Adaptively Secure Coin-Flipping, Revisited

Shafi Goldwasser[1], Yael Tauman Kalai[2], and Sunoo Park[3]

[1]MIT and the Weizmann Institute of Science
[2]Microsoft Research
[3]MIT

Abstract

The full-information model was introduced by Ben-Or and Linial in 1985 to study collective coin-flipping: the problem of generating a common bounded-bias bit in a network of n players with $t = t(n)$ faults. They showed that the majority protocol, in which each player sends a random bit and the output is the majority of the players' bits, can tolerate $t(n) = O(\sqrt{n})$ even in the presence of *adaptive* corruptions, and they conjectured that this is optimal for such adversaries. Lichtenstein, Linial, and Saks proved that the conjecture holds for protocols in which each player sends only a single bit. Their result has been the main progress on the conjecture during the last 30 years.

In this work we revisit this question and ask: what about protocols where players can send longer messages? Can increased communication allow for a larger fraction of corrupt players?

We introduce a model of *strong adaptive* corruptions, in which an adversary sees all messages sent by honest parties in any given round and, based on the message content, decides whether to corrupt a party (and alter its message or sabotage its delivery) or not. This is in contrast to the (classical) adaptive adversary who can corrupt parties only based on past messages, and cannot alter messages already sent.

We prove that any one-round coin-flipping protocol, *regardless of message length*, can be secure against at most $\widetilde{O}(\sqrt{n})$ strong adaptive corruptions. Thus, increased message length does not help in this setting.

We then shed light on the connection between adaptive and strongly adaptive adversaries, by proving that for any symmetric one-round coin-flipping protocol secure against t adaptive corruptions, there is a symmetric one-round coin-flipping protocol secure against t strongly adaptive corruptions. Going back to the standard adaptive model, we can now prove that any symmetric one-round protocol with arbitrarily long messages can tolerate at most $\widetilde{O}(\sqrt{n})$ adaptive corruptions.

At the heart of our results there is a new technique for converting any one-round secure protocol with arbitrarily long messages into a secure one where each player sends only $\mathsf{polylog}(n)$ bits. This technique may be of independent interest.

1 Introduction

A collective coin-flipping protocol is one where a set of n players use private randomness to generate a common random bit b. Several protocol models have been studied in the literature. In this work, we focus on the model of *full information* [BL85] where all parties communicate via a single broadcast channel.

The challenge is that $t = t(n)$ of the parties may be corrupted and aim to bias the protocol outcome (i.e. the "coin") in a particular direction. We focus on *Byzantine faults*, where once a party is corrupted, the adversary completely controls the party and can send any message on its behalf. Two types of Byzantine adversaries have been considered in the literature: *static* adversaries and *adaptive* adversaries. A static adversary is one that chooses which t players to corrupt *before the protocol begins*. An adaptive adversary is one who may choose which t players to corrupt adaptively, as the protocol progresses.

Collective coin-flipping in the case of static adversaries is well understood (see section 1.2). In this work, our focus is on the setting of adaptive adversaries, which has received considerably less attention. We say that a collective coin-flipping protocol is secure against t adaptive (resp. static) corruptions if for any adaptive adversary corrupting t parties, there is a constant $\varepsilon > 0$ such that the probability that the protocol outputs 0 (and the probability that the protocol outputs 1) is at least ε, where the probability is taken over the randomness of the players and the adversary.

The question we study is: *What is the maximum number of adaptive corruptions that a secure coin-flipping protocol can tolerate?* On the positive side, it has been shown by Ben-Or and Linial [BL85] in 1985 that the majority protocol (where each party sends a random bit, and the output is equal to the majority of the bits sent), is resilient to $\Theta(\sqrt{n})$ adaptive corruptions. Ben-Or and Linial conjectured that this is in fact optimal.

Conjecture 1.1 ([BL85])**.** *Majority is the optimal coin-flipping protocol against adaptive adversaries. In particular, any coin-flipping protocol is resilient to at most $O(\sqrt{n})$ adaptive corruptions.*

Shortly thereafter, Lichtenstein, Linial, and Saks [LLS89] proved the conjecture for a restricted class of protocols: namely, those in which each player sends only a single bit. Their result has been the main progress on the conjecture of [BL85] during the last 30 years.

1.1 Our contribution

We first define a new adversarial model of *strong adaptive* corruptions. Informally, an adversary is strongly adaptive if he can corrupt players depending on the content of their messages. More precisely, in each round, he can see all the messages that honest players "would" send, and then decide which of them to corrupt. This is in contrast to a (traditionally defined) adaptive adversary who can, at any point in the protocol, corrupt any player who has not yet spoken based on the history of communication, but cannot alter the message of a player who has already spoken. Thus, strong adaptive adversaries are more powerful than adaptive adversaries.

We believe that the notion of *strong adaptive security* gives rise to a natural and interesting new adversarial model in which to study multi-party protocols in general. Indeed, it is a realistic concern in many settings that malicious parties may decide to stop or alter messages sent by honest players *depending on message content*, and it is a shortcoming that existing adversarial models fail to take such behavior into account.

Our main result is that the conjecture of [BL85] holds (up to polylogarithmic factors) for any one-round coin-flipping protocol in the presence of *strong adaptive* corruptions.

Theorem. Any secure one-round coin-flipping protocol Π can tolerate at most $t = \widetilde{O}(\sqrt{n})$ strong adaptive corruptions.

This is shown by a generic reduction of communication in the protocol: first, we prove that any strongly adaptively secure protocol Π can be converted to one where players send messages of no

more than polylogarithmic length, while preserving the number of corruptions that can be tolerated. Then, we show that any protocol with messages of polylogarithmic length can be converted to one where each player sends only a single bit, at the cost of a polylogarithmic factor in the number of corruptions. Finally, we reach the *single-bit* setting in which the bound of Lichtenstein et al. [LLS89] can be applied to obtain the theorem. We believe that our technique of converting any protocol into one with short messages is of independent interest and will find other applications.

