explanation.

2 Terminology: (non-)local, (no-)signalling and entangled

The terms "communicating" and "signalling" are used equivalently throughout this work and should have the obvious meaning of information transfer between two or several parties. Signalling Provers are essentially the same as a single Prover because we put no restriction whatsoever on their communication (potential interesting sub-cases arise when we restrict the amount of communication they can actually use, but we do not consider them here). In the context of several parties, we consider that signalling is taking place even if no individual communicates with any other individual. In cryptographic terms, if someone uses secret-sharing to distribute a message to several parties excluding himself (even if all of them are required to communicate to reconstruct the original secret) then signalling is considered to have taken place between the sender and the secret-share-holders.

However, as soon as we restrict communication we would need to define what non-communication (or no-signalling) actually means. The initial intuition was that non-communication = locality, meaning that the Provers are allowed to share arbitrary amount of randomness before being restricted to computations involving only these local random variables. However, because of entanglement, it was later understood that certain classes of probability distributions cannot be shared in a local fashion, but they do not allow communication. The term *no-signalling* was coined to define "everything but signalling". Of course this includes locality, but also strictly more. Typical examples are the CHSH Game (on inputs a, b output x, y s. t. $x \oplus y = a \times b$) and the Magic Square Game [4].

This terminology mostly originates from physics. The acclaimed work from John Bell [8] in the 1960s can be summarized thus, "It seems that quantum entanglement allows for non-local yet no-signalling distributions". However, it turns out that quantum physics does not allow *all* no-signalling distributions. For instance, the CHSH game cannot be achieved from quantum entanglement.

An approximation that succeeds roughly 85.4% of the time can be achieved using entanglement, whereas any local strategy can only succeed up to 75% of the time [4]. Winning the CHSH game 100% of the time is impossible even using quantum entanglement.

It may seem that the only models which make sense are "local" and "entangled" because they are motivated by physical models of reality. Nevertheless, the no-signalling model turns out to be useful under certain circumstances as explained in [7] Section 1.2 : "We show that any MIP that is sound against no-signalling cheating Provers can be converted into a 1-round delegation scheme, using a fully homomorphic encryption scheme (FHE), or alternatively, using a computational private information retrieval (PIR) scheme."

3 Issues With Existing Protocols

First, we illustrate that MIPs may be sound on their own but not when composed. It was shown in [4] that the Magic Square Game may be turned into a language which has a MIP that is sound classically but unsound when Provers share entanglement.

We present a variant of this MIP below (as Figure 1). Given a string of six bits $r_0, r_1, r_2, c_0, c_1, c_2$, the success probability of the classical Provers is one when there exists such a matrix M and at most 8/9 when no such matrix M exists. By repeating this protocol many times, V will be able to decide with high probability whether such a matrix exists or not. However, if P_0, P_1 can win the Magic Square Game, they can systematically break the soundness of this protocol and succeed with probability one whether such a matrix exists or not.

Construction 31 *On input six bits $r_0, r_1, r_2, c_0, c_1, c_2$,*
P_0 and P_1 claim that there exists a 3×3 binary matrix

$$M := \begin{bmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{bmatrix}$$

such that for $0 \leq t \leq 2$, $m_{t0} \oplus m_{t1} \oplus m_{t2} = r_t$ and $m_{0t} \oplus m_{1t} \oplus m_{2t} = c_t$.

- *V chooses two trits a, b uniformly and sends a to P_0 and b to P_1.*
- *P_0 (P_1) replies with row $[m_{a0}, m_{a1}, m_{a2}]$ (column $[m_{0b}, m_{1b}, m_{2b}]$).*
- *V checks that $m_{a0} \oplus m_{a1} \oplus m_{a2} = r_a$, that $m_{0b} \oplus m_{1b} \oplus m_{2b} = c_b$, and that m_{ab} is the same unique value from both P_0 and P_1.*

Fig. 1. A MIP for a language on six bits strings.

3.1 Issues with current proofs of composability

A problem with prior MIPs' proofs of soundness is that different protocols (each of which do not allow communication) can break each other.

For instance, the MIP from [7] is resistant to no-signalling strategies. Therefore if we change [7] by appending at its end an implementation of the CHSH box by the Verifier, we would still have provable soundness. However, this new MIP, when concurrently composed with any MIP vulnerable to no-signalling strategies will result in a protocol that is unsound.

