Coin Tossing with Lazy Defense: Hardness of Computation Results

Hamidreza Amini Khorasgani
Department of Computer Science, Purdue University, USA
haminikh@purdue.edu

Hemanta K. Maji
Department of Computer Science, Purdue University, USA
hmaji@purdue.edu

Mingyuan Wang
Department of Computer Science, Purdue University, USA
wang1929@purdue.edu

Abstract

There is a significant interest in securely computing functionalities with guaranteed output delivery, a.k.a., fair computation. For example, consider a 2-party n-round coin-tossing protocol in the information-theoretic setting. Even if one party aborts during the protocol execution, the other party has to receive her outcome. Towards this objective, every round, the sender of that round's message, preemptively prepares a defense coin, which is her output if the other party aborts prematurely. Cleve and Impagliazzo (1993), Beimel, Haitner, Makriyannis, and Eran Omri (2018), and Khorasgani, Maji, and Mukherjee (2019) show that a fail-stop adversary can alter the distribution of the outcome by $\Omega(1/\sqrt{n})$.

However, preparing the defense coin is computationally expensive. So, the parties would prefer to update their defense coin only sparingly or when indispensable. Furthermore, if parties delegate their coin-tossing task to an external server, it is even infeasible for the parties to stay abreast of the progress of the protocol and keep their defense coins in sync with the protocol evolution. Therefore, this paper considers lazy coin-tossing protocols, where parties update their defense coins only a total of d times during the protocol execution. Is it possible that using only $d \ll n$ defense coin updates a fair coin-tossing protocol is robust to $\mathcal{O}(1/\sqrt{n})$ change in their output distribution?

This paper proves that being robust to $\mathcal{O}(1/\sqrt{n})$ change in the output distribution necessarily requires that the defense complexity $d = \Omega(n)$, thus ruling out the possibility mentioned above. More generally, our work proves that a fail-stop adversary can bias the outcome distribution of a coin-tossing protocol by $\Omega\left(1/\sqrt{d}\right)$, a qualitatively better attack than the previous state-of-the-art when $d = \mathrm{o}(n)$. That is, the defense complexity of a coin-tossing protocol, not its round complexity, determines its security. We emphasize that the rounds where parties calculate their defense coins need not be a priori fixed; they can depend on the protocol's evolution itself. Finally, we translate this fail-stop adversarial attack into black-box separation results for lazy coin-tossing protocols.

The proof relies on an inductive argument using a carefully crafted potential function to precisely account for the quality of the best attack on coin-tossing protocols. Previous approaches fail when the protocol evolution reveals information about the defense coins of both the parties, which is inevitable in lazy coin-tossing protocols. Our analysis decouples the defense complexity of coin-tossing protocols from its round complexity to guarantee fail-stop attacks whose performance depends only on the defense complexity of the coin-tossing protocol; irrespective of their round complexity.

Keywords and phrases Discrete-time Martingale, Coin-tossing Protocols, Fair Computation, Defense Complexity, Fail-stop Adversary, Black-box Separation

Funding The research effort is supported in part by an NSF CRII Award CNS–1566499, an NSF SMALL Award CNS–1618822, the IARPA HECTOR project, MITRE Innovation Program Academic Cybersecurity Research Award, a Purdue Research Foundation (PRF) Award, and The Center for Science of Information, an NSF Science and Technology Center, Cooperative Agreement CCF–0939370.

1 Introduction

Guaranteed output delivery is a desirable attribute of secure computation protocols. Secure computation of functionalities with guaranteed output delivery ensures that even if a party aborts during the execution of the protocol, the other party still obtains her output. Defining security and constructing secure protocols in this setting for general functionalities has been a field of highly influential research [Cle86, CI93, GHKL08, GK10, BLOO11, ALR13, Ash14, Mak14, ABMO15]. This paper studies the fundamental 2-party coin-tossing functionality in this setting. Our motivation underlying the study and our contributions are best exemplified in the context of the following representative problem.

Representative Motivating Problem. Alice and Bob are interested in generating multiple (shared) random coins; each of them is heads (independently) with probability X_0. Instead of undertaking this computationally heavy task themselves, they delegate it to a dedicated third-party cloud computing service (refer to Figure 1 for illustration). This cloud computing service spawns two processes to generate each coin (one on behalf of Alice and one on behalf of Bob) and runs one instance of an n-round 2-party coin-tossing protocol between these two processes. Upon the completion of the protocol, Alice's process reports its outcome back to Alice and, likewise, Bob's process reports its outcome back to Bob.

However, there is a threat that the cloud computing platform gets infected by a virus that can eavesdrop on the communication between the processes participating in the coin-tossing protocol instances. Based on the communication, the adversary can terminate the processes before they report back their respective outcomes to Alice and Bob. In order to defend against the complete loss of their computation, the processes report back intermediate coin outcomes every t rounds to safeguard the partial progress of their computation. Ideally, one would like to set $t = 1$ to keep Alice/Bob abreast of the progress in the coin-tossing protocol.

The computation of these *defense coins*, however, is computationally expensive. Each process has to sample a complete transcript that honestly extends the partial transcript generated so far. This extension is computationally expensive without the knowledge of the other process's internal state [JVV86, BGP00]. Furthermore, reporting back to Alice/Bob introduces network latency that is multiple orders of magnitude larger than the low-latency of the coin-tossing protocol among processes on the same processor (see, for example, https://gist.github.com/hellerbarde/2843375). Inevitably, by the time Alice/Bob receive their defense coin, the coin-tossing protocol would have already progressed significantly ahead. Consequently, only large values of t are achievable.

Given n, t, and X_0, the *insecurity* of any coin-tossing protocol instance is the maximum change in the output distribution that the virus causes by killing the processes. In this scenario, the following questions are but natural. How much insecurity (as a function of n, t, and X_0) should Alice and Bob anticipate? Equivalently, given a particular tolerance for security, how frequently should Alice and Bob update their defense coins?

Looking Ahead. All previous works [CI93, BHMO18, KMM19] are applicable only for the particular case of $t = 1$. They prove that the insecurity in the coin-tossing protocol mentioned above (qualitatively) behaves as $\frac{X_0(1-X_0)}{\sqrt{n}}$. This bound leaves open the possibility

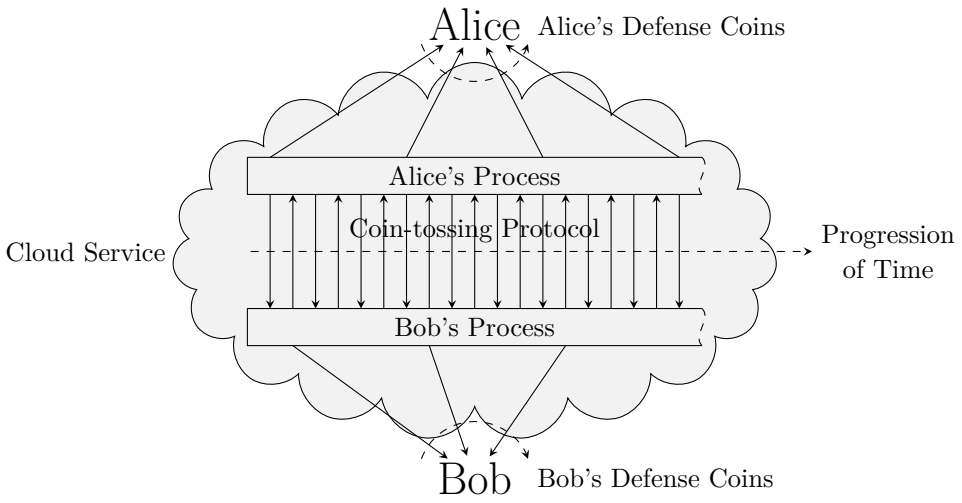

Figure 1 Illustration of Alice and Bob delegating the generation of one coin toss to a cloud service where parties are lazy to obtain their defense coins.

that one might increase the time t between updating defenses (a.k.a., parties *lazily* update their defenses) without sacrificing the security of the protocol. Existing proof techniques break down entirely when the evolution of the coin-tossing protocol between consecutive updates of Alice's and Bob's defense coins reveals information about both of their defense coins, which is indeed possible for $t \geqslant 2$ (see Figure 4 for a concrete example). To circumvent this challenge, we introduce a new inductive proof strategy demonstrating that the insecurity is at least $\frac{X_0(1-X_0)}{\sqrt{n/t}}$, a qualitatively better lower bound when $t = \mathrm{o}(1)$. Note that $d := n/t$, referred to as the *defense complexity* of the protocol, is the total number of defense coins received by Alice and Bob during the protocol execution. In general, we demonstrate that the defense complexity of the protocol, *not* the round complexity, is key to determining the insecurity of a coin-tossing protocol. Intuitively, for example, our result implies that a high round-complexity coin tossing protocol is vulnerable if parties do not frequently update their defense coins. We emphasize that the decision of whether to update the defense coin or not in a round may depend on the evolution of the protocol itself.

1.1 Discussion on Previous Approaches

Consider a 2-party n-round coin-tossing protocol such that the probability of the outcome being 1 is X_0, and the probability of the outcome being 0 is $1-X_0$. Let $\mathcal{X} = (X_0, X_1, \ldots, X_n)$ represent the Doob's martingale corresponding to this protocol where X_i represents the expected outcome conditioned on the first i messages of the transcript. Note that $X_n \in \{0, 1\}$, because at the end of the protocol, both parties agree on the outcome being 0 or 1 with certainty. Previous works [CI93, KKR18, BHMO18, KMM19] prove the existence of a (randomized) round $\tau \in \{1, 2, \ldots, n\}$ such that the expected magnitude of the gap $|X_\tau - X_{\tau-1}|$ is $\Omega\left(\frac{X_0(1-X_0)}{\sqrt{n}}\right)$. We clarify that the round τ being randomized implies that it can depend on the partial transcript generated during the protocol. Such a round τ, intuitively, is susceptible to attacks because there is a significant gap between the knowledge of the two parties regarding the (expected) outcome of the protocol.

In *fair* coin-tossing protocols [Cle86, CI93, GHKL08] (i.e., coin-tossing protocols with

guaranteed output delivery), if one of the parties aborts prematurely, then the other party still has to output 0 or 1. Intuitively, the two parties carry private defense coins, which they regularly update as the protocol progresses. If a party aborts, then the other party outputs her defense coin. Without loss of generality, one can assume that the parties update their defense coin as part of their next message computation in the protocol execution. For example, without loss of generality, assume that Alice plays the role of the party that sends the first message in the coin-tossing protocol. Then, Alice updates her defense coin every odd round, and Bob updates his defense coin every even round.

A crucial property of information-theoretic protocols is the following. The expectation of Alice's defense coin (conditioned on the partial transcript) does not change after Bob sends the next message in the protocol, and (likewise) the expectation of Bob's defense coin does not change after Alice sends the next message in the protocol. For example, the expected value of Bob's defense coin immediately before and after Alice sends her message in round 3 is the same. Previous works consider the *message exposure filtration* $\{\emptyset, \mathcal{T}\} = M_0 \subseteq M_1 \subseteq \cdots \subseteq M_n = 2^{\mathcal{T}}$ corresponding to the protocol.[1] They identify the susceptible (randomized) round τ witnessing an $\Omega\left(\frac{X_0(1-X_0)}{\sqrt{n}}\right)$ gap in the martingale. Next, they use this round τ as a template to identify a fail-stop attack on the coin-tossing protocol and change the output distribution by $\Omega\left(\frac{X_0(1-X_0)}{\sqrt{n}}\right)$. This transference crucially relies on the fact that the expectation of the defense coin of the receiver in round τ immediately before and after the τ-th message is identical.