Furthermore, we prove that strongly adaptively secure protocols are a more general class of protocols than symmetric adaptively secure protocols. A symmetric protocol Π is a one that is oblivious to the order of its inputs: that is, where for any permutation $\pi : [n] \to [n]$ of the players, it holds that the protocol outcome $\Pi(r_1, \ldots, r_n) = \Pi(r_{\pi(1)}, \ldots, r_{\pi(n)})$ is the same.

Theorem. For any symmetric one-round coin-flipping protocol Π secure against $t = t(n)$ adaptive corruptions, there is a symmetric one-round coin-flipping protocol Π' secure against $\Omega(t)$ strong adaptive corruptions.

Curiously, this proof makes a novel use of the Minimax Theorem [NM44; Nas50] from game theory, in order to take any symmetric, adaptively secure protocol and convert it to a new protocol which is strongly adaptively secure. This technique views the protocol as a zero-sum game between two players $\mathcal{A}_0$ and $\mathcal{A}_1$, where $\mathcal{A}_0$ wins if the protocol outcome is 0 and $\mathcal{A}_1$ wins if the outcome is 1. We analyze the "minimax strategy" in which the players try to minimize their maximum loss, in order to deduce the strong adaptive security of the new protocol. To our knowledge, this is the first use of these game-theoretic concepts in the construction of distributed protocols.

Finally, using the above results as stepping stones, we return to the classical conjecture of [BL85], in the model of adaptive adversaries, and show that the conjecture holds (up to polylogarithmic factors) for any *symmetric* one-round protocol with arbitrarily long messages.

Theorem. Any secure symmetric one-round coin-flipping protocol Π can tolerate at most $t = \widetilde{O}(\sqrt{n})$ adaptive corruptions.

1.2 Related work

The full-information model (also known as the *perfect information model*) was introduced by Ben-Or and Linial [BL85] to study the problem of collective coin-flipping when no secret communication is possible between honest players.

In the static setting. Protocols for collective coin-flipping in the presence of static corruptions have been constructed in a series of works that variously focus on improving the fault-tolerance, round complexity, and/or bias of the output bit. Feige [Fei99] gave a protocol that is $(\delta^{1.65}/2)$-secure in the presence of $t = (1 + \delta) \cdot n/2$ static corruptions for any constant $0 < \delta < 1$. Russell, Saks, and Zuckerman [RSZ02] then showed that any protocol that is secure in the presence of linearly many corruptions must either have at least $(1/2 - o(1)) \cdot \log^*(n)$ rounds, or communicate many bits per round.

Interestingly, nearly all proposed *multi-round* protocols for collective coin-flipping first run a *leader election* protocol in which one of the n players is selected as a "leader", who then outputs a bit that is taken as the protocol outcome. We remark that this approach is inherently unsuitable for adaptive adversaries, which can always corrupt the leader after he is elected, and thereby surely control the protocol outcome.

In the adaptive setting. The study of coin-flipping protocols has been predominantly in the static setting. The problem of adaptively secure coin-flipping was introduced by Ben-Or and Linial [BL85] and further examined by Lichtenstein, Linial, and Saks [LLS89] as described in the previous section. In addition, Dodis [Dod00] proved that through "black-box" reductions from *non-adaptive* coin-flipping, it is not possible tolerate significantly more corruptions than the majority protocol. The definition of "black-box" used in [Dod00] is rather restricted: it only considers sequential composition of non-adaptive coin-flipping protocols, followed by a (non-interactive) function computation on the coin-flips thus obtained.

In the computational setting. The problem of generating a shared random bit has also been studied in the setting where players are computationally bounded, and in different communication network models. Blum [Blu81] introduced the coin-flipping problem in the two-player computational setting; and Goldreich, Micali, and Wigderson [GMW87] subsequently showed that it is possible to efficiently generate a shared bit with negligible bias, in the presence of static adversaries.

Another line of work shows that the existence of any coin-flipping protocol for computationally bounded players which achieves a sufficiently small bias implies the existence of one-way functions. The latest result in this line of work, due to Berman, Haitner, and Tentes [BHT14], proves that if there exists a two-player coin-flipping protocol that achieves any constant bias, then one-way functions exist.

2 Preliminaries

We consider coin-flipping protocols in the *full-information model* (also known as the *perfect information model*), where n computationally unbounded players communicate via a single broadcast channel. The network is synchronized between rounds, but is asynchronized within each round (that is, there is no guarantee on message ordering within a round, and an adversary can see the messages of all honest players in a round before deciding his own messages).

In this work, we focus on *one-round* protocols, and we consider protocols that terminate (and produce an output) with probability 1. In particular, we focus on *coin-flipping* protocols, which are defined as follows.

Definition 2.1 (Coin-flipping protocol)**.** *A* coin-flipping protocol $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ *is a family of protocols where each* Π_n *is a* n*-player protocol which outputs a bit in* $\{0,1\}$*.*

Notation. We write $\overset{s}{\approx}$ for statistical indistinguishability of distributions. We denote by $\Pr^{\Pi}(b)$ the probability that an honest execution of Π will lead to the outcome $b \in \{0,1\}$. We denote by $\Pr^{\Pi,\mathcal{A}}(b)$ the probability that an execution of Π in the presence of an adversary $\mathcal{A}$ will lead to the outcome $b \in \{0,1\}$. The probability is over the random coins of the honest players and the adversary.

For one-round protocols, we write $\Pi_n(r_1,\ldots,r_n)$ to denote the outcome of the protocol Π_n when each player i sends message r_i. (The vector $(r_1,\ldots,r_n)$ is a protocol *transcript*.)

2.1 Properties of protocols

Definition 2.2 (Symmetric protocol)**.** *A protocol* Π *is* symmetric *if the outcome of a protocol execution is the same no matter how the messages within each round are permuted. In particular,*

a one-round protocol Π *is symmetric if for all* $n \in \mathbb{N}$ *and any permutation* $\pi \in [n] \to [n]$,

$$\Pi_n(r_1, \ldots, r_n) = \Pi_n(r_{\pi(1)}, \ldots, r_{\pi(n)}).$$

We remark, for completeness, that in the multi-round case, the outcome of a symmetric protocol should be unchanged even if different permutations are applied in different rounds.