The same problem exists for protocols which are vulnerable to entanglement. The MIP from [6] is resistant to entangled Provers. We can modify this protocol into one which asks the Verifier to implement some Magic Square Games [4] without affecting soundness. This new protocol, concurrently composed with the protocol of Figure 1, breaks the soundness of the latter protocol.

In either of the above two cases, the no-communication assumption of the composed protocols is not broken. While the above examples illustrate problems with *concurrent* composition, we consider that this is indicative of the incompleteness of existing MIPs and their analyses.

3.2 Issues Specific to ZKMIPs

In this section we explain issues with the specific construction found in [2, 9] which transforms an arbitrary MIP for language L into a Zero-Knowledge version of the same proof. The technique involves the Provers using commitments to show the Verifier that "if you were to see the contents of these (committed) discussions, you would accept that $x \in L$.". In the case where local (or entangled) Provers are involved, it is possible to construct bit (or trit) commitment schemes that are perfectly concealing and statistically binding [1]. One of these (Construction 33) rests on the Magic Square game and is binding against classical Provers but not against entangled Provers, while a second (Construction 32) rests on the CHSH Game and is binding against classical and entangled Provers but not against No-signalling Provers.

We summarize the construction of Kilian (and BGKW) as Fig. 2. The purpose of this protocol is to convert a generic MIP $\langle P_0, P_1, V \rangle$ into a specific format given in this Figure and then compile it using Bit Commitments to make it Zero-Knowledge. The issue at hand with this construction is the Steps 1. and 3. of this protocol where P_1' send messages a_0 and a_1 to V'.

In those Steps the Prover P_1' may send V' arbitrary messages as long as V' does not reject them and abort in Step 4. Imagine if we had modified the Verifier V into V^* in such a way that it ignores whatever the Provers say unless it starts with "Simon says" and the Provers P_0, P_1 accordingly. Clearly, the resulting MIP $\langle P_0^*, P_1^*, V^* \rangle$ will be as good and sound as the original $\langle P_0, P_1, V \rangle$. However, when transformed by the protocol of Fig. 2 the new MIP will allow dishonest Provers to send arbitrary messages that V^* will ignore. In Step 5., V' will imbed these arbitrary messages into Q deterministically (see Figure 5) and feed it to P_2'. This is a communication channel P_1' may use to send arbitrary messages to P_2'. The issue here is that nowhere is it verified that *any* of these Verifiers are Isolating. The Verifier of this protocol allows P_1' to send messages to P_2' and adding Commitments will not fix that.

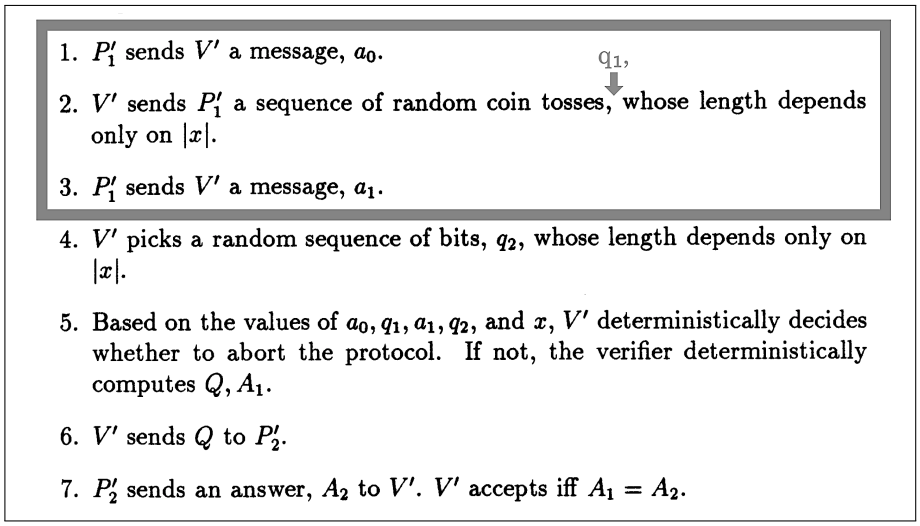

Fig. 2. Figure from [9] chapter 6, page 207.

This issue may be used in several different ways to break soundness of the protocol. We explain only one here, but will illustrate it with several examples in the complete version of this paper.