Now consider the scenario where parties update their defense coins lazily. Suppose the parties update their defenses in rounds $1 \leqslant i_1 < i_2 < \cdots < i_d \leqslant n$. We clarify that the rounds $\{i_1, i_2, \ldots, i_d\}$ can be randomized as well, i.e., they depend on the partial transcripts during the protocol execution, refer to Figure 2. Furthermore, note that the parity of the round i_k implicitly identifies the party updating her defense coin. The *randomized defense rounds* are very natural to consider. For example, continuing the motivating example from the introduction, the next message computation of a delegated protocol may depend on the partial transcript of the delegated protocol. If, for instance, the protocol evolves into a state where the next-message-generation becomes time consuming for the processes of the cloud, then Alice and Bob can use this opportunity to reduce their lag in the knowledge of the protocol's evolution.

Suppose one considers the message exposure filtration $\{\emptyset, \mathcal{T}\} = M_0 \subseteq M_1 \subseteq \cdots \subseteq M_n = 2^{\mathcal{T}}$, then the fail-stop attack shall ensure that the output distribution changes only by $\Omega(X_0(1-X_0)/\sqrt{n})$. On the other hand, one can instead consider the filtration $\{\emptyset, \mathcal{T}\} = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_d \subseteq F_{d+1} = 2^{\mathcal{T}}$, where F_k (for $1 \leqslant k \leqslant d$) corresponds to exposing all the protocol messages up to (the randomized) round i_k, and F_{d+1} represents exposing the full transcript. We emphasize that the σ-field F_k may simultaneously expose multiple rounds of messages sent by both the parties in addition to the messages already exposed by F_{k-1}. Let $\mathcal{Y} = (Y_0 = X_0, Y_1, \ldots, Y_d, Y_{d+1})$ represent the martingale such that Y_k is the expectation of the outcome conditioned on the first i_k messages in the protocol. Note that Y_{d+1} is the expected outcome at the end of the protocol and, therefore, we have $Y_{d+1} \in \{0, 1\}$.

Indeed, by applying [CI93, KKR18, BHMO18, KMM19], there exists a $\tau \in \{1, 2, \ldots, d, d+1\}$ for this filtration such that the gap in parties' knowledge of the outcome between rounds

[1] The set $\mathcal{T}$ represents the set of all possible transcripts in the protocol. The set $2^{\mathcal{T}}$ represents the set of all possible subsets of $\mathcal{T}$.

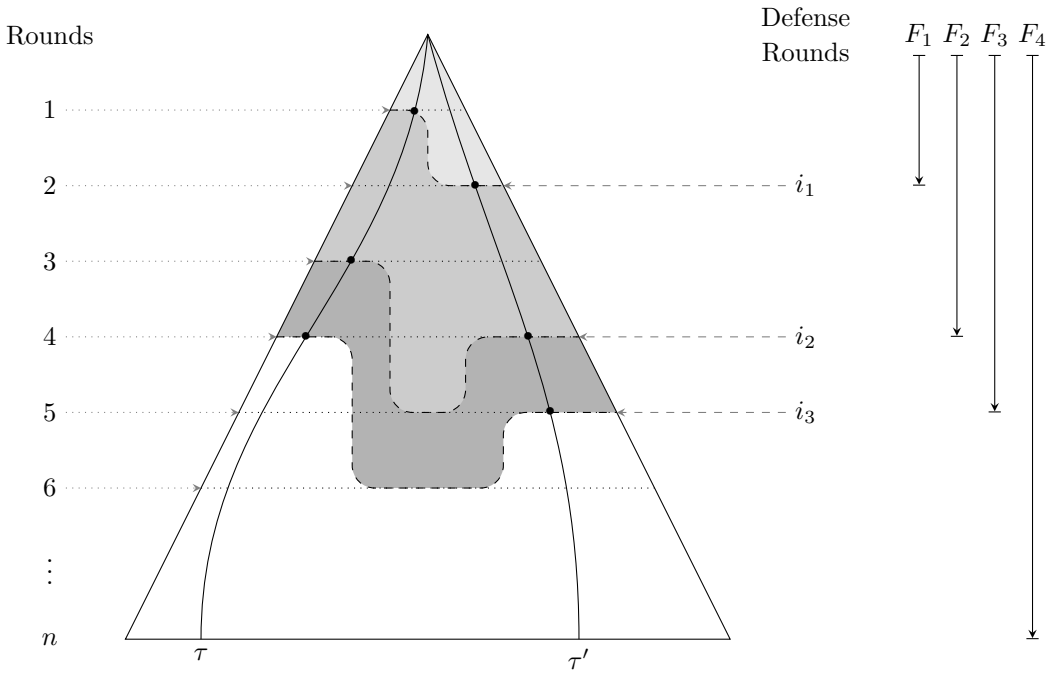

Figure 2 The tree represents the protocol tree of the coin-tossing protocol. Gray dotted lines represent the n rounds of the protocol. The dashed lines represent the $d = 3$ defense rounds. Any complete protocol execution (root to leaf path in the tree) encounters the defense rounds i_1, i_2, and i_3 in that particular order. For example, we consider two transcripts τ and τ', and illustrate that they encounter i_1, i_2, and i_3 in that order. The σ-fields F_1, F_2, and F_3 expose messages till encountering i_1, i_2, and i_3, respectively. The σ-field F_4 reveals the entire protocol transcript.

$i_{\tau-1}$ and i_τ is $\Omega\left(X_0(1-X_0)/\sqrt{d}\right)$. However, the transference of τ into a fail-stop attack on the coin-tossing protocol fails. This failure is attributable to the fact that the expectation of the defense coins of *both* parties may change between rounds $i_{\tau-1}$ and i_τ because F_τ may expose messages by both parties in addition to the messages already exposed by $F_{\tau-1}$ (refer to Figure 4 and Figure 5 for concrete examples).

Towards resolving this impasse, we employ a new potential function enabling an inductive proof for this problem; generalizing the approach of Khorasgani et al. [KMM19]. This new proof considers the message exposure filtration $\{\emptyset, \mathcal{T}\} = M_0 \subseteq M_1 \subseteq \cdots \subseteq M_n = 2^{\mathcal{T}}$ while, simultaneously, ensuring a fail-stop attack on (information-theoretic) coin-tossing protocols that changes their output distribution by $\Omega\left(X_0(1-X_0)/\sqrt{d}\right)$. Finally, these attacks naturally translate into black-box separation results for (appropriately restricted-versions of) fair coin-tossing protocols as considered in [DLMM11, HOZ13, DMM14].

1.2 Our Contributions

A *2-party* (n, d)*-coin-tossing protocol with bias-*X_0 is an n-round 2-party coin-tossing protocol (with output $\{0, 1\}$) such that the expected outcome of the protocol is X_0, and parties update their defense coins in d rounds. The defense complexity d is a function of the round complexity n. Furthermore, the decision of a party to update her defense coin may depend on the partial transcript of the protocol itself. A protocol is ε*-unfair* if there exists a fail-stop

strategy for one of the parties to deviate the output distribution of the honest party by ε in statistical distance. Our main result is the following theorem.

▶ **Theorem 1** (Attacks on Coin-Tossing Protocol with Lazy Defense). *There exists a universal positive constant c, such that for any $X_0 \in [0,1]$ and 2-party (n,d)-coin-tossing protocol with bias-X_0 is (at least) $c \cdot X_0(1-X_0)/\sqrt{d}$-unfair.*

Before our result, we knew that 2-party (n,n)-coin-tossing protocol with bias-X_0 is $\Omega(X_0(1-X_0)/\sqrt{n})$ unfair [CI93, BHMO18, KMM19]. Our work, motivated by interesting cryptographic applications discussed in the introduction, decouples the round complexity and the defense complexity of coin-tossing protocols for a more fine-grained study of fair coin-tossing functionalities. We show that the defense complexity, not the round complexity, of coin-tossing protocols determines the security of coin-tossing protocols. For example, a coin-tossing protocol with high round complexity but a small defense complexity shall be very unfair. In particular, when $d = o(n)$ it is impossible for an n-round coin-tossing protocol to be $\mathcal{O}(1/\sqrt{n})$-unfair, for a constant $X_0 \in (0,1)$.

Finally, this fail-stop attack on coin-tossing protocols translates into black-box separation results. Existing techniques leverage the fail-stop attack of Theorem 1 to rule out the construction of fair coin-tossing protocols by using one-way functions in a black-box manner for a broad class of protocols.

▶ **Corollary 1** (Black-box Separation). *There exists a universal positive constant c such that, for any $X_0 \in [0,1]$, there is no construction of a 2-party (n,d)-coin-tossing protocols with bias-X_0 that is $< c \cdot X_0(1-X_0)/\sqrt{d}$-unfair and uses one-way functions in a black-box manner (restricted to the classes of protocols considered by [DLMM11, HOZ13, DMM14]).*

When $d = o(n)$, our corollary provides new black-box separation results that do not follow from prior techniques.

1.3 Technical Overview

Our proof proceeds by induction on the defense complexity d of the coin-tossing protocol to lower bound the performance of the *best fail-stop attack* on the protocol. The proof of the inductive step for this result proceeds by another induction on the number of rounds m until the first time a party updates her defense coins. In particular, this second-level induction crucially *avoids degrading* the estimate of the best fail-stop attack's performance on the coin-tossing protocol *as a function of* m. In effect, the quality of the fail-stop attack depends only on d and is insensitive to the round complexity n of the protocol, thus circumventing the hurdles encountered by previous works.

1.3.1 Score Function & Inductive Hypothesis

Consider any n-round coin-tossing protocol π with bias-X_0 and defense complexity d, the inductive argument maintains a *lower bound* to the performance of the *best cumulative attack* possible on this coin-tossing protocol.

For any stopping time[2] τ in the protocol, we associate a score to τ measuring its susceptibility to fail-stop attacks. For a partial transcript $v \in \tau$, its contribution to the score is the sum of the change in the output distribution that Alice can cause by aborting

[2] A stopping time in a protocol is a set of prefix-free partial transcripts.

at that partial transcript, and the change in the output distribution that Bob can cause by aborting at that partial transcript. We emphasize that the same partial transcript $v \in \tau$ may contribute to both Alice and Bob's attacks. Explicitly, the two possible attacks are as follows. The sender of the message can abort after generating (but *not sending*) the last message of the partial transcript v. The receiver may abort immediately after receiving the last message of the partial transcript v. Both these fail-stop strategies may be effective attacks in the scenario of coin-tossing protocols with lazy defense. The score of a stopping time τ is the sum of all the contributions of $v \in \tau$. The optimal score associated with a protocol π, represented by $\mathsf{Opt}(\pi)$, is the maximum score achievable by a stopping time in that protocol.

Using induction on d, we prove that, for any protocol π with bias-X_0 and defense complexity d,

$$\mathsf{Opt}(\pi) \geqslant \Gamma_{2d} \cdot X_0(1 - X_0),$$

where $\Gamma_i = \frac{1}{\sqrt{(\sqrt{2}+1)(i+2)}}$ (refer to Theorem 3). We remark that, indeed, it is possible to tighten the constants involved in the lower bound with a more careful analysis; however, such a tighter analysis does not qualitatively improve the bound.[3]

1.3.2 Base Case: $d = 0$

In the case when the defense complexity of π is $d = 0$, parties begin with their respective default defense coins and never update them. Irrespective of the round complexity n, our objective is to demonstrate a fail-stop attack on the $(n, d = 0)$-coin-tossing protocol with bias-X_0 that changes the output distribution of the honest party by $\Omega(X_0(1 - X_0))$ statistical distance. Note that obtaining an attack whose effectiveness does not degrade with the round complexity n even for this simple case cannot be obtained from previous techniques [CI93, BHMO18, KMM19] (refer to the protocol in Figure 4).