Definition 2.3 (Single-bit/multi-bit protocol)**.** *A protocol is* single-bit *if each player sends at most one bit over the course of the protocol execution. Similarly, a protocol is* m-bit *if each player sends at most* m *bits over the course of the protocol execution. More generally, a protocol which is not single-bit is called* multi-bit*.*

Definition 2.4 (Public-coin protocol)**.** *A protocol is* public-coin *if all honest players' messages consist only of random bits.*

2.2 Adversarial models in the literature

The type of adversary that has been by far the most extensively studied in the coin-flipping literature is the static adversary, which chooses a subset of players to corrupt *before* the protocol execution begins, and controls the behavior of the corrupt players arbitrarily throughout the protocol execution.

A stronger type of adversary is the *adaptive* adversary, which may choose players to corrupt at any point during protocol execution, and controls the behavior of the corrupt players arbitrarily from the moment of corruption until protocol termination.

Definition 2.5 (Adaptive adversary)**.** *Within each round, the adversary chooses players one-by-one to send their messages; and he can perform corruptions at any point during this process.*

2.3 Security of coin-flipping protocols

The security of a coin-flipping protocol is usually measured by the extent to which an adversary can, by corrupting a subset of parties, bias the protocol outcome towards his desired bit.

Definition 2.6 (ε-security)**.** *A coin-flipping protocol* Π *is* ε-secure *against* $t = t(n)$ *adaptive (or static or strong adaptive) corruptions if for all* $n \in \mathbb{N}$*, it holds that for any adaptive (resp. static or strong adaptive) adversary* $\mathcal{A}$ *that corrupts at most* $t = t(n)$ *players,*

$$\min\left(\Pr^{\Pi_n,\mathcal{A}}(0), \Pr^{\Pi_n,\mathcal{A}}(1)\right) \geq \varepsilon.$$

We remark that this definition of ε-security is sometimes referred to as *ε-control* or *ε-resilience* in other works. We next define a *secure* protocol to be one with "minimal" security properties (that is, one where the adversary does not almost always get the outcome he wants).

Definition 2.7 (Security)**.** *A coin-flipping protocol is* secure *against* $t = t(n)$ *corruptions if it is* ε*-secure against* t *corruptions for some constant* $0 < \varepsilon < 1$*.*

In this work, we investigate the maximum proportion of adaptive corruptions that can be tolerated by *any* secure protocol.

3 Our results

3.1 Strongly adaptive adversaries

In this work, we propose a new, stronger adversarial model than those that have been studied thus far (see section 2.2), in which the adversary can see all honest players' messages within any given round, and *subsequently* decide which players to corrupt. That is, he can see all the messages that the honest players "would have sent" in a round, and then selectively intercept and alter these messages.

Definition 3.1 (Strong adaptive adversary)**.** *Within each round, the adversary sees all the messages that honest players would have sent, then gets to choose which (if any) of those messages to corrupt (i.e. replace with messages of his choice).*

This notion is an essential tool underlying the proof techniques in our work. Moreover, we believe that the notion of *strong adaptive security* gives rise to a natural and interesting new adversarial model in which to study multi-party protocols, which is of independent interest beyond the scope of this work.

3.2 Corruption tolerance in secure coin-flipping protocols

Our main contributions consist of the following three results. These can be viewed as partial progress towards proving the 30-year-old conjecture of [BL85].

Theorem 3.2. *Any one-round coin-flipping protocol* Π *can be secure against at most* $t = \widetilde{O}(\sqrt{n})$ *strong adaptive corruptions.*

Theorem 3.3. *For any symmetric one-round coin-flipping protocol* Π *secure against* $t = t(n)$ *adaptive corruptions, there is a symmetric one-round coin-flipping protocol* Π' *secure against* $\Omega(t)$ *strong adaptive corruptions.*

Corollary 3.4. *Any symmetric one-round coin-flipping protocol* Π *can be secure against at most* $t = \widetilde{O}(\sqrt{n})$ *adaptive corruptions.*

Proof. Suppose, for contradiction, that there exists a symmetric one-round coin-flipping protocol Π which is secure against more than $\widetilde{O}(\sqrt{n})$ adaptive corruptions. Then, by Theorem 3.3, there exists a one-round coin-flipping protocol Π' which is secure against more than $\widetilde{O}(\sqrt{n})$ *strong* adaptive corruptions. This contradicts Theorem 3.2. □

In the next sections, we proceed to give detailed proofs of the theorems.

3.3 Proof of Theorem 3.2

We begin by recalling the result of Lichtenstein et al. [LLS89] which proves that the maximum number of adaptive corruptions for any secure *single-bit* coin-flipping protocol is $O(\sqrt{n})$. Note that the *majority protocol* is the one-round protocol in which each player broadcasts a random bit, and the majority of broadcasted bits is taken to be the protocol outcome.

Theorem 3.5 ([LLS89])**.** *Any coin-flipping protocol in which each player broadcasts at most one bit can be secure against at most* $t = O(\sqrt{n})$ *corruptions. Moreover, the majority protocol achieves this bound.*

Next, we establish some definitions and supporting lemmas.

Definition 3.6 (Distance between message-vectors)**.** *For vectors $\vec{r}, \vec{r'} \in \mathcal{M}^n$, let $\mathsf{dist}(\vec{r}, \vec{r'})$ be equal to the number of coordinates $i \in [n]$ for which $r_i \neq r'_i$.*

Definition 3.7 (Robust sets)**.** *Let Π be a one-round coin-flipping protocol in which each player sends a message from a message space $\mathcal{M}$. For any $n \in \mathbb{N}$ and $b \in \{0,1\}$, define the set $\mathsf{Robust}^{\Pi_n}(b,t)$ as follows:*

$$\mathsf{Robust}^{\Pi_n}(b,t) = \left\{ \vec{r} \in \mathcal{M}^n \ : \ \forall \vec{r'} \in \mathcal{M}^n \ s.t. \ \mathsf{dist}(\vec{r}, \vec{r'}) \leq t, \ \Pi_n(\vec{r}) = \Pi_n(\vec{r'}) = b \right\}.$$