If this protocol is composed with any other one that uses one of the Commitments of Fig. 3-4 where P'_1 receives the string z, he can communicate it to its partner Prover. Once this has happened, the Commitments are no longer binding and whatever proof they are using loses its soundness completely.

3.3 Synchronous VS Asynchronous MIPs

In all the MIP literature, it is never clearly specified whether MIPs are *synchronous* or *asynchronous*. Can the Verifier interact with the Provers at its own (chosen) pace independently of any clock or is the whole thing very accurately clocked? Does it even matter? We argue that being *asynchronous* is much more desirable than being *synchronous*.

In the asynchronous setting, V can interact with the Provers in any order it likes, at any rate it likes. The Provers will be allowed to communicate only when both of their respective protocols are finished with V. If the Provers are not allowed to communicate and if V does not help them in that sense, we expect this asynchronous property to be satisfied. On the contrary, in the synchronous setting V must interact with the Provers in the exact order of the protocol. If the protocol has rounds, V must complete the first round before moving on to the second round and so on. If V and the Provers have a common clock they can actually have each step of the protocol happen at a very precise time and abort if any party is not ready at the expected time or if messages did not arrive by their prescribed deadline. All MIPs for local, entangled or no-signalling provers

Construction 32 *All parties agree on a security parameter k.*
P_0 and P_1 partition their private random tape into a $k+1$-bit string $\langle c, w\rangle$, where c is a bit and $|w| = k$.

Pre-computation phase:

- *V chooses a k-bit string z uniformly at random and sends it to P_1.*
- *P_1 responds with $d = w \oplus c \times z$, where $c \times z$ is thought of as the product between a scalar c and a vector z, over $\mathbb{Z}_2$.*

Commit phase:

- *P_0 commits b to V as $\boxed{b} = b \oplus c$.*

Unveil phase:

- *P_0 sends w to V.*
- *V computes $c = 1$ if $d \oplus w = z$, or $c = 0$ if $d \oplus w = \mathbf{0}$ and recovers $b = \boxed{b} \oplus c$. V rejects if $d \oplus w$ does not equal to either z or $\mathbf{0}$.* □

Fig. 3. Statistically binding, perfectly concealing bit-commitment protocol.

should be writable in the asynchronous model because for such provers it should not matter who goes first and last.

Clearly, a protocol that is provably correct and sound in the asynchronous model will also be correct and sound in the synchronous model, but certainly not the other way around. For instance, the construction of Kilian in Fig. 2 is *not* clearly defined in the *asynchronous* model because Step 6 requires Steps 1–5 to have been completed before it can be performed. Any attempt to defining an asynchronous version has lead to either incorrect or unsound protocols. It is another reason why we became suspicious of the constructions leading ZKMIPs.

We strongly believe that we should require all MIPs to be sound in the asynchronous model. Moreover, when defining Zero-Knowledge, arbitrary Verifiers should be asynchronous. This will result in a stronger notion of Zero-Knowledge as opposed to restricting the participants to be synchronous.

3.4 A concrete example

We give a somewhat contrived example in existing literature which is a striking example of the consequences of lacking isolation. Consider the protocol in Fig. 5. The first Prover P_1' will simulate a number of transcripts of a MIP involving a simulated Verifier and a number of simulated Provers. The actual Verifier V' will then send a random, partial transcript to the second Prover P_2'. This partial transcript contains only one of the simulated Provers' questions and answers up to a random point. P_2' must then be able to complete the transcript in an

Construction 33 *All parties agree on a security parameter k.*
P_0 and P_1 partition their private random tape into a trit and a $4k$-bit string $\langle e, w\rangle$, where e is the trit and $|w| = 4k$. P_0 and P_1 use each block of 4 bits $w_{4j}...w_{4j+3}$ to construct a 3×3 binary matrix (where $c_x = 1$ iff $x \neq e$.)

$$\begin{bmatrix} m^j_{00} & m^j_{01} & m^j_{02} \\ m^j_{10} & m^j_{11} & m^j_{12} \\ m^j_{20} & m^j_{21} & m^j_{22} \end{bmatrix} := \begin{bmatrix} w_{4j} & w_{4j+1} & w_{4j} \oplus w_{4j+1} \\ w_{4j+2} & w_{4j+3} & w_{4j+2} \oplus w_{4j+3} \\ c_0 \oplus w_{4j} \oplus w_{4j+2} & c_1 \oplus w_{4j+1} \oplus w_{4j+3} & c_2 \oplus w_{4j} \oplus w_{4j+1} \oplus w_{4j+2} \oplus w_{4j+3} \end{bmatrix}$$