We proceed by induction on the round complexity n. Consider the base case of $n = 1$, a one-round protocol where Alice sends the first message in the protocol. Section 4.1.1 proves that $\mathsf{Opt}(\pi) \geqslant X_0(1 - X_0)$.

The inductive step for $n \geqslant 2$ is the non-trivial part of the proof. The case of $n = 2$ is representative enough to highlight all the key ideas to handle any general $n \geqslant 2$. Alice and Bob have defense coins such that their respective expected values are D_0^{A} and D_0^{B} before the protocol began. Alice sends the first message in the protocol, and Bob sends the second message in the protocol. Suppose the first message sent by Alice is $M_1 = i$, which happens with probability $p^{(i)}$. Conditioned on the first message being $M_1 = i$, the expected value of (a) the outcome be $x_1^{(i)}$, (b) Alice's defense coin be $d_1^{\mathsf{A},(i)}$, and (c) Bob's defense coin be D_0^{B}. By aborting at the message $M_1 = i$, we obtain the following contribution to the score[4]

$$\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|.$$

By deferring the attack to the residual $(n - 1)$ round protocol conditioned on $M_1 = i$, by the

[3] One can convert the optimal d-round protocol of Khorasgani et al. [KMM19] to construct an (n, d)-protocol that makes progress only when parties update their defense coin; thus, demonstrating the qualitative optimality of our lower bounds.

[4] Recall that we are considering the sum of the change in the output distribution caused by Alice when she aborts (which is $\left|x_1^{(i)} - D_0^{\mathsf{B}}\right|$) and the change in the output distribution caused by Bob when he aborts (which is $\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right|$).

inductive hypothesis, we obtain the following contribution to the score

$$\geqslant x_1^{(i)}\left(1-x_1^{(i)}\right).$$

The optimal stopping time can ensure the maximum of these two contributions, thus, obtaining a contribution of

$$\geqslant \max\left(\left|x_1^{(i)}-d_1^{\mathsf{A},(i)}\right|+\left|x_1^{(i)}-D_0^{\mathsf{B}}\right|\ ,\ x_1^{(i)}\left(1-x_1^{(i)}\right)\right).$$

We prove a key technical lemma[5] (Lemma 1) proving the following lower bound to the quantity above.

$$\geqslant \frac{1}{2}\cdot\left(x_1^{(i)}\left(1-x_1^{(i)}\right)+\left(x_1^{(i)}-d_1^{\mathsf{A},(i)}\right)^2+\left(x_1^{(i)}-D_0^{\mathsf{B}}\right)^2\right).$$

Overall, at the root of the protocol tree, the score of the optimal stopping-time is lower-bounded by

$$\sum_i p^{(i)}\cdot\frac{1}{2}\cdot\left(x_1^{(i)}\left(1-x_1^{(i)}\right)+\left(x_1^{(i)}-d_1^{\mathsf{A},(i)}\right)^2+\left(x_1^{(i)}-D_0^{\mathsf{B}}\right)^2\right).$$

Let us define the multivariate function f below

$$f(x,y,z) := x(1-x)+(x-y)^2+(x-z)^2.$$

Let $f_z(x,y)$ represent the function $f(x,y,z)$ where z is a constant. Then, the function $f_z(x,y)$ is convex. Likewise, the function $f_y(x,z)$ is also convex. Recall that $\sum_i p^{(i)}\cdot x_1^{(i)} = X_0$ and $\sum_i p^{(i)}\cdot d_1^{\mathsf{A},(i)} = D_0^{\mathsf{A}}$. Note that D_0^{B} is a constant, and, therefore, one can use Jensen's inequality on f, to push the expectation inside, obtaining the following lower bound.

$$\geqslant \frac{1}{2}\cdot\left(X_0\left(1-X_0\right)+\left(X_0-D_0^{\mathsf{A}}\right)^2+\left(X_0-D_0^{\mathsf{B}}\right)^2\right).$$

This bound is minimized when $D_0^{\mathsf{A}} = X_0$ and $D_0^{\mathsf{B}} = X_0$. So, we obtain the lower-bound

$$\geqslant \frac{1}{2}\cdot X_0\left(1-X_0\right).$$

For $n > 2$, we rely on the fact that the expected value of the receiver's defense coin in every round does not change. So, Jensen's inequality applies to f, and we move the lower-bound one round closer to the root. Iterative application of Jensen's inequality brings the lower-bound to the root, where it is identical to the expression above and is independent of the round complexity n of the protocol. Subsection 4.1 and Appendix C.1 provides full proof.

1.3.3 Inductive Step for $d \geqslant 1$

The inductive step of the proof shall proceed by induction on $m \geqslant 1$, the round where a party first updates her defense coins. Again, our objective is to obtain a lower bound that is independent of m.

[5] As an aside, we remark that this technical lemma is sufficiently powerful and immediately subsumes the lower bounds of Khorasgani et al. [KMM19].

Consider the base case of $m = 1$. So, we have an (n, d)-coin-tossing protocol with bias-X_0 and Alice sends the first message and updates her defense. Suppose the first message set by Alice is $M_1 = i$, which happens with probability $p^{(i)}$. Conditioned on the first message being $M_1 = i$, the expected value of (a) the outcome be $x_1^{(i)}$, (b) Alice's *updated* defense coin be $d_1^{\mathsf{A},(i)}$, and (c) Bob's defense coin be D_0^{B}. In the remaining subprotocol, there are only $d-1$ defense updates. Therefore, the score in that subprotocol is at least $\Gamma_{2(d-1)} \cdot x_1^{(i)}(1 - x_1^{(i)})$, by the induction hypothesis. So, using arguments similar to the case of $d = 0$ and $n = 2$ presented above, by appropriately deciding to either abort at $M_1 = i$ or deferring the attack to the subtree we get a score of

$$\geqslant \max\left(\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right| \ , \ \underbrace{\Gamma_{2(d-1)} \cdot x_1^{(i)}\left(1 - x_1^{(i)}\right)}_{\text{By inductive hypothesis}}\right).$$

Using Lemma 1 and Jensen's inequality (because the first message does not reveal any information regarding Bob's default defense), we conclude that there is a stopping time with score

$$\geqslant \Gamma_{2(d-1)+1} \cdot \left(X_0\left(1 - X_0\right) + \left(X_0 - \sum_i p^{(i)} \cdot d_1^{\mathsf{A},(i)}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2\right) \geqslant \Gamma_{2d-1} \cdot X_0\left(1 - X_0\right).$$

Observe that the expected value of the "updated Alice defense coin" appears in the first lower bound above; instead of the expected value of Alice's default defense D_0^{A}. However, the final lower bound does *not* depend on the updated defense.

Finally, consider the inductive step of $m \geqslant 2$. The special case of $m = 2$ illustrates all the primary ideas. So, we have an (n, d)-coin-tossing protocol with bias-X_0, Alice sends the first message, and Bob sends the second message and updates his defense coin. Suppose the first message set by Alice is $M_1 = i$, which happens with probability $p^{(i)}$. Conditioned on the first message being $M_1 = i$, the expected value of (a) the outcome be $x_1^{(i)}$, (b) Alice's defense coin be $d_1^{\mathsf{A},(i)}$, and (c) Bob's defense coin be D_0^{B}. For every i, using the above argument, we get that there exists a stopping time in the subprotocol rooted at $M_1 = i$ with a score of

$$\geqslant \Gamma_{2d-1} \cdot x_1^{(i)}(1 - x_1^{(i)}).$$

So, a stopping time by deciding to either abort at $M_1 = i$ or deferring the attack to a later point in time can obtain score of

$$\begin{aligned}
&\geqslant \max\left(\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right| \ , \ \underbrace{\Gamma_{2d-1} \cdot x_1^{(i)}\left(1 - x_1^{(i)}\right)}_{\text{By previous argument}}\right) \\
&\geqslant \Gamma_{2d} \cdot \left(X_0\left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2\right)
\end{aligned}$$

The last inequality is an application of Lemma 1 and the fact that the expected value of Alice defense coins at the end of first round is identical to D_0^{A} (because Alice does not update her defense coins in the first round).

For $m > 2$, we use Jensen's inequality on f and the fact that the receiver's defense coins do not update in the protocol to rise one round up in the protocol tree. In this step, the constant Γ_{2d} does not change. So, iterating this procedure, we reach the root of the protocol

where we get a lower-bound of

$$\Gamma_{2d} \cdot \left(X_0 \left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right) \geqslant \Gamma_{2d} \cdot X_0(1 - X_0)$$

for the maximum score of the stopping time. Subsection 4.2 and Appendix C.2 provides the full proof.

1.3.4 Randomized Defense Coin Update Rounds

In general, the round where parties update their defense coin in a coin-tossing protocol may be randomized. More formally, parties in a coin-tossing protocol decide on updating their defense coins as follows. Suppose the partial transcript generated so far in the protocol is $(M_1, M_2, \ldots, M_i)$. The party sending the next message M_{i+1} in the protocol decides whether to update her defense coin or not based on the partial transcript $(M_1, M_2, \ldots, M_i)$. If the party decides to update her defense coin, then she updates her defense coin based on her private view.

The defense complexity of a coin-tossing protocol with randomized rounds for defense coin updates is (at most) d if during the generation of any complete transcript $(M_1, M_2, \ldots, M_n)$ the total number of defense coin updates is $\leqslant d$. The proofs mentioned above generalize to this setting naturally. Section 5 (using Figure 3 for intuition) extends the proofs outlined above to this general setting.

2 Preliminaries

2.1 Martingales and Related Definitions

Suppose (X, Y) is a discrete joint distribution, then the conditional expectation of X given that $Y = y$, for any y such that $\Pr[Y = y] > 0$, is defined as $\mathrm{E}[X|Y = y] = \sum_x x \cdot \Pr[X = x|Y = y]$ where $\Pr[X = x|Y = y] = \frac{\Pr[X=x, Y=y]}{\Pr[Y=y]}$. The conditional expectation of X given Y, denoted by $\mathrm{E}[X|Y]$, is defined as the random variable that takes value $\mathrm{E}[X|Y = y]$ with probability $\Pr[Y = y]$.

A discrete time random process $\{X_i\}_{i=0}^n$ is a sequence of random variables where the random variable X_k denotes the value of process at time k.