Lemma 3.8. *Let Π be a one-round coin-flipping protocol in which each player sends a random message from a message space $\mathcal{M}$. Π is secure against $t = t(n)$ strong adaptive corruptions if and only if there exists a constant $0 < \varepsilon < 1$ such that for all $n \in \mathbb{N}$ and each $b \in \{0,1\}$,*

$$\Pr_{\vec{r} \leftarrow \mathcal{M}} \left[\vec{r} \in \mathsf{Robust}^{\Pi_n}(b,t) \right] \geq \varepsilon.$$

Proof. ("IF") Suppose that there exists a constant $0 < \varepsilon < 1$ such that for all $n \in \mathbb{N}$ and all $b \in \{0,1\}$, it holds that

$$\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\vec{r} \in \mathsf{Robust}^{\Pi_n}(b,t) \right] \geq \varepsilon. \tag{1}$$

Let $\mathcal{A}$ be any strong adaptive adversary making up to t corruptions. For n-vector of (honest) messages $\vec{r} \in \mathcal{M}^n$, let $\mathcal{A}(\vec{r}) \in \mathcal{M}^n$ denote the corresponding corrupted message-vector, where up to t of the messages have been modified by $\mathcal{A}$. By the definition of the set $\mathsf{Robust}^{\Pi_n}(b,t)$, it holds that

$$\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\Pi_n(\mathcal{A}(\vec{r})) = b \mid \vec{r} \in \mathsf{Robust}^{\Pi_n}(b,t) \right] = 1. \tag{2}$$

Combining equations (1) and (2), it follows that for each outcome $b \in \{0,1\}$,

$$\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\Pi_n(\mathcal{A}(\vec{r})) = b \right] \geq \varepsilon.$$

We have shown that for each $b \in \{0,1\}$, $\Pr^{\Pi,\mathcal{A}}(b) \geq \varepsilon$, as required.

("ONLY IF") Suppose, on the other hand, that there is no constant $0 < \varepsilon < 1$ such that for all $b \in \{0,1\}$, it holds that $\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\vec{r} \in \mathsf{Robust}^{\Pi_n}(b,t) \right] = \varepsilon$. That is, there exists some $\varepsilon' = o(1)$ such that for some $b \in \{0,1\}$ and infinitely many values of $n \in \mathbb{N}$, it holds that

$$\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\vec{r} \in \mathsf{Robust}^{\Pi_n}(b,t) \right] \leq \varepsilon'. \tag{3}$$

Without loss of generality, let $b = 0$ be the bit for which equation (3) holds. By the definition of $\mathsf{Robust}^{\Pi_n}(b,t)$, it holds that for any $\vec{r} \notin \mathsf{Robust}^{\Pi_n}(b,t)$, there exists a vector $\vec{r}_{bad} \in \mathcal{M}^n$ such that $\mathsf{dist}(\vec{r}, \vec{r}_{bad}) \leq t$ and $\Pi_n(\vec{r}) \neq \Pi_n(\vec{r}_{bad})$. In other words, if the honest players' messages $\vec{r}$ do not fall in $\mathsf{Robust}^{\Pi_n}(0,t)$, then it is possible for a strong adaptive adversary $\mathcal{A}$ to *force* the outcome to be 1, by doing as follows:

$$\mathcal{A}(\vec{r}) = \begin{cases} \vec{r} & \text{if } \Pi_n(\vec{r}) = 1 \\ \vec{r}_{bad} & \text{if } \Pi_n(\vec{r}) = 0 \end{cases}$$

Note that since $\mathsf{dist}(\vec{r}, \vec{r}_{bad}) \leq t$, it is always possible for the adversary to change from $\vec{r}$ to $\vec{r}_{bad}$ using t or fewer corruptions. Moreover, if $\Pi_n(\vec{r}) = 0$, then it must be that $\Pi_n(\vec{r}_{bad}) = 1$, by construction of $\vec{r}_{bad}$. Hence,

$$\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\Pi_n(\mathcal{A}(\vec{r})) = 1 \mid \vec{r} \notin \mathsf{Robust}^{\Pi_n}(0,t) \right] = 1. \tag{4}$$

Combining equations (3) and (4) (for $b = 0$), we obtain:

$$\Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\Pi_n(\mathcal{A}(\vec{r})) = 1 \right] = \Pr_{\vec{r} \leftarrow \mathcal{M}^n} \left[\vec{r} \notin \mathsf{Robust}^{\Pi_n}(0,t) \right] \geq 1 - \varepsilon'.$$

Hence, $\Pr^{\Pi,\mathcal{A}}(1) \geq 1 - \varepsilon'$, and so $\Pr^{\Pi,\mathcal{A}}(0) \leq \varepsilon' = o(1)$. Therefore, Π is not secure against t strong adaptive corruptions. The lemma follows. □

Since players are computationally unbounded and we consider one-round protocols, we may without loss of generality consider public-coin protocols[1]: for any one-round protocol Π in the full-information model, there is a protocol Π' with an identical output distribution (in the presence of any adversary), in which honest players send random messages in $\{0,1\}^k$ for some $k = \mathsf{poly}(n)$.

The following lemma serves as a stepping-stone to our final theorem.

Lemma 3.9. *For any one-round multi-bit coin-flipping protocol Π secure against $t = t(n)$ strong adaptive corruptions, and any constant $\delta > 0$, there is a one-round ℓ-bit coin-flipping protocol Π' that is secure against t strong adaptive corruptions, where $\ell = O(\log^{1+\delta}(n))$.*

Proof. Without loss of generality, we consider only public-coin protocols, and assume that each player sends a message of the same length (say, $k = k(n)$ bits). Let $\delta > 0$ be any constant, let $\ell = O(\log^{1+\delta}(n))$, and let $\ell' = 2^\ell$.