Pre-computation phase:

- *V chooses a k-trit string z uniformly at random and sends it to P_1.*
- *P_1 responds with row $\left[m^j_{z_j 0}, m^j_{z_j 1}, m^j_{z_j 2}\right]$, for $1 \leq j \leq k$.*
- *V checks that $m^j_{z_j 0} \oplus m^j_{z_j 1} \oplus m^j_{z_j 2} = 0$, for $1 \leq j \leq k$.*

Commit phase:

- *P_0 commits t to V as $\boxed{t} = t + e \bmod 3$.*

Unveil phase:

- *P_0 sends e to V.*
- *V chooses a k-bit string r uniformly at random and sends it to P_0.*
- *Both compute $s_j := e + r_j + 1 \bmod 3$, for $1 \leq j \leq k$.*
- *P_0 responds with column $\left[m^j_{0 s_j}, m^j_{1 s_j}, m^j_{2 s_j}\right]$, for $1 \leq j \leq k$.*
- *V checks that $m^j_{0 s_j} \oplus m^j_{1 s_j} \oplus m^j_{2 s_j} = 1$, and that $m^j_{z_j s_j}$ is the same unique value from both P_0 and P_1, for $1 \leq j \leq k$.* □

Fig. 4. Statistically binding, perfectly concealing trit-commitment protocol.

identical way, otherwise the proof is rejected. The simulated Provers and Verifiers are deterministic, which should allow this consistency check with P'_2 to succeed.

The first problem we would like to point out is that the simulated *answers* from the simulated Provers cannot be authenticated. There is no *a priori* reason why the Verifier would suspect these answers to be attempts at communication. In addition, not only is the Verifier not isolating, in this case the protocol *requires* that the Verifier actually courier some messages from one Prover to another.

Suppose that the simulated protocol has a "header" section where the simulated Provers can say anything and it will be ignored by the Verifier, but would nevertheless be part of a valid transcript. In the compositional form of Kilian's protocol, the Verifier has an auxiliary input tape (which is normally used to model prior knowledge a Verifier might have). The real Provers can use this auxiliary tape to communicate; in particular, P'_1 can send to P'_2 the value of R, the random coins V' is forcing P'_1 to use. This fixes the simulated Verifier's