Let $(M_1, M_2, \ldots, M_n)$ be a joint distribution defined over sample space $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ such that for any $i \in \{1, \ldots, n\}$, M_i is a random variable over Ω_i. A random variable X_j defined over Ω is said to be $M_1, \ldots, M_j$ measurable if there exists a deterministic function $f_j : \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n \to \mathbb{R}$ such that $X_j = f_j(M_1, M_2, \ldots, M_j)$ i.e. the value of X_j is determined by the random variables $M_1, M_2, \ldots, M_j$ and in particular, it does not depend on random variables $M_{j+1}, \ldots, M_n$. A discrete time random process $\{X_i\}_{i=0}^n$ is said to be a discrete time martingale with respect to another sequence $\{M_i\}_{i=1}^n$ if it satisfies the two following conditions for any time values $1 \leqslant k \leqslant n$ and $0 \leqslant r \leqslant \ell$:

$$\mathrm{E}[|X_k|] < \infty$$
$$\mathrm{E}[X_\ell | M_1, M_2, \ldots, M_r] = X_r$$

which means that at any time, given the current value and all values from the past, the conditional expectation of random process at any time in the future is equal to the current value. For such a martingale, a random variable $\tau \colon \Omega \to \{0, 1, \ldots, n\}$ is called a stopping time if the random variable $1_{\{\tau \leqslant k\}}$ is $M_1, \ldots, M_k$ measurable. One can verify that for a given function $g : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$, the random sequence $\{Z_i\}_{i=0}^n$ where for each i,

$Z_i = \mathrm{E}[f(M_1, \ldots, M_n)|M_1, \ldots, M_i]$ is a martingale with respect to the sequence $\{M_i\}_{i=1}^n$. This martingale is called the *Doob's martingale*.

2.2 Coin-tossing Protocols with Fixed Defense Rounds

Let us first define the coin-tossing protocol with fixed defense rounds.

▶ **Definition 1** (($X_0, n, \mathcal{A}, \mathcal{B}$)-coin tossing protocol)**.** Let π be an n-round coin tossing protocol, where Alice and Bob speak in alternate rounds to determine the outcome of the tossing of a X_0-bias coin, i.e., the probability of head is X_0. Without loss of generality, assume that Alice sends the first message. Therefore, Alice (resp., Bob) will be speaking in the odd (resp., even) rounds. Let $\mathcal{A} \subseteq [n] \cap \mathsf{Odd}$ and $\mathcal{B} \subseteq [n] \cap \mathsf{Even}$.[6] During the protocol execution, Alice and Bob shall defend in the following manner.

- Alice and Bob both prepare a defense before the beginning of the protocol based on their private tape. We refer to this defense as Alice's and Bob's defense at round 0.
- At any round $i \in [n]$, if Alice is supposed to speak (i.e., $i \in \mathsf{Odd}$) and $i \in \mathcal{A}$, she shall prepare a new defense based on her private view, which is, her private tape and the first $i-1$ messages exchanged. Otherwise, i.e., $i \notin \mathcal{A}$, she shall not prepare a new defense and simply set her defense for the previous round as her defense for this round. That is, Alice keeps her defense unchanged for this round. Bob's defense is prepared in the similar manner.
- At an odd round $i \in [n]$, Alice is supposed to speak and she might decide to abort the protocol. If Alice aborts, Bob shall output his defense for this round as defined above. Alice's output when Bob aborts is defined in the similar manner.

For brevity, we refer to such coin-tossing protocols as an $(X_0, n, \mathcal{A}, \mathcal{B})$-*coin tossing protocol*. We refer to the expectation of the outcome of the protocol, i.e., X_0, as the *root-color*. We refer to the size of the set $\mathcal{A} \cup \mathcal{B}$ as the *defense complexity* of the coin-tossing protocol.[7]

We provide a few representative examples in Appendix A. The following remarks provide additional perspectives to this definition.

▶ **Remark 1.** We clarify that a party *does not* update her defense during a round where she does not send a message in the protocol. For example, at an odd round i, Bob does not update his defense. This is because Bob's private view at round i, i.e., Bob's private tape, and the first $i-1$ messages, is a *deterministic function* of Bob's private view at round $i-1$, i.e., Bob's private tape and the first $i-2$ messages. Therefore, Bob's defense strategy to update his defense at round i is *simulatable* by a defense strategy to update his defense at round $i-1$. Hence, without loss of generality, parties only update their respective defenses during a round that they are supposed to speak. This simplification *shall not* make the protocol any more vulnerable.

▶ **Remark 2.** In particular, if we set $\mathcal{A}$ to be $[n] \cap \mathsf{Odd}$ and $\mathcal{B}$ to be $[n] \cap \mathsf{Even}$, this is the *fair coin-tossing protocol* that has been widely studied in the literature.

---

[6] We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. Odd (resp., Even) represents the set of all odd (resp., even) positive integers.

[7] Note that the defense complexity is less than or equal to the round complexity.

2.2.1 Notation

Let us denote the message exchanged between two parties in an n-round protocol by $M_1, M_2, \ldots, M_n$. For $i \in [n]$, let X_i be the expected outcome conditioned on the first i messages, i.e., $M_1, \ldots, M_i$. We also refer to the expected outcome X_i as the color at time i. Let D_i^{A} (resp., D_i^{B}) represents the expectation of the Alice's (resp., Bob's) defense at round i conditioned on the first i messages. Note that X_i, D_i^{A} and D_i^{B} are $M_1, \ldots, M_i$ measurable. In particular, X_0, D_0^{A} and D_0^{B} are constants.

Throughout our proof, the following inequality will be useful.

▶ **Theorem 2** (Jensen's inequality). *If f is a multivariate convex function, then $\mathrm{E}\left[f\left(\vec{X}\right)\right] \geqslant f\left(\mathrm{E}\left[\vec{X}\right]\right)$, for all probability distributions $\vec{X}$ over the domain of f.*

3 Our Results on Fixed Defense Rounds

In this section, we shall present our main results on the coin-tossing protocols with *fixed* defense rounds. In Section 5, we present how one can generalize the proof strategies to coin-tossing protocols with *randomized* defense rounds.

Intuitively, our results state that the vulnerability of a coin-tossing protocol depends solely on the defense complexity and is irrespective of the round complexity.

Let us first define the following score function which captures the susceptibility of a protocol with respect to a stopping time.

▶ **Definition 2.** Let π be a $(X_0, n, \mathcal{A}, \mathcal{B})$-coin tossing protocol. Let $\mathsf{P} \in \{\mathsf{A}, \mathsf{B}\}$ be the party who sends the last message of the protocol. For any stopping time τ, define

$$\mathsf{Score}(\pi, \tau) := \mathrm{E}\left[\mathbb{1}_{(\tau \neq n) \vee (\mathsf{P} \neq \mathsf{A})} \cdot \left|X_\tau - D_\tau^{\mathsf{A}}\right| + \mathbb{1}_{(\tau \neq n) \vee (\mathsf{P} \neq \mathsf{B})} \cdot \left|X_\tau - D_\tau^{\mathsf{B}}\right|\right].$$

We clarify that the binary operator $\vee$ in the expression above represents the boolean OR operation.

The following remarks provide additional perspectives to this definition.

▶ **Remark 3.** Suppose we are in a round τ, where Alice is supposed to speak. The color X_τ corresponds to Alice's message being m_τ^*. We note that, in a coin-tossing protocol with lazy defense, *both* Alice and Bob can deviate the outcome by aborting appropriately. Alice can attack by aborting when her next message turns out to be m_τ^* without sending it to Bob. By our definition, this attack ensures a deviation of $\left|X_\tau - D_\tau^{\mathsf{B}}\right|$. On the other hand, Bob can also attack this message by aborting the next round upon receiving the message m_τ^*. This attack might be successful because Alice's defense is *lazy* and she *does not* update her defense at round τ. Bob's attack will deviate the distribution of the outcome by $\left|X_\tau - D_{\tau+1}^{\mathsf{A}}\right|$. However, note that Alice is not supposed to speak at the $(\tau+1)^{th}$ round, her defense at $(\tau+1)^{th}$ round is identical to her defense at τ^{th} round. Hence, the deviation of Bob's attack is also $\left|X_\tau - D_\tau^{\mathsf{A}}\right|$. We emphasize that, in fair coin-tossing protocols where parties update their defenses *every round*, this attack by Bob, possibly, is ineffective.

▶ **Remark 4.** We note that the above remark has a boundary case, i.e., the last message of the protocol. Without loss of generality, assume that Alice sends the last message of the protocol. Note that, unlike previous messages, Bob cannot abort anymore after receiving the last message from Alice, since the protocol has ended. Therefore, our score function should exclude $\left|X_\tau - D_\tau^{\mathsf{A}}\right|$ when $\tau = n$. Hence, in the definition of our score function, we

have an indicator function $\mathbb{1}$. Intuitively, this boundary case needs to be accounted in our score; however, we emphasize that, this boundary case does not significantly alter our proof strategy.

▶ **Remark 5.** Looking ahead, we elaborate how one translates our score function into fail-stop attacks by Alice and Bob. Fix a stopping time τ that witnesses a large susceptibility. To construct the attacks by Alice, we *partition* the stopping time τ into two sets depending on whether $X_\tau \geqslant D^{\mathsf{A}}_\tau$ or not. Similarly, for Bob's attacks, we *partition* the stopping time τ into two sets depending on whether $X_\tau \geqslant D^{\mathsf{B}}_\tau$ or not. These four (fail-stop) attack strategies correspond to Alice or Bob deviating the outcome towards 0 or 1, respectively. Note that the sum of the biases achieved by these four attacks is identical to the score function. Therefore, by averaging arguments, one of these four attacks can deviate the protocol by at least $\frac{1}{4} \cdot \mathsf{Score}\,(\pi, \tau)$. We clarify that, in light of Remark 3, the portions of the stopping time τ that contribute to Alice attacks and the portions that contribute to Bob attacks *need not be* mutually exclusive.

Given an $(X_0, n, \mathcal{A}, \mathcal{B})$-coin-tossing protocol π, we are interested in the optimal stopping time τ that maximizes $\mathsf{Score}\,(\pi, \tau)$. This quantity represents the susceptibility of the protocol. Hence, we have the following definition.

▶ **Definition 3.** For any coin-tossing protocol π, we define

$$\mathsf{Opt}(\pi) := \max_{\tau} \mathsf{Score}(\pi, \tau).$$

With these definitions, we are ready to present our main theorem, which states the following.

▶ **Theorem 3.** *For all root-color $X_0 \in [0,1]$ and defense complexity $d \in \mathbb{N}$, and any $(X_0, n, \mathcal{A}, \mathcal{B})$-coin-tossing protocol π where $d = |\mathcal{A} \cup \mathcal{B}|$, we have*

$$\mathsf{Opt}(\pi) \geqslant \Gamma_{2d} \cdot X_0(1 - X_0),$$

where $\Gamma_i := \frac{1}{\sqrt{(\sqrt{2}+1)(i+2)}}$ for all $i \in \{0, 1, \dots\}$.

Asymptotically, we have $\Gamma_i \gtrapprox 0.64/\sqrt{i}$. Note that the lower bound is only associated with the root-color X_0 and defense complexity d of the protocol π.

We present the proof of Theorem 3 in Section 4. In light of Remark 5 above, we can directly translate this theorem into a fail-stop attack strategy.

▶ **Corollary 2.** *For any $(X_0, n, \mathcal{A}, \mathcal{B})$-coin-tossing protocol, with defense complexity d, there exists a fail-stop attack strategy for either Alice or Bob that deviates the protocol by at least*

$$\frac{1}{4} \cdot \frac{X_0(1 - X_0)}{\sqrt{(\sqrt{2}+1)\,(2d+2)}}.$$

4 Proof of Theorem 3

In this section, we shall prove Theorem 3 using mathematical induction on the defense complexity d of the coin-tossing protocol. In Subsection 4.1, we prove the base case, i.e., $d = 0$. In Subsection 4.2, we prove the inductive step. We stress that although the base case is conceptually simple, its proof already captures most of the technical challenges involved in proving the general inductive step.

Throughout the proof, we use the following key technical lemma repeatedly. We defer the proof of Lemma 1 to Appendix B.