For an $\ell' \times n$ matrix of messages $M \in (\{0,1\}^k)^{\ell' \times n}$, we define the protocol Π^M as follows: each player P_i broadcasts a random integer $a_i \leftarrow [\ell']$, and the protocol outcome is defined by

$$\Pi_n^M(a_1, \ldots, a_n) = \Pi_n(M_{(a_1,1)}, \ldots, M_{(a_n,n)}),$$

where $M_{(i,j)}$ denotes the message at the i^{th} row and j^{th} column of the matrix M. For notational convenience, define $\vec{M}(a_1, \ldots, a_n) = (M_{(a_1,1)}, \ldots, M_{(a_n,n)})$. Notice that by construction of the protocol Π^M, it holds that for any message-vector $\vec{a} \in [\ell']^n$,

$$\vec{M}(\vec{a}) \in \mathsf{Robust}^{\Pi_n}(b,t) \implies \vec{a} \in \mathsf{Robust}^{\Pi_n^M}(b,t). \tag{5}$$

Suppose each entry of the matrix M is a uniformly random message in $\{0,1\}^k$. Note that the length of each player's message in Π^M is $\log(\ell') = \ell$. We want to show that Π^M is a secure coin-flipping protocol against t strong adaptive corruptions, for some M. By Lemma 3.8, it is sufficient to show that there exists $M \in (\{0,1\}^k)^{\ell' \times n}$ such that for all $b \in \{0,1\}$,

$$\Pr_{\vec{a} \leftarrow [\ell']^n} \left[\vec{a} \in \mathsf{Robust}^{\Pi_n^M}(b,t) \right] \geq \varepsilon, \tag{6}$$

[1] This is without loss of generality: each player can simply send his random coin tosses, and security holds since we are in the full-information model.

where $0 < \varepsilon < 1$ is constant. Using implication (5), it actually suffices to prove:

$$\exists M \in (\{0,1\}^k)^{\ell' \times n} \text{ s.t. } \forall b \in \{0,1\}, \quad \Pr_{\vec{a} \leftarrow [\ell']^n} \left[\vec{M}(\vec{a}) \in \mathsf{Robust}^{\Pi_n}(b,t) \right] \geq \varepsilon. \tag{7}$$

Suppose the matrix M is chosen uniformly at random. Let $q = \mathsf{poly}(n)$ and let $\vec{a}_1, \ldots \vec{a}_q$ be sampled independently and uniformly from $[\ell']^n$. Since, the number of matrix rows $\ell' = 2^{O(\log^{1+\delta}(n))}$ is super-polynomial, whereas $q = \mathsf{poly}(n)$, it is overwhelmingly likely that $\vec{a}_1, \ldots \vec{a}_q$ will be composed of distinct elements in $[\ell']$. That is, to be precise,

$$\Pr_{\vec{a}_1,\ldots,\vec{a}_q} \left[\forall (i,j) \neq (i',j') \in [q] \times [n],\ (\vec{a}_i)_j \neq (\vec{a}_{i'})_{j'} \right] \geq 1 - \mathsf{negl}(n).$$

If $\vec{a}_1, \ldots, \vec{a}_q$ are indeed composed of distinct elements, the message-vectors $\vec{M}(\vec{a}_1), \ldots, \vec{M}(\vec{a}_q)$ are independent random elements in $(\{0,1\}^k)^n$. Thus,

$$(\vec{M}(\vec{a}_1), \ldots, \vec{M}(\vec{a}_q)) \overset{s}{\approx} (\vec{r}_1, \ldots, \vec{r}_q), \tag{8}$$

when M is a random matrix in $(\{0,1\}^k)^{\ell' \times n}$, the (short) message-vectors $\vec{a}_1, \ldots, \vec{a}_q$ are random in $[\ell']^n$, and the (long) message-vectors $\vec{r}_1, \ldots, \vec{r}_q$ are random in $(\{0,1\}^k)^n$.

Since Π is a secure coin-flipping protocol, there is a constant $0 < \varepsilon' < 1$ such that for all $n \in \mathbb{N}$ and $b \in \{0,1\}$ and $i \in [q]$,

$$\Pr_{\vec{r}_i} \left[\vec{r}_i \in \mathsf{Robust}^{\Pi_n}(b,t) \right] \geq \varepsilon'.$$

The rest of the proof follows from a series of Chernoff bounds.

For $i \in [q]$ and $b \in \{0,1\}$, let $Z_{i,b}$ be an indicator variable for the event that $\vec{r}_i \in \mathsf{Robust}^{\Pi_n}(b,t)$. Since the $\vec{r}_i$ are independent, we apply a Chernoff bound to obtain the following (for all $b \in \{0,1\}$):

$$\Pr_{\vec{r}_1,\ldots,\vec{r}_q} \left[\frac{1}{q} \cdot \sum_{i \in [q]} Z_{i,b} < \varepsilon' - \varepsilon'' \right] \leq \mathsf{negl}(n), \tag{9}$$

for any constant $0 < \varepsilon'' < \varepsilon'$.

Let $Y_{i,b}$ be an indicator variable for the event that $\vec{M}(\vec{a}_i) \in \mathsf{Robust}^{\Pi_n}(b,t)$. It follows from (8) and (9) that with overwhelming probability over the choice of the random matrix M, it holds for all $b \in \{0,1\}$ that

$$\Pr_{\vec{a}_1,\ldots,\vec{a}_q} \left[\frac{1}{q} \cdot \sum_{i \in [q]} Y_{i,b} < \varepsilon' - \varepsilon'' \right] \leq \mathsf{negl}(n). \tag{10}$$

For $b \in \{0,1\}$, let α_b denote the probability $\Pr_{\vec{a}_i} \left[\vec{M}(\vec{a}_i) \in \mathsf{Robust}^{\Pi_n}(b,t) \right]$. Note that for any given $b \in \{0,1\}$ the variables $Y_{i,b}$ are independently and identically distributed, each taking value 1 with probability α_b and value 0 with probability $1 - \alpha_b$. By a Chernoff bound, for any constant $0 < \varepsilon''' < 1$, it holds that (with overwhelming probability over the choice of M):

$$\Pr_{\vec{a}_1,\ldots,\vec{a}_q} \left[\left| \frac{1}{q} \cdot \sum_{i \in [q]} Y_{i,b} - \alpha_b \right| \geq \varepsilon''' \right] \leq \mathsf{negl}(n). \tag{11}$$