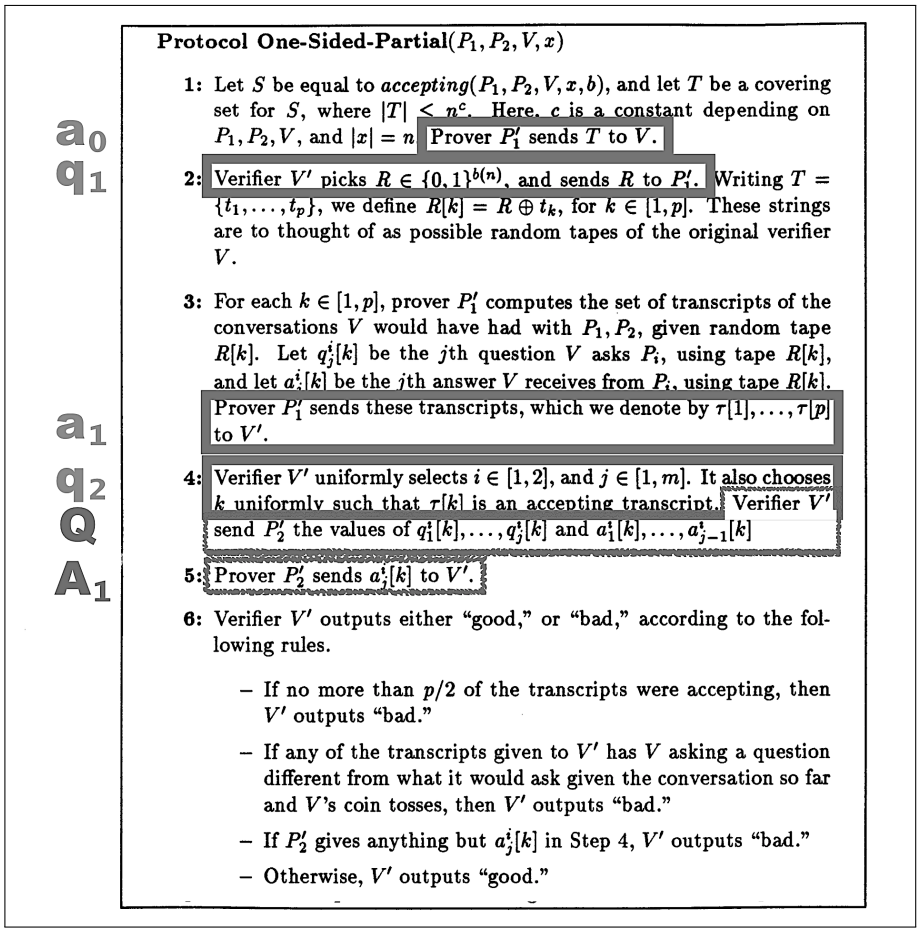

Fig. 5. Figure from [9] chapter 6, page 205.

random tape, which allows the real second Prover P_2' to break any consistency checks, and therefore soundness.

Even if there is no auxiliary tape, the simulated Provers' *first* question cannot be in general authenticated by the Verifier as genuine (as the Provers are all-powerful but the Verifier is only polynomial-time). Therefore P_1' can always produce a transcript where R is disguised as the first question. Soundness breaks by the same reasoning as above. Luckily, in this case, the cheating Provers would be detected if V' asks P_2' to produce the *first* question from scratch, which it cannot since it does not know R.

Now the second problem is that in [9], the proof of soundness does not mention this near-miss. This is symptomatic of proofs of soundness of MIPs in the literature: their proofs are considerably incomplete; fixing these proofs require proper definitions of many concepts taken for granted in these papers.

4 Discussion

Considering the many issues described in the previous sections, we believe that there is a need to rethink MIPs/ZKIPs with respect to locality, synchronicity, composability and isolation. The attacks that we have demonstrated may be somewhat contrived, but they demonstrate the incompleteness of existing work.

In particular, we think that new definitions and proofs are necessary to capture the counter-intuitiveness of non-locality, including entanglement and other no-signalling tasks. Existing results must be revalidated under an upgraded model. We will explore this idea in detail in our follow-up paper.

Acknowledgements

We would like to thank Serge Fehr, Gilles Brassard, Samuel Ranellucci, Christian Schaffner, and Louis Salvail for various fruitful discussions about this work. Finally, we are grateful to Raphael C.-W. Phan and Moti Yung for inviting us to submit our work here. Special thanks to Justin Holmgren (MIT) for pointing out an incorrect/unclear statement at the end of Section 3.3 which is now corrected.

References

1. C. Crépeau, L. Salvail, J.-R. Simard, and A. Tapp, *Two Provers in Isolation*, pp. 407–430. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
2. M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson, "Multi-prover interactive proofs: How to remove intractability assumptions," in *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, STOC '88, (New York, NY, USA), pp. 113–131, ACM, 1988.
3. L. Babai, L. Fortnow, and C. Lund, "Non-deterministic exponential time has two-prover interactive protocols," *Comput. Complex.*, vol. 2, pp. 374–374, Dec. 1992.
4. R. Cleve, P. Hoyer, B. Toner, and J. Watrous, "Consequences and limits of nonlocal strategies," in *Proceedings of the 19th IEEE Annual Conference on Computational Complexity*, CCC '04, (Washington, DC, USA), pp. 236–249, IEEE Computer Society, 2004.
5. T. Ito, "Polynomial-space approximation of no-signaling provers," in *Proceedings of the 37th International Colloquium Conference on Automata, Languages and Programming*, ICALP'10, (Berlin, Heidelberg), pp. 140–151, Springer-Verlag, 2010.
6. T. Ito and T. Vidick, "A multi-prover interactive proof for nexp sound against entangled provers," in *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, FOCS '12, (Washington, DC, USA), pp. 243–252, IEEE Computer Society, 2012.
7. Y. T. Kalai, R. Raz, and R. D. Rothblum, "How to delegate computations: The power of no-signaling proofs," in *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, (New York, NY, USA), pp. 485–494, ACM, 2014.
8. J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, pp. 195–200, 1964.
9. J. Kilian, *Uses of randomness in algorithms and protocols*. MIT Press, 1990.