▶ **Lemma 1** (Key technical Lemma). *For all $P \in [0,1]$ and $Q \in [0,1/2]$, if P, Q satisfies*

$$P - Q - P^2 Q \geqslant 0,$$

then, for all $x, \alpha, \beta \in [0,1]$, we have

$$\max\left(P \cdot x(1-x)\ ,\ |x-\alpha| + |x-\beta|\right) \geqslant Q \cdot \left(x(1-x) + (x-\alpha)^2 + (x-\beta)^2\right).$$

In particular, for any $k \geqslant 1$, the constraints are satisfied, if we set $P = \Gamma_{k-1} := \frac{1}{\sqrt{(\sqrt{2}+1)(k+1)}}$ and $Q = \Gamma_k := \frac{1}{\sqrt{(\sqrt{2}+1)(k+2)}}$.

4.1 Base Case: $d = 0$

The base case is that the defense complexity d is 0, i.e., both $\mathcal{A}$ and $\mathcal{B}$ are empty sets, and hence parties only prepare their defenses before the beginning of the protocol and never update it (see the example in Figure 4).

To prove the base case, we shall prove the following stronger statement that clearly implies that Theorem 3 is correct for the base case. We prove the following lemma by induction on the round complexity n, where $n = 1$ and $n = 2$ serve as the base cases.

▶ **Lemma 2** (Base Case of $d = 0$). *For any n-round protocol π with defense complexity $d = 0$,*

1. *If $n = 1$,*
$$\mathsf{Opt}(\pi) \geqslant X_0\,(1 - X_0)\,.$$
2. *If $n \geqslant 2$,*
$$\mathsf{Opt}(\pi) \geqslant \frac{1}{2} \cdot \left(X_0\,(1-X_0) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2\right).$$

▶ **Remark 6.** We remark that $D_0^{\mathsf{A}} = D_0^{\mathsf{B}} = X_0$ is the only Alice's and Bob's defense that optimizes our lower bound for the $n \geqslant 2$ case. In general, we *do not* claim that they are the *optimal defenses* that minimize the score of the optimal stopping time. Our bound is simply a lower bound.

4.1.1 Round Complexity $n = 1$

Let us start with the simplest case, i.e., when $n = 1$. Here, we have a one-round protocol π. Without loss of generality, assume that Alice sends the only message. The only attack is by Alice to abort her message and thus we pick our stopping time to be $\tau = 1$. This gives us

$$\mathsf{Score}(\pi, \tau) = \mathrm{E}\left[\left|X_1 - D_1^{\mathsf{B}}\right|\right].$$

Recall that $X_1 \in \{0,1\}$ and $\Pr[X_1 = 1] = X_0$. Moreover, regardless of what Alice's first message is, the expectation of Bob's defense for the first round, i.e., D_1^{B}, remains the same and is exactly the expectation of his defense at the beginning of the protocol, i.e., D_0^{B}. Therefore,

$$\mathsf{Score}(\pi, \tau) = (1 - X_0) \cdot \left|0 - D_0^{\mathsf{B}}\right| + X_0 \cdot \left|1 - D_0^{\mathsf{B}}\right|.$$

To lower-bound the score mentioned above, observe that

$$(1 - X_0)D_0^{\mathsf{B}} + X_0(1 - D_0^{\mathsf{B}}) \geqslant X_0(1 - X_0) + (X_0 - D_0^{\mathsf{B}})^2 \geqslant X_0(1 - X_0).$$

Hence, for any coin-tossing protocol π with $n = 1$, $\mathsf{Opt}(\pi) \geqslant X_0\,(1 - X_0)$.

4.1.2 Round Complexity $n = 2$

Next, we consider the case when $n = 2$. Let π be a two-round protocol, where Alice sends the first message and Bob sends the second message. Without loss of generality, assume that there are ℓ possible first messages that Alice can send, namely $\{1, 2, \ldots, \ell\}$. The probability of the first message being i, i.e., $M_1 = i$, is $p^{(i)}$. For all $i \in [\ell]$, conditioned on first message being i, let $X_1 = x_1^{(i)}$ and $D_1^{\mathsf{A}} = d_1^{\mathsf{A},(i)}$. Again, regardless of what Alice's first message is, the expectation of Bob's defense D_1^{B} remains the same as D_0^{B}. Therefore, if we stop at message $M_1 = i$, this contributes to our score function by

$$\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|.$$

On the other hand, conditioned on Alice's first message being i, the remaining protocol is exactly a one-round protocol with root-color $x_1^{(i)}$. By our analysis above, the optimal stopping time for this sub-protocol will yield a score of at least $x_1^{(i)}\left(1 - x_1^{(i)}\right)$. Hence, the optimal stopping time will decide on whether to stop at first message being i or continue to a stopping time in the mentioned sub-protocol, depending on which of these two strategies yield a larger score. This will contribute to the score function by at least

$$\max\left(\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|, x_1^{(i)}\left(1 - x_1^{(i)}\right)\right).$$

Using Lemma 1 with $P = 1$ and $Q = 1/2$, we get

$$\begin{aligned}
&\max\left(\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|, x_1^{(i)}\left(1 - x_1^{(i)}\right)\right) \\
\geqslant &\frac{1}{2} \cdot \left(x_1^{(i)}\left(1 - x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2\right).
\end{aligned}$$

Therefore, the optimal stopping time will have score

$$\begin{aligned}
&\sum_{i=1}^{\ell} p^{(i)} \cdot \max\left(\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|, x_1^{(i)}\left(1 - x_1^{(i)}\right)\right) \\
\geqslant &\frac{1}{2} \cdot \sum_{i=1}^{\ell} p^{(i)} \cdot \left(x_1^{(i)}\left(1 - x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2\right) \\
\overset{\text{(i)}}{\geqslant} &\frac{1}{2} \cdot \left(X_0\left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2\right),
\end{aligned}$$

Let us elaborate on inequality (i).

1. One can verify that for any constant c, the function $f(x, y) := x(1-x)+(x-y)^2+(x-c)^2$ is a bivariate convex function. The Hessian matrix of f is positive semi-definite.
2. Since (X_0, X_1) forms a martingale, we have $\sum_{i=1}^{\ell} p^{(i)} \cdot x_1^{(i)} = \mathrm{E}[X_1] = X_0$.
3. Since Alice never updates her defense, Alice's defense $(D_0^{\mathsf{A}}, D_1^{\mathsf{A}})$ forms a martingale as well, which impies that $\sum_{i=1}^{\ell} p^{(i)} \cdot d_1^{\mathsf{A},(i)} = \mathrm{E}\left[D_1^{\mathsf{A}}\right] = D_0^{\mathsf{A}}$.

Given these observations, applying Jensen's inequality on $f(x, y) := x(1 - x) + (x - y)^2 + \left(x - D_0^{\mathsf{B}}\right)^2$ gives us inequality (i).

This completes the proof of Lemma 2 for $n = 2$. In general, for the case when $n > 2$, the proof is essentially the same as $n = 2$ case and hence we omit it here. Appendix C.1 presents the complete proof.

4.2 Inductive Step

In this section, we prove that for all $d_0 \geqslant 1$, if Theorem 3 holds for defense complexity $d = d_0 - 1$, then it is also correct for $d = d_0$. Together, with the proof of base case, i.e., $d = 0$, we complete the proof of Theorem 3.

Our analysis is based on the index of the round that, *for the first time*, some party updates her defense. Let us call the index of this round m. To prove the inductive step, we shall prove the following stronger statement that clearly implies the inductive step. We prove the following lemma by induction on the index of the first defense round m, where $m = 1$ and $m = 2$ serve as the base cases.

▶ **Lemma 3** (Inductive Step of any $d \geqslant 1$)**.** *For any coin-tossing protocol π with defense complexity $d = d_0$,*

1. *If $m = 1$,*
$$\mathsf{Opt}(\pi) \geqslant \Gamma_{2d_0-1} \cdot (X_0 (1 - X_0)) .$$
2. *If $m \geqslant 2$,*
$$\mathsf{Opt}(\pi) \geqslant \Gamma_{2d_0} \cdot \left(X_0 (1 - X_0) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right) .$$

4.2.1 First defense round: $m = 1$

Let us start with $m = 1$. In this case, we have some $(X_0, n, \mathcal{A}, \mathcal{B})$ protocol π, with defense complexity $d_0 = |\mathcal{A} \cup \mathcal{B}|$ and assume, without loss of generality, Alice sends the first message. $m = 1$ implies that Alice updates her defense in the first round, i.e., $1 \in \mathcal{A}$. Assume that there are ℓ possible first messages that Alice can send, namely $\{1, 2, \ldots, \ell\}$. For all $i \in [\ell]$, the probability of the first message being i is $p^{(i)}$ and conditioned on the first message being i, $X_1 = x_1^{(i)}$ and $D_1^{\mathsf{A}} = d_1^{\mathsf{A},(i)}$ and the rest $(n-1)$ rounds forms a sub-protocol π_i that is a $(x_1^{(i)}, n-1, \mathcal{A}', \mathcal{B}')$ protocol where $\mathcal{A}'$ and $\mathcal{B}'$ are obtained respectively by reducing each index inside $\mathcal{A} \backslash \{1\}$ and $\mathcal{B}$ by 1. Clearly, the defense complexity of π_i is $|\mathcal{A}' \cup \mathcal{B}'| = d_0 - 1$. By our induction hypothesis (that Theorem 3 is true for $d = d_0 - 1$), there exists a stopping time of this sub-protocol that yields a score of at least

$$\Gamma_{2(d_0-1)} \cdot x_1^{(i)} \left(1 - x_1^{(i)}\right) .$$

On the other hand, if we stop when message i happens as the first message, the score will increase by

$$\left| x_1^{(i)} - d_1^{\mathsf{A},(i)} \right| + \left| x_1^{(i)} - D_0^{\mathsf{B}} \right| .$$

Again, note that, regardless of Alice's messages, the expectation of Bob's defense shall remain the same and equals to D_0^{B}. The optimal stopping time will decide on whether to stop at first message being i, by comparing which one yields a higher score. Therefore, it will contribute to our score by at least

$$\max \left(\Gamma_{2(d_0-1)} \cdot x_1^{(i)} \left(1 - x_1^{(i)}\right) , \left| x_1^{(i)} - d_1^{\mathsf{A},(i)} \right| + \left| x_1^{(i)} - D_0^{\mathsf{B}} \right| \right) .$$

By invoking Lemma 1 with $P = \Gamma_{2(d_0-1)}$ and $Q = \Gamma_{2d_0-1}$, we get that, for any i

$$\begin{aligned} &\max \left(\Gamma_{2(d_0-1)} \cdot x_1^{(i)} \left(1 - x_1^{(i)}\right) , \left| x_1^{(i)} - d_1^{\mathsf{A},(i)} \right| + \left| x_1^{(i)} - D_0^{\mathsf{B}} \right| \right) \\ &\geqslant \Gamma_{2d_0-1} \cdot \left(x_1^{(i)} \left(1 - x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2 \right) . \end{aligned}$$

Hence, the score corresponding to optimal stopping time will be at least

$$\begin{aligned}
&\sum_{i=1}^{\ell} p^{(i)} \cdot \max\left(\Gamma_{2(d_0-1)} \cdot x_1^{(i)}\left(1-x_1^{(i)}\right), \left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|\right) \\
\geqslant\ & \Gamma_{2d_0-1} \cdot \sum_{i=1}^{\ell} p^{(i)} \cdot \left(x_1^{(i)}\left(1-x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2\right) \\
\overset{\text{(ii)}}{\geqslant}\ & \Gamma_{2d_0-1} \cdot \left(X_0\left(1-X_0\right) + \left(X_0 - \mathrm{E}\left[D_1^{\mathsf{A}}\right]\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2\right) \\
\geqslant\ & \Gamma_{2d_0-1} \cdot X_0\left(1-X_0\right).
\end{aligned}$$

Similar to the previous cases, inequality (ii) is also a consequence of Jensen's inequality. However, we emphasize a crucial point, which is that, since Alice updates her defense in the first round, in general, $(D_0^{\mathsf{A}}, D_1^{\mathsf{A}})$ *need not be* a martingale and so $\mathrm{E}\left[D_1^{\mathsf{A}}\right]$ does not necessarily equal to D_0^{A}.