From (10) and (11), it follows that with overwhelming probability over the random choice of M, for all $b \in \{0,1\}$ and any constant $0 < \varepsilon'' < 1$ and $0 < \varepsilon''' < 1$,

$$\Pr_{\vec{a}_1,\ldots,\vec{a}_q}\left[\alpha_b < \varepsilon' - \varepsilon'' - \varepsilon'''\right] \leq \mathsf{negl}(n).$$

By taking $\varepsilon'' + \varepsilon''' \leq \varepsilon'/2$, we have that with overwhelming probability over M, it holds that $\alpha_b < \varepsilon'/2$ for all $b \in \{0,1\}$. Finally, the α_b correspond exactly to the probability expression in (7), so we have shown statement (7) as required. □

Having reduced the length of players' messages to $\mathsf{polylog}(n)$ in Lemma 3.9, we now prove the following lemma which reduces the required communication even further, so that each player sends only one bit. This comes at the cost of a polylogarithmic factor reduction in the number of corruptions.

Before the lemma, we recall the statement of the Chernoff bound.

Theorem 3.10 (Chernoff bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$, which all have the same expectation $\mu = \mathbb{E}[X_i]$. Then, for every $0 < \varepsilon < 1$,*

$$\Pr\left[\left|\frac{1}{n} \cdot \sum_{i\in[n]} X_i - \mu\right| \geq \varepsilon\right] \leq 2e^{-2n\varepsilon^2}.$$

Lemma 3.11. *For any one-round ℓ-bit coin-flipping protocol Π secure against $t = t(n)$ strong adaptive corruptions, there is a one-round single-bit coin-flipping protocol Π' that is secure against t/ℓ strong adaptive corruptions.*

Proof. Let Π be any one-round ℓ-bit coin-flipping protocol secure against $t = t(n)$ strong adaptive corruptions. We define our new single-bit protocol[2] Π' as follows, for each $n \in \mathbb{N}$:

$$\Pi'_{n\cdot\ell}(r_1, \ldots, r_{n\cdot\ell}) = \\ \Pi_n\left((r_1||\ldots||r_\ell), (r_{\ell+1}||\ldots||r_{2\ell}), \ldots, (r_{(n-1)\cdot\ell+1}||\ldots||r_{n\cdot\ell})\right),$$

where the messages $r_i \in \{0,1\}$ are bits and $||$ denotes concatenation. Informally speaking, there are n groups of ℓ players in the single-bit protocol $\Pi'_{n\cdot\ell}$, each of which "corresponds to" a single player in the protocol Π_n.

We show that Π' is secure against t/ℓ corruptions. Let G_i denote the i^{th} group of ℓ players: to be precise, $G_i = \{i\cdot\ell + 1, \ldots, (i+1)\cdot\ell\}$. If all of the players in the set G_i are honest, then the i^{th} "combined message" $(r_{i\cdot\ell+1}||\ldots||r_{(i+1)\cdot\ell})$ is distributed identically to an honest message of the i^{th} player in the protocol Π_n. By the construction of the protocol Π', it follows that for any $b \in \{0,1\}$ and $n \in \mathbb{N}$,

$$\Pr_{\vec{r}\leftarrow\{0,1\}^{n\cdot\ell}}\left[\vec{r} \in \mathsf{Robust}^{\Pi'_{n\cdot\ell}}(b, t(n))\right] \geq \Pr_{\vec{r'}\leftarrow(\{0,1\}^\ell)^n}\left[\vec{r'} \in \mathsf{Robust}^{\Pi_n}(b, t(n))\right]. \tag{12}$$

[2]We remark that the protocol Π' that we construct does not strictly adhere to Definition 2.1, because $\Pi' = \{\Pi_n\}_{n\in\ell\cdot\mathbb{N}}$ does not define an n-player protocol for every $n \in \mathbb{N}$. We consider this to be a very minor technical detail that we bury for clarity of exposition.

By Lemma 3.8, since Π is secure against t strong adaptive corruptions, there is a constant $0 < \varepsilon < 1$ such that for all $b \in \{0,1\}$ and $n \in \mathbb{N}$, the right-hand side of inequality (12) is at least ε. Hence we obtain

$$\Pr_{\vec{r} \leftarrow \{0,1\}^{n \cdot \ell}} \left[\vec{r} \in \mathsf{Robust}^{\Pi'_{n \cdot \ell}}(b, t(n)) \right] \geq \varepsilon.$$

It follows (by applying Lemma 3.8 again) that Π' is secure against t/ℓ strong adaptive corruptions. □

Finally, we bring together Lemmas 3.9 and 3.11 to prove the theorem.

Theorem 3.2. Any one-round coin-flipping protocol Π can be secure against at most $t = \widetilde{O}(\sqrt{n})$ strong adaptive corruptions.

Proof. Suppose, for contradiction, that there exists a one-round coin-flipping protocol Π which is secure against t corruptions, where $t = \omega(\sqrt{n} \cdot \mathsf{polylog}(n))$. Then, by Lemma 3.9, there is an ℓ-bit one-round coin-flipping protocol Π' that is secure against t strong adaptive corruptions, where $\ell = \mathsf{polylog}(n)$. By applying Lemma 3.11 to the protocol Π', we deduce that there is a single-bit one-round coin-flipping protocol Π'' which is secure against $t/\ell = \widetilde{\Omega}(t)$ strong adaptive corruptions. Since a *strongly adaptive* adversary can perfectly simulate any strategy of an *adaptive* adversary, it follows that Π'' is secure against $\widetilde{\Omega}(t)$ adaptive corruptions. Since Π'' is single-bit, this contradicts Theorem 3.5. □

3.4 Proof of Theorem 3.3

In this section, we show that for any symmetric one-round coin-flipping protocol secure against t *adaptive* corruptions, there is a one-round coin-flipping protocol secure against $\Omega(t)$ corruptions by *strong adaptive* adversaries. That is, one-round strong adaptively secure protocols are a more general class than one-round symmetric, adaptively secure protocols.