4.2.2 First defense round: $m = 2$

Next, we consider the case $m = 2$. Let π be a $(X_0, n, \mathcal{A}, \mathcal{B})$ protocol. Without loss of generality, assume Alice sends the first message and Bob sends the second message. $m = 2$ implies that Alice does not update her defense in the first round, while Bob does update his defense in the second round, i.e. $1 \notin \mathcal{A}$ and $2 \in \mathcal{B}$. Again, assume that there are ℓ different messages that Alice can send as the first message, namely $\{1, 2, \ldots, \ell\}$. For all $i \in [\ell]$, the probability of first message being i is $p^{(i)}$ and conditioned on first message being i, $X_1 = x_1^{(i)}$ and $D_1^{\mathsf{A}} = d_1^{\mathsf{A},(i)}$. Furthermore, conditioned on the first message being i, the rest $(n-1)$ rounds forms a $(x_1^{(i)}, n-1, \mathcal{A}', \mathcal{B}')$ sub-protocol π_i. Here, $\mathcal{A}'$ is obtained by reducing each index inside $\mathcal{A}$ by 1. Similarly, $\mathcal{B}'$ is obtained by reducing each index inside $\mathcal{B}$ by 1. Clearly, π_i has the same defense complexity as π, which is d_0. Plus, it falls into the category $m = 1$, since Bob speaks first now and he does update his defense in the first round, i.e., $1 \in \mathcal{B}'$. By our analysis in the $m = 1$ case, there exists a stopping time for π_i that guarantees a score of at least

$$\Gamma_{2d_0-1} \cdot x_1^{(i)}\left(1-x_1^{(i)}\right).$$

On the other hand, if we stop when message i happens, the score will increase by

$$\left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|.$$

Again, we note that, regardless of Alice's message, the expectation of Bob's defense remains the same and equals D_0^{B}. Therefore, the optimal stopping time will decide on whether to stop at first message being i depending on which quantity is larger, i.e.,

$$\max\left(\Gamma_{2d_0-1} \cdot x_1^{(i)}\left(1-x_1^{(i)}\right), \left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|\right).$$

By invoking Lemma 1 with $P = \Gamma_{2d_0-1}$ and $Q = \Gamma_{2d_0}$, we get

$$\begin{aligned}
&\max\left(\Gamma_{2d_0-1} \cdot x_1^{(i)}\left(1-x_1^{(i)}\right), \left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|\right) \\
\geqslant\ & \Gamma_{2d_0} \cdot \left(x_1^{(i)}\left(1-x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2\right).
\end{aligned}$$

This will yield a total score of at least

$$\begin{aligned}
&\sum_{i=1}^{\ell} p^{(i)} \cdot \max\left(\Gamma_{2d_0-1} \cdot x_1^{(i)}\left(1-x_1^{(i)}\right), \left|x_1^{(i)} - d_1^{\mathsf{A},(i)}\right| + \left|x_1^{(i)} - D_0^{\mathsf{B}}\right|\right) \\
\geqslant\ & \Gamma_{2d_0} \cdot \sum_{i=1}^{\ell} p^{(i)} \cdot \left(x_1^{(i)}\left(1-x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2\right) \\
\overset{\text{(iii)}}{\geqslant}\ & \Gamma_{2d_0} \cdot \left(X_0\left(1-X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2\right).
\end{aligned}$$

Here, inequality (iii) is again the consequence Jensen's inequality. And, in comparison to the analysis when $m = 1$, here, since Alice does not update her defense in the first round, $(D_0^{\mathsf{A}}, D_1^{\mathsf{A}})$ *indeed forms* a martingale.

This proves that Lemma 3 holds for $m = 2$. In general, for the case when $m > 2$, the proof is essentially the same as the case $m = 2$, and hence we omit it here. Appendix C.2 presents the complete proof.

5 Generalization to Protocols with Randomized Rounds for Updating Defense

In this section, we present a proof overview of how one can generalize our proof strategies to protocols with randomized defense rounds.

In an n-round coin-tossing protocol with d-randomized defense rounds, each party will decide on whether to update their defenses based on the transcript so far. The upper bound d ensures that, for any full execution of the protocol, i.e., $M_1 = m_1^*, M_2 = m_2^*, \ldots, M_n = m_n^*$, the total number of defense updates from both parties is bounded by d.

We use $i_1, i_2, \ldots, i_d$ to represent the $1^{st}, 2^{nd}, \ldots, d^{th}$ round, in which parties update their defenses. Unlike fixed defense round case, $i_1, \ldots, i_d$ are random variables depending on the transcript of the protocol. Moreover, for all $j \in [d]$, whether $i_j \leqslant k$ is $(M_1, \ldots, M_{k-1})$-measurable.

▶ **Remark 7.** If during a full execution of the protocol, i.e., $M_1 = m_1^*, M_2 = m_2^*, \ldots, M_n = m_n^*$, parties update their defenses $d^* (< d)$ times, without loss of generality, we can simply pick any $d - d^*$ rounds where parties do not update their defense and consider them to be the rounds where parties do update their defense. Therefore, $i_1, \ldots, i_d$ are always well-defined.

For a bias-X_0 coin-tossing protocol with d-randomized defense rounds, we shall prove the same results as the fixed defense round case. That is, either Alice or Bob has a fail-stop attack strategy that deviates the protocol by

$$\frac{1}{4} \cdot \Gamma_{2d} \cdot X_0(1 - X_0).$$

We devote the rest of this section to prove this result. Since the proof is essentially identical to the fixed defense case, we shall present only a proof overview in this submission.

In the same manner, the proof will show a lower bound on the score of the optimal stopping time. Translating this score into a fail-stop attack strategy is identical to the fixed defense round case (see Remark 5). The proof on the lower bound will again use mathematical induction on the defense complexity d.

Firstly, the base case is when $d = 0$, i.e., both parties only prepare defenses before the beginning of the protocol and never update them. In this case, there is no difference between randomized defense rounds and fixed defense rounds. Hence, the proof will be identical.

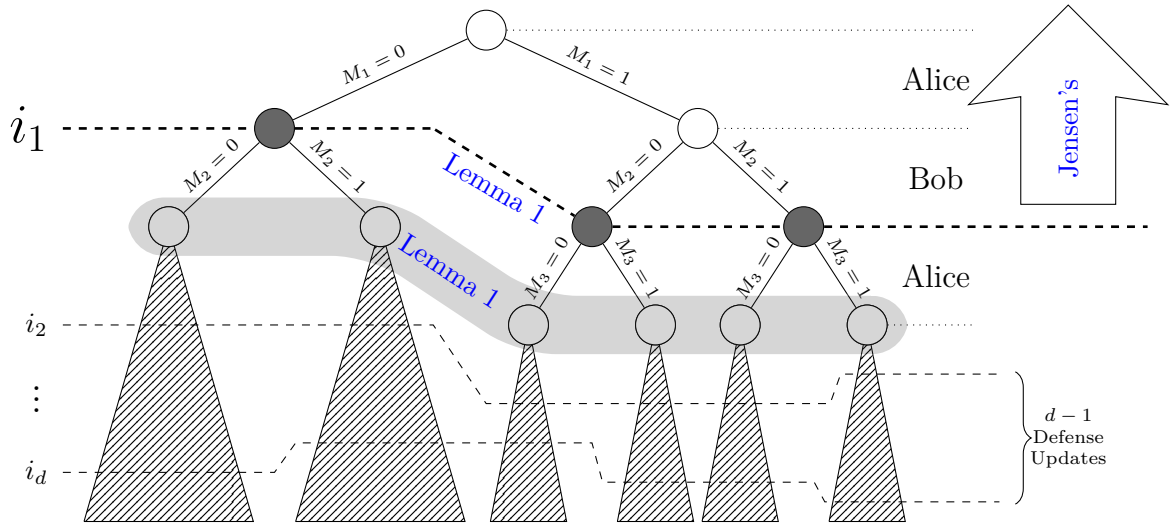

Figure 3 A representative example of a protocol with randomized rounds for updating defense coins. Black nodes represent the first time party updates their defense. For instance, when Alice's first message $M_1 = 0$, Bob will update his defense in round 2. Our proof proceeds by first applying Lemma 1 on the nodes at round $i_1 + 1$ and then again applying Lemma 1 on the nodes at round i_1. Finally, one can "lift" the lower bound on each node at round i_1 all the way to the root of the tree using Jensen's inequality.

Secondly, for the inductive step, let us use Figure 3 as a representative example. The proof shall proceed in the following steps.

1. Consider the subtree rooted at round $i_1 + 1$, i.e., the shaded subtree in Figure 3. By our definition, this subtree will be a sub-protocol with $(d-1)$-randomized defense rounds. Hence, by our induction hypothesis, there exists a stopping time that yields a score of at least $\Gamma_{2d-2} \cdot X(1-X)$, where X is the color at the root, i.e., the node at round $i_1 + 1$.
2. Secondly, consider whether we pick the root of this subtree, i.e., the node at round $i_1 + 1$ as our stopping time, or we continue on this node. Similar to the proof in fixed defense rounds, by invoking Lemma 1 and applying Jensen's inequality, one can prove that for each subtree rooted at nodes at round i_1, i.e., the black node in Figure 3, there exists a stopping time that yields a score of at least $\Gamma_{2d-1} \cdot X(1-X)$.
3. Next, we consider whether we pick the node at round i_1 as our stopping time, or we continue to the subtree rooted at this node. By invoking Lemma 1, one can show that, for each node at round i_1, either we stop at this node or we pick a stopping time for the subtree rooted at this node, this will yield a score of at least

$$\Gamma_{2d} \cdot \left(X(1-X) + (X - d^{\mathsf{A}})^2 + (X - d^{\mathsf{B}})^2 \right).$$

Here, X, d^{A} and d^{B} are the expected outcome, expected Alice's defense and expected Bob's defense, respectively, at this node.

The crucial point is that at these nodes (at round i_1), *no party* has updated their defense yet.[8] Therefore, d^{A} (resp., d^{B}) is the expectation of the defense Alice (resp., Bob) prepares

[8] Recall our score function. By picking a node as stopping time, our score function considers two types of attack. Let m^* be the last message of the path from the root to this node. Either the party who

before the beginning of the protocol conditioned on the transcript so far, i.e., the path from the root to the node at round i_1.

4. Finally, one can repetitively use Jensen's inequality to "lift" this lower bound to the root of the tree and show that the optimal stopping time yields a score of at least

$$\Gamma_{2d} \cdot \left(X_0(1-X_0) + (X_0 - D_0^{\mathsf{A}})^2) + (X_0 - D_0^{\mathsf{B}})^2\right).$$

This can be done because (i) Since no party update their defenses, the expectation of Alice's and Bob's defenses form a martingale; (ii) for every message exposure filtration, information of at most one party's defense will be revealed; (iii) the convexity of our lower bound, that is, function $f(x,y) := x(1-x) + (x-y)^2 + (x-c)^2$ is convex for any constant c.
(Take Figure 3 as an example. One shall first apply Jensen's inequality at the node in round 1 with $M_1 = 1$. And then apply Jensen's inequality at the root of the tree.)

This completes the proof overview.

prepares m^* aborts without sending m^* or the party who receives m^* aborts immediately after receiving m^*. For a node at round i_1, when those two attacks happen, no party has updated their defenses yet.

References

ABMO15 Gilad Asharov, Amos Beimel, Nikolaos Makriyannis, and Eran Omri. Complete characterization of fairness in secure two-party computation of Boolean functions. In Yevgeniy Dodis and Jesper Buus Nielsen, editors, *TCC 2015: 12th Theory of Cryptography Conference, Part I*, volume 9014 of *Lecture Notes in Computer Science*, pages 199–228, Warsaw, Poland, March 23–25, 2015. Springer, Heidelberg, Germany. doi:10.1007/978-3-662-46494-6_10. 2

ALR13 Gilad Asharov, Yehuda Lindell, and Tal Rabin. A full characterization of functions that imply fair coin tossing and ramifications to fairness. In Amit Sahai, editor, *TCC 2013: 10th Theory of Cryptography Conference*, volume 7785 of *Lecture Notes in Computer Science*, pages 243–262, Tokyo, Japan, March 3–6, 2013. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-36594-2_14. 2

Ash14 Gilad Asharov. Towards characterizing complete fairness in secure two-party computation. In Yehuda Lindell, editor, *TCC 2014: 11th Theory of Cryptography Conference*, volume 8349 of *Lecture Notes in Computer Science*, pages 291–316, San Diego, CA, USA, February 24–26, 2014. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-54242-8_13. 2

BGP00 Mihir Bellare, Oded Goldreich, and Erez Petrank. Uniform generation of np-witnesses using an np-oracle. *Inf. Comput.*, 163(2):510–526, 2000. 2

BHMO18 Amos Beimel, Iftach Haitner, Nikolaos Makriyannis, and Eran Omri. Tighter bounds on multi-party coin flipping via augmented weak martingales and differentially private sampling. In Mikkel Thorup, editor, *59th Annual Symposium on Foundations of Computer Science*, pages 838–849, Paris, France, October 7–9, 2018. IEEE Computer Society Press. doi:10.1109/FOCS.2018.00084. 2, 3, 4, 6, 7

BLOO11 Amos Beimel, Yehuda Lindell, Eran Omri, and Ilan Orlov. $1/p$-Secure multiparty computation without honest majority and the best of both worlds. In Phillip Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *Lecture Notes in Computer Science*, pages 277–296, Santa Barbara, CA, USA, August 14–18, 2011. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-22792-9_16. 2

CI93 Richard Cleve and Russell Impagliazzo. Martingales, collective coin flipping and discrete control processes (extended abstract). 1993. 2, 3, 4, 6, 7

Cle86 Richard Cleve. Limits on the security of coin flips when half the processors are faulty (extended abstract). In *18th Annual ACM Symposium on Theory of Computing*, pages 364–369, Berkeley, CA, USA, May 28–30, 1986. ACM Press. doi:10.1145/12130.12168. 2, 3

DLMM11 Dana Dachman-Soled, Yehuda Lindell, Mohammad Mahmoody, and Tal Malkin. On the black-box complexity of optimally-fair coin tossing. In Yuval Ishai, editor, *TCC 2011: 8th Theory of Cryptography Conference*, volume 6597 of *Lecture Notes in Computer Science*, pages 450–467, Providence, RI, USA, March 28–30, 2011. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-19571-6_27. 5, 6

DMM14 Dana Dachman-Soled, Mohammad Mahmoody, and Tal Malkin. Can optimally-fair coin tossing be based on one-way functions? In Yehuda Lindell, editor, *TCC 2014: 11th Theory of Cryptography Conference*, volume 8349 of *Lecture Notes in Computer Science*, pages 217–239, San Diego, CA, USA, February 24–26, 2014. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-54242-8_10. 5, 6

GHKL08 S. Dov Gordon, Carmit Hazay, Jonathan Katz, and Yehuda Lindell. Complete fairness in secure two-party computation. In Richard E. Ladner and Cynthia Dwork, editors, *40th Annual ACM Symposium on Theory of Computing*, pages 413–422, Victoria, BC, Canada, May 17–20, 2008. ACM Press. doi:10.1145/1374376.1374436. 2, 3

GK10 S. Dov Gordon and Jonathan Katz. Partial fairness in secure two-party computation. In Henri Gilbert, editor, *Advances in Cryptology – EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 157–176, French Riviera, May 30 – June 3, 2010. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-13190-5_8. 2

HOZ13 Iftach Haitner, Eran Omri, and Hila Zarosim. Limits on the usefulness of random oracles. In Amit Sahai, editor, *TCC 2013: 10th Theory of Cryptography Conference*, volume 7785 of *Lecture Notes in Computer Science*, pages 437–456, Tokyo, Japan, March 3–6, 2013. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-36594-2_25. 5, 6

JVV86 Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.*, 43:169–188, 1986. URL: https://doi.org/10.1016/0304-3975(86)90174-X, doi:10.1016/0304-3975(86)90174-X. 2

KKR18 Yael Tauman Kalai, Ilan Komargodski, and Ran Raz. A lower bound for adaptively-secure collective coin-flipping protocols. In Ulrich Schmid and Josef Widder, editors, *32nd International Symposium on Distributed Computing, DISC 2018, New Orleans, LA, USA, October 15-19, 2018*, volume 121 of *LIPIcs*, pages 34:1–34:16. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018. URL: https://doi.org/10.4230/LIPIcs.DISC.2018.34, doi:10.4230/LIPIcs.DISC.2018.34. 3, 4

KMM19 Hamidreza Amini Khorasgani, Hemanta K. Maji, and Tamalika Mukherjee. Estimating gaps in martingales and applications to coin-tossing: Constructions and hardness. In Dennis Hofheinz and Alon Rosen, editors, *TCC 2019: 17th Theory of Cryptography Conference, Part II*, volume 11892 of *Lecture Notes in Computer Science*, pages 333–355, Nuremberg, Germany, December 1–5, 2019. Springer, Heidelberg, Germany. doi:10.1007/978-3-030-36033-7_13. 2, 3, 4, 5, 6, 7, 8

Mak14 Nikolaos Makriyannis. On the classification of finite Boolean functions up to fairness. In Michel Abdalla and Roberto De Prisco, editors, *SCN 14: 9th International Conference on Security in Communication Networks*, volume 8642 of *Lecture Notes in Computer Science*, pages 135–154, Amalfi, Italy, September 3–5, 2014. Springer, Heidelberg, Germany. doi:10.1007/978-3-319-10879-7_9. 2

A Some Examples

We use Maj to denote the majority function. In this section we present coin-tossing protocols where the message in the protocol divulge information about Alice's and Bob's defense coins because they prepare the defense coins lazily.

Figure 4 is a $(X_0 = 1/2, n = 3, \mathcal{A} = \emptyset, \mathcal{B} = \emptyset)$-coin-tossing protocol. The defense complexity, i.e., $d = |\mathcal{A} \cup \mathcal{B}|$, is 0.

Alice		Bob
$x_1, u_2, x_3 \xleftarrow{\$} \{0,1\}$		$u_1, x_2, u_3 \xleftarrow{\$} \{0,1\}$
Set $D_0^{\mathsf{A}} = \mathsf{Maj}(x_1, u_2, x_3)$		Set $D_0^{\mathsf{B}} = \mathsf{Maj}(u_1, x_2, u_3)$
$D_1^{\mathsf{A}} = D_0^{\mathsf{A}}$	$M_1 = x_1 \longrightarrow$	$D_1^{\mathsf{B}} = D_0^{\mathsf{B}}$
$D_2^{\mathsf{A}} = D_1^{\mathsf{A}}$	$M_2 = x_2 \longleftarrow$	$D_2^{\mathsf{B}} = D_1^{\mathsf{B}}$
$D_3^{\mathsf{A}} = D_2^{\mathsf{A}}$	$M_3 = x_3 \longrightarrow$	$D_3^{\mathsf{B}} = D_2^{\mathsf{B}}$
Outcome $:= \mathsf{Maj}(x_1, x_2, x_3)$		

Figure 4 A 3-round Majority protocol where both parties never update their defense.

The following Figure 5 is a $(X_0 = 1/2, n = 5, \mathcal{A} = \{3\}, \mathcal{B} = \{2\})$-coin-tossing protocol. The defense complexity, i.e., $d = |\mathcal{A} \cup \mathcal{B}|$, is 2.

Alice		Bob
$x_1, u_2, x_3, u_4, x_5 \xleftarrow{\$} \{0,1\}$		$u_1, x_2, u_3, x_4, u_5 \xleftarrow{\$} \{0,1\}$
Set $D_0^{\mathsf{A}} = \mathsf{Maj}(x_1, u_2, x_3, u_4, x_5)$		Set $D_0^{\mathsf{B}} = \mathsf{Maj}(u_1, x_2, u_3, x_4, u_5)$
$D_1^{\mathsf{A}} = D_0^{\mathsf{A}}$	$M_1 = x_1 \longrightarrow$	$D_1^{\mathsf{B}} = D_0^{\mathsf{B}}$
$D_2^{\mathsf{A}} = D_1^{\mathsf{A}}$	$M_2 = x_2 \longleftarrow$	Update $D_2^{\mathsf{B}} = \mathsf{Maj}(x_1, x_2, u_3, x_4, u_5)$
Update $D_3^{\mathsf{A}} = \mathsf{Maj}(x_1, x_2, x_3, u_4, x_5)$	$M_3 = x_3 \longrightarrow$	$D_3^{\mathsf{B}} = D_2^{\mathsf{B}}$
$D_4^{\mathsf{A}} = D_3^{\mathsf{A}}$	$M_4 = x_4 \longleftarrow$	$D_4^{\mathsf{B}} = D_3^{\mathsf{B}}$
$D_5^{\mathsf{A}} = D_4^{\mathsf{A}}$	$M_5 = x_5 \longrightarrow$	$D_5^{\mathsf{B}} = D_4^{\mathsf{B}}$
Outcome $:= \mathsf{Maj}(x_1, x_2, x_3, x_4, x_5)$		

Figure 5 A 5-round Majority protocol where Alice updates her defense at round 3 and Bob updates his defense at round 2.

B Proof of Lemma 1

In this section, we prove Lemma 1, which states the following.

▶ Lemma (Restatement of Lemma 1). For all $P \in [0,1]$ and $Q \in [0,1/2]$, if P, Q satisfies

$$P - Q - P^2 Q \geqslant 0,$$

then for all $x, \alpha, \beta \in [0,1]$, we have

$$\max\left(P \cdot x(1-x)\ ,\ |x-\alpha| + |x-\beta|\right) \geqslant Q \cdot \left(x(1-x) + (x-\alpha)^2 + (x-\beta)^2\right).$$

In particular, for any $n \geqslant 1$, the constraints are satisfied, if we set $P = \Gamma_{n-1} = \frac{1}{\sqrt{(\sqrt{2}+1)(n+1)}}$ and $Q = \Gamma_n = \frac{1}{\sqrt{(\sqrt{2}+1)(n+2)}}$.