The Minimax Theorem – a classic tool in game theory – will be an important tool in our proof. The statement of the Minimax Theorem and supporting game-theoretic definitions are given below.

Definition 3.12 (Two-player strategic game)**.** *A* two-player finite strategic game $\Gamma = \langle (A_1, A_2), (u_1, u_2) \rangle$ *is defined by: for each player $i \in \{1,2\}$, a non-empty set of possible* actions A_i *and a* utility function $u_i : A_1 \times A_2 \to \mathbb{R}$.

Definition 3.13 (Zero-sum game)**.** *A two-player finite strategic game $\Gamma = \langle (A_1, A_2), (u_1, u_2) \rangle$ is* zero-sum *if for any pair of actions $a_1 \in A_1$ and $a_2 \in A_2$, it holds that $u_1(a_1, a_2) + u_2(a_1, a_2) = 0$.*

Theorem 3.14 (Minimax [NM44; Nas50])**.** *Let $\Gamma = \langle (A_1, A_2), (u_1, u_2) \rangle$ be a zero-sum two-player finite strategic game. Then*

$$\max_{a_2 \in \Delta(A_2)} \min_{a_1 \in \Delta(A_1)} u_2(a_1, a_2) = \min_{a_1 \in \Delta(A_1)} \max_{a_2 \in \Delta(A_2)} u_1(a_1, a_2),$$

where $\Delta(A_i)$ denotes the set of distributions over A_i (in game-theoretic terminology, this corresponds to the set of "mixed strategies" for player i.)

Theorem 3.3. For any symmetric one-round coin-flipping protocol Π secure against $t = t(n)$ adaptive corruptions, there is a symmetric one-round coin-flipping protocol Π' secure against $s = t/2$ strong adaptive corruptions.

Proof. Let Π be a symmetric one-round coin-flipping protocol secure against $t = t(n)$ adaptive corruptions, and define $s(n) = t(n)/2$. We define a new protocol $\Pi' = \{\Pi'_n\}_{n\in\mathbb{N}}$ as follows:

$$\Pi'_n(r_1, \ldots, r_n) = \min_{r'_1,\ldots,r'_s} \max_{r''_1,\ldots,r''_s} \Pi_{n+2s}\left(r_1, \ldots, r_n, r'_1, \ldots, r'_s, r''_1, \ldots, r''_s\right),$$

where $s = s(n)$ and honest players in Π'_n must send messages according to the same distributions as in Π_{n+2s}.

Observe that Π_{n+2s} is secure against $t(n + 2s(n)) > t(n)$ corruptions. We show that Π'_n is secure against $s(n) = t(n)/2$ strong adaptive corruptions.

Case 1. Suppose that the adversary aims to bias the outcome towards 0. By the security of Π_{n+2s}, there is a constant $0 < \varepsilon < 1$ such that $\Pr^{\Pi_{n+2s},\mathcal{A}}(1) \geq \varepsilon$ for any adaptive adversary $\mathcal{A}$ that corrupts up to $t = 2s$ players. Without loss of generality (since the protocol is symmetric), suppose that the adversary corrupts the last $2s$ players in Π_{n+2s}.

We say that the honest players' messages $r_1, \ldots, r_n$ "fix" the outcome of Π_{n+2s} to be 1 if for any possibly malicious messages $\hat{r}_1, \ldots, \hat{r}_{2s}$, it holds that $\Pi_{n+2s}(r_1, \ldots, r_n, \hat{r}_1, \ldots, \hat{r}_{2s}) = 1$. Then, with probability at least ε, the honest players' messages $r_1, \ldots, r_n$ "fix" the outcome of Π_{n+2s} to be 1. (To see this: suppose not. Then there would exist an adversary which could set the corrupt messages $\hat{r}_1, \ldots, \hat{r}_{2s}$ so that the protocol outcome is 0 with probability $1 - \varepsilon$. But this cannot be, since we already established that $\Pr^{\Pi_{n+2s},\mathcal{A}}(1) \geq \varepsilon$.)

Define the set $R_1 \stackrel{\text{def}}{=} \{(r_1, \ldots, r_n) : \forall \hat{r}_1, \ldots, \hat{r}_{2s},\ \ \Pi_{n+2s}(r_1, \ldots, r_n, \hat{r}_1, \ldots, \hat{r}_{2s}) = 1\}$ to consist of those honest message-vectors that fix the output of Π_{n+2s} to be 1.

Take any $(r_1, \ldots, r_n) \in R_1$. We now show that the outcome of Π'_n when the honest players send messages $r_1, \ldots, r_n$ is equal to 1, even in the presence of a strong adaptive adversary $\mathcal{A}'$ that corrupts up to s players and aims to bias the outcome towards 0. Without loss of generality, suppose that $\mathcal{A}'$ corrupts the first s players in Π'_n, and replaces their honest messages $r_1, \ldots, r_s$ with some maliciously chosen messages $\hat{r}_1, \ldots, \hat{r}_s$. In this case, the outcome of Π'_n is

$$\begin{aligned}
&\Pi'_n(\hat{r}_1, \ldots, \hat{r}_s, r_{s+1}, \ldots, r_n) \\
&\quad = \min_{r'_1,\ldots,r'_s} \max_{r''_1,\ldots,r''_s} \Pi_{n+2s}\left(\hat{r}_1, \ldots, \hat{r}_s, r_{s+1}, \ldots, r_n, r'_1, \ldots, r'_s, r''_1, \ldots, r''_s\right) \\
&\quad \geq \min_{r'_1,\ldots,r'_s} \Pi_{n+2s}\left(\hat{r}_1, \ldots, \hat{r}_s, r_{s+1}, \ldots, r_n, r'_1, \ldots, r'_s, r_1, \ldots, r_s\right) \\
&\quad = \min_{r'_1,\ldots,r'_s} \Pi_{n+2s}\left(r_1, \ldots, r_n, \hat{r}_1, \ldots, \hat{r}_s, r'_1, \ldots, r'_s\right) \qquad \text{(by symmetry)} \\
&\quad = 1,
\end{aligned}$$

where the last line follows from the definition of R_1, since we started with $(r_1, \ldots, r_n) \in R_1$.