Proof. We first note that it suffices to show that

$$\max\left(P \cdot x(1-x), |x-\alpha|\right) \geqslant Q \cdot x(1-x) + 2Q \cdot (x-\alpha)^2. \tag{1}$$

If this is correct, then we also have

$$\max\left(P \cdot x(1-x), |x-\beta|\right) \geqslant Q \cdot x(1-x) + 2Q \cdot (x-\beta)^2.$$

Together, they imply that

$$\begin{aligned}
&\max\left(P \cdot x(1-x), |x-\alpha| + |x-\beta|\right) \\
\geqslant\ &\frac{1}{2}\Big(\max\left(P \cdot x(1-x), |x-\alpha|\right) + \max\left(P \cdot x(1-x), |x-\beta|\right)\Big) \\
\geqslant\ &Q \cdot \left(x(1-x) + (x-\alpha)^2 + (x-\beta)^2\right).
\end{aligned}$$

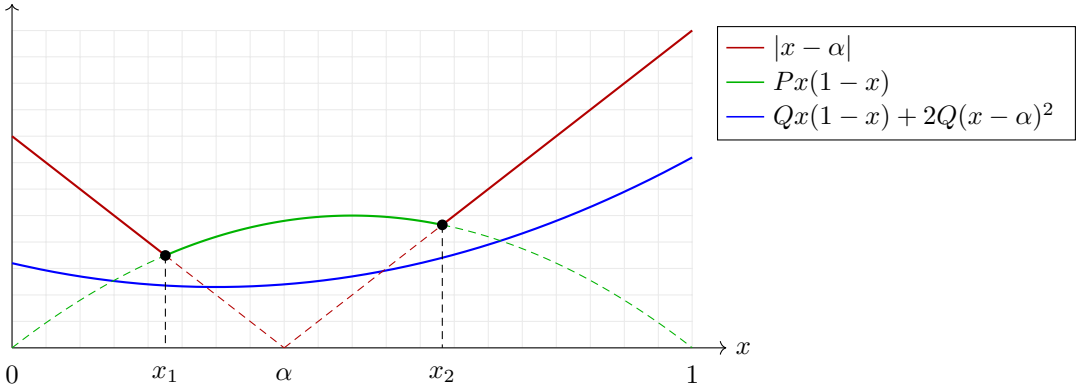

Figure 6 Pictorial summary of Equation 1.

To show Equation 1, let x_1 be the x-coordinate of the left intersection point of

$$\begin{cases} y = Px(1-x) \\ y = \alpha - x \end{cases}$$

and x_2 be the x-coordinate of the right intersection point of

$$\begin{cases} y = Px(1-x) \\ y = x - \alpha \end{cases}$$

Note that, for Equation 1, the RHS is convex on the entire domain, i.e., $[0,1]$. And LHS is piece-wise concave on $[0, x_1]$, $[x_1, x_2]$ and $[x_2, 1]$ respectively. Therefore, to prove Equation 1, it suffices to verify it at $x = 0, x_1, x_2$ and 1. It is trivial to verify it for $x = 0$ and $x = 1$ since $2Q \leqslant 1$ and $|x - \alpha| \leqslant 1$. Furthermore, because of symmetry along $x = 1/2$ axis, it suffices to show this inequality just for $x = x_1$ and all $\alpha \in [0,1]$.[9]

Specifically, we get

$$x_1 = \frac{P + 1 - \sqrt{(P+1)^2 - 4P\alpha}}{2P}.$$

And this inequality is equivalent to, for all $\alpha \in [0,1]$,

$$Px_1(1-x_1) \geqslant Qx_1(1-x_1) + 2Q(x_1 - \alpha)^2,$$

which is equivalent to

$$(P-Q)\left((P+1) - \sqrt{(P+1)^2 - 4P\alpha}\right)\left((P-1) + \sqrt{(P+1)^2 - 4P\alpha}\right) - 2Q\left((P+1) - 2P\alpha - \sqrt{(P+1)^2 - 4P\alpha}\right)^2 \geqslant 0$$

Define

$$\gamma := \sqrt{(P+1)^2 - 4P\alpha},$$

which means

$$\alpha = \frac{(P+1)^2 - \gamma^2}{4P}.$$

And since $\alpha \in [0,1]$, we have $\gamma \in [1-P, 1+P]$. Now, we can simplify the above inequality as, for all $\gamma \in [1-P, 1+P]$,

$$h(\gamma) := (P-Q)(P+1-\gamma)(P-1+\gamma) - 2Q\left(P + 1 - \gamma - \frac{(P+1)^2 - \gamma^2}{2}\right)^2 \geqslant 0 \qquad (2)$$

Note that

$$h(1-P) = h(1+P) = 0,$$

and hence, to prove Equation 2, it suffices to show that $h''(\gamma) \leqslant 0$ on $[1-P, 1+P]$. We get that

$$\begin{aligned} h''(\gamma) &= -2(P-Q) - 2Q\left(1 \cdot \left(P + 1 - \gamma - \frac{(P+1)^2 - \gamma^2}{2}\right) + 2(\gamma - 1)^2 + 1 \cdot \left(P + 1 - \gamma - \frac{(P+1)^2 - \gamma^2}{2}\right)\right) \\ &= -2(P-Q) - 2Q\left(-P^2 + 3(\gamma-1)^2\right) \end{aligned}$$

Hence, for all $\gamma \in [1-P, 1+P]$,

$$h''(\gamma) \leqslant h''(1) = -2(P - Q - P^2Q) \leqslant 0.$$

This completes the proof. ◀

[9] If we verify the inequality for $x = x_1$ when we set $\alpha = c$, this would imply the correctness of the inequality for $x = x_2$ when we set $\alpha = 1 - c$.

C Missing Proofs

C.1 Base Case

In this section, we complete the proof of Lemma 2. We have already shown the lemma is correct for $n = 1$ and $n = 2$. Here, we show how one can inductively prove that, for all $n \geqslant 2$, and any n-round protocol π with defense complexity $d = 0$,

$$\mathsf{Opt}(\pi) \geqslant \frac{1}{2} \cdot \left(X_0 \left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right).$$

We only need to show the inductive step.

Now, suppose the statement is correct for $n = n_0 - 1$ and consider an arbitrary n_0-round protocol π. Without loss of generality, assume Alice sends the first message and there are ℓ possible first messages, namely $\{1, 2, \ldots, \ell\}$. For all $i \in [\ell]$, the probability of the first message being i is $p^{(i)}$ and conditioned on the first message being i, $X_1 = x_1^{(i)}$ and $D_1^{\mathsf{A}} = d_1^{\mathsf{A},(i)}$. Again, regardless of Alice's message, Bob's defense D_1^{B} remains the same and is equal to D_0^{B}. Note that by conditioning on i occurs as the first message, the remaining protocol forms a $(n_0 - 1)$-round protocol π_i with root-color $x_1^{(i)}$. And Alice's and Bob's defense prepared before the beginning of this sub-protocol are Alice's and Bob's defense prepared for the first round in the original protocol, which are $d_1^{\mathsf{A},(i)}$ and D_0^{B} respectively. Using our induction hypothesis, for all $i \in [\ell]$, there exists a stopping times τ_i for this sub-protocol π_i, such that

$$\mathsf{Score}\left(\pi_i, \tau_i\right) \geqslant \frac{1}{2} \cdot \left(x_1^{(i)} \left(1 - x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2 \right).$$

Now, by picking our stopping time τ as the combination of $\tau_1, \tau_2, \ldots, \tau_\ell$, we have

$$\begin{aligned}
\mathsf{Score}\left(\pi, \tau\right) &= \sum_{i=1}^{\ell} p^{(i)} \cdot \mathsf{Score}\left(\pi_i, \tau_i\right) \\
&\geqslant \frac{1}{2} \cdot \sum_{i=1}^{\ell} p^{(i)} \cdot \left(x_1^{(i)} \left(1 - x_1^{(i)}\right) + \left(x_1^{(i)} - d_1^{\mathsf{A},(i)}\right)^2 + \left(x_1^{(i)} - D_0^{\mathsf{B}}\right)^2 \right) \\
&\geqslant \frac{1}{2} \cdot \left(X_0 \left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right).
\end{aligned}$$

The last inequality follows from the same reasoning as inequality (i), i.e., by applying Jensen's inequality on function $x(1-x) + (x-y)^2 + (x-c)^2$ and using the fact that $(D_0^{\mathsf{A}}, D_1^{\mathsf{A}})$ and (X_0, X_1) both are martingales and thus $D_0^{\mathsf{A}} = \mathrm{E}\left[D_1^{\mathsf{A}}\right]$ and $X_0 = \mathrm{E}[X_1]$.

This completes the proof of Lemma 2.

C.2 Inductive Step

In this section, we complete the proof of Lemma 3. We have already shown the lemma is correct for $m = 1$ and $m = 2$. Here, we show how one can inductively prove that, for all $m \geqslant 2$, let π be an $(X_0, n, \mathcal{A}, \mathcal{B})$ protocol that has defense complexity $d_0 = |\mathcal{A} \cup \mathcal{B}|$ and the very first defense update happens at round m. Then

$$\mathsf{Opt}(\pi) \geqslant \Gamma_{2d_0} \cdot \left(X_0 \left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right).$$

We only need to show the inductive step.

Assume that the statement is correct for $m = m_0 - 1$ and let us consider the case $m = m_0$. Let π be an $(X_0, n, \mathcal{A}, \mathcal{B})$ protocol that has defense complexity $d_0 = |\mathcal{A} \cup \mathcal{B}|$ and the very

first defense update happens at round m_0. Without loss of generality, assume Alice sends the first message that has ℓ possibilities, namely $\{1, 2, \ldots, \ell\}$. For all $i \in [\ell]$, the probability of the first message being i is $p^{(i)}$. And conditioned on the first message being i, $X_1 = x_1^{(i)}$ and $D_1^{\mathsf{A}} = d_1^{\mathsf{A},(i)}$. Furthermore, conditioned on the first message being i, we are left with a sub-protocol π_i that has defense complexity d_0 and the first defense update happens at round $m_0 - 1$. Note that Alice's and Bob's defense prepared before the beginning of this sub-protocol are exactly equal to their defense in the first round of the original protocol, that is $d_1^{\mathsf{A},(i)}$ and D_0^{B}. Using our induction hypothesis, we know there exists a stopping time τ_i such that

$$\mathsf{Score}\left(\pi_i, \tau_i\right) \geqslant \Gamma_{2d_0} \cdot \left(X_0\left(1 - X_0\right) + \left(X_0 - d_1^{\mathsf{A},(i)}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right).$$

Now, we pick our stopping time of protocol π as the combination of all the stopping times τ_i of sub-protocol π_i. This would yield a score of at least

$$\begin{aligned}
\mathsf{Score}(\pi, \tau) &= \sum_{i=1}^{\ell} p^{(i)} \cdot \mathsf{Score}\left(\pi_i, \tau_i\right) \\
&\geqslant \Gamma_{2d_0} \cdot \sum_{i=1}^{\ell} p^{(i)} \cdot \left(X_0\left(1 - X_0\right) + \left(X_0 - d_1^{\mathsf{A},(i)}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right) \\
&\geqslant \Gamma_{2d_0} \cdot \left(X_0\left(1 - X_0\right) + \left(X_0 - D_0^{\mathsf{A}}\right)^2 + \left(X_0 - D_0^{\mathsf{B}}\right)^2 \right).
\end{aligned}$$

Again, we apply Jensen's inequality and use the fact that, since Alice does not update her defense in the first round, $(D_0^{\mathsf{A}}, D_1^{\mathsf{A}})$ is a martingale.

This completes the proof of Lemma 3.