We already established that the probability that the honest players' messages fall in R_1 is at least ε. Thus we deduce that with probability at least ε, the outcome of the new protocol Π'_n is equal to 1, even in the presence of a strong adaptive adversary corrupting s players and aiming to bias towards 0.

Case 2. Suppose instead that the adversary $\mathcal{A}'$ aims to bias the outcome towards 1. We apply the Minimax Theorem to a zero-sum game where player 1 chooses the messages $r'_1, \ldots, r'_s$ and player 2 chooses the messages $r''_1, \ldots, r''_s$, and player 1 "wins" if the protocol outcome is 0, and player 2 wins otherwise. By the Minimax Theorem,

$$\Pi'_n(r_1, \ldots, r_n) = \max_{r''_1,\ldots,r''_s} \min_{r'_1,\ldots,r'_s} \Pi_{n+2s}\left(r_1, \ldots, r_n, r'_1, \ldots, r'_s, r''_1, \ldots, r''_s\right).$$

Given this new and equivalent definition of Π'_n, we can apply exactly the same argument structure as that given for Case 1 above, to deduce that

- There is a constant $0 < \varepsilon' < 1$ such that $\Pr^{\Pi_{n+2s},\mathcal{A}}(0) = 1 - \Pr^{\Pi_{n+2s},\mathcal{A}}(1) = \varepsilon'$ for any adaptive $\mathcal{A}$ performing up to $2s$ corruptions, and hence there is a non-empty set

$$R_0 \stackrel{\text{def}}{=} \{(r_1, \ldots, r_n) : \forall \hat{r}_1, \ldots, \hat{r}_{2s}, \;\; \Pi_{n+2s}(r_1, \ldots, r_n, \hat{r}_1, \ldots, \hat{r}_{2s}) = 0\}, \text{ and}$$

- by the adaptive security of Π_{n+2s}, the messages of honest players will fall in R_0 with probability at least ε', and
- if the honest players' messages fall in R_0, then the outcome of Π'_n is equal to 0, even in the presence of a strong adaptive adversary corrupting s players and aiming to bias towards 1.

We have established that both outcomes 0 and 1 occur with constant probability in Π'_n, even in the presence of an arbitrary strong adaptive adversary corrupting up to s players. Therefore, Π'_n is secure against $s = t/2$ corruptions. □

4 Conclusion

We have introduced a new adversarial model for multi-party protocols and an associated security notion, *strong adaptive security*. We have made use of a novel and widely applicable technique for reducing the amount of communication in a protocol, to show that any one-round strongly adaptively secure coin-flipping protocol can tolerate at most $\widetilde{O}(\sqrt{n})$ corruptions. We believe that this work paves the way to a number of little-explored research directions. We highlight some interesting questions for future work:

- To study the extent to which *communication can be reduced in protocols in general*, and to extend our communication-reduction techniques to the settings of multi-round protocols and/or adaptive security.
- To apply the *strong adaptive security notion* in the context of other types of protocols and settings, and to design protocols secure in the presence of strong adaptive adversaries.
- To extend this work to prove (or disprove) the long-open conjecture of Lichtenstein et al. [LLS89] that *any* adaptively secure coin-flipping protocol can tolerate at most $O(\sqrt{n})$ corruptions.

References

[BL85] Michael Ben-Or and Nathan Linial. "Collective Coin Flipping, Robust Voting Schemes and Minima of Banzhaf Values". In: *FOCS*. IEEE Computer Society, 1985, pp. 408–416.

[BHT14] Itay Berman, Iftach Haitner, and Aris Tentes. "Coin flipping of *any* constant bias implies one-way functions". In: *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*. Ed. by David B. Shmoys. ACM, 2014, pp. 398–407. ISBN: 978-1-4503-2710-7. DOI: 10.1145/2591796.2591845. URL: http://doi.acm.org/10.1145/2591796.2591845.

[Blu81] Manuel Blum. "Coin Flipping by Telephone". In: *Advances in Cryptology: A Report on CRYPTO 81, CRYPTO 81, IEEE Workshop on Communications Security, Santa Barbara, California, USA, August 24-26, 1981.* Ed. by Allen Gersho. U. C. Santa Barbara, Dept. of Elec. and Computer Eng., ECE Report No 82-04, 1981, pp. 11–15.

[Dod00] Yevgeniy Dodis. "Impossibility of Black-Box Reduction from Non-Adaptively to Adaptively Secure Coin-Flipping". In: *Electronic Colloquium on Computational Complexity (ECCC)* 7.39 (2000).

[Fei99] Uriel Feige. "Noncryptographic Selection Protocols". In: *FOCS*. IEEE Computer Society, 1999, pp. 142–153.

[GMW87] Oded Goldreich, Silvio Micali, and Avi Wigderson. "How to Play any Mental Game or A Completeness Theorem for Protocols with Honest Majority". In: *STOC*. Ed. by Alfred V. Aho. ACM, 1987, pp. 218–229. ISBN: 0-89791-221-7.

[LLS89] David Lichtenstein, Nathan Linial, and Michael E. Saks. "Some extremal problems arising form discrete control processes". In: *Combinatorica* 9.3 (1989), pp. 269–287.

[Nas50] John F. Nash. "Equilibrium points in n-person games". In: *Proceedings of the National Academy of Sciences* 36.1 (1950), pp. 48–49. DOI: 10.1073/pnas.36.1.48. eprint: http://www.pnas.org/content/36/1/48.full.pdf+html. URL: http://www.pnas.org/content/36/1/48.short.

[NM44] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944. ISBN: 0691119937.

[RSZ02] Alexander Russell, Michael E. Saks, and David Zuckerman. "Lower Bounds for Leader Election and Collective Coin-Flipping in the Perfect Information Model". In: *SIAM J. Comput.* 31.6 (2002), pp. 1645–1662. DOI: 10.1137/S0097539700376007. URL: http://dx.doi.org/10.1137/S0097539700376007.