Lower Bounds on Anonymous Whistleblowing

Willy Quach * LaKyah Tyner † Daniel Wichs ‡

Abstract

Anonymous transfer, recently introduced by Agrikola, Couteau and Maier [ACM22] (TCC '22), allows a sender to leak a message anonymously by participating in a public non-anonymous discussion in which everyone knows who said what. This opens up the intriguing possibility of using cryptography to ensure strong anonymity guarantees in a seemingly non-anonymous environment.

The work of [ACM22] presented a lower bound on anonymous transfer, ruling out constructions with strong anonymity guarantees (where the adversary's advantage in identifying the sender is negligible) against arbitrary polynomial-time adversaries. They also provided a (heuristic) upper bound, giving a scheme with weak anonymity guarantees (the adversary's advantage in identifying the sender is inverse in the number of rounds) against *fine-grained* adversaries whose run-time is bounded by some fixed polynomial that exceeds the run-time of the honest users. This leaves a large gap between the lower bound and the upper bound, raising the intriguing possibility that one may be able to achieve weak anonymity against arbitrary polynomial time adversaries, or strong anonymity againstfine grained adversaries.

In this work, we present improved lower bounds on anonymous transfer, that rule out both of the above possibilities:

- We rule out the existence of anonymous transfer with any non-trivial anonymity guarantees against general polynomial time adversaries.
- Even if we restrict ourselves to fine-grained adversaries whose run-time is essentially equivalent to that of the honest parties, we cannot achieve strong anonymity, or even quantitatively improve over the inverse polynomial anonymity guarantees (heuristically) achieved by [ACM22].

Consequently, constructions of anonymous transfer can only provide security against fine-grained adversaries, and even in that case they achieve at most weak quantitative forms of anonymity.

*Weizmann Institute of Science. Email: willy.quach@weizmann.ac.il.
†Northeastern University. Email: tyner.l@northeastern.edu.
‡Northeastern University and NTT Research. Email: wichs@ccs.neu.edu.

Contents

1 Introduction

Consider the following question:

Can a sender leak a message anonymously*, by exclusively participating in a public non-anonymous discussion where everyone sees who said what?*

In particular, we consider a setting where the participants are having a seemingly innocuous discussion (e.g., about favorite cat videos). The discussion is public and non-anonymous, meaning that the participants are using their real identities and everyone knows who said what.[1] The non-sender participants are having a real conversation about this topic. On the other hand, the sender is carefully choosing what to say in a way that looks like she is participating in the conversation, but her real goal is to leak a secret document (e.g., NSA's polynomial-time factoring algorithm). At the end of the discussion, anyone should be able to look at the transcript of the conversation and reconstruct the secret document, without learning anything about which of the participants was the actual sender responsible for leaking it. Despite its conceptual importance and simplicity, this question has not been studied until recently, perhaps because it may appear "obviously impossible".

A formal study of the question was recently initiated by Agrikola, Couteau and Maier in [ACM22], who, perhaps surprisingly, raise the intriguing possibility of answering it positively using cryptography. They do so by introducing a new cryptographic primitive, dubbed anonymous transfer (henceforth AT), to capture the setting above. An anonymous transfer involves a sender with a secret document, along with unaware dummy participants who send uniformly random messages.[2] The parties run for some number of rounds, where in each round the sender and each participant sends a message. At the end of the protocol anyone can reconstruct the secret document with high probability given the transcript. However, the transcript cannot be used to identify who the sender is among the participants.

Crucially, anonymous transfer does not rely on the availability of any (weak) anonymous channels, nor on the availability of trusted third parties during the execution. Instead, all protocol messages are assumed to be traceable to their respective senders, and all other dummy participants only passively send random messages. The simplicity of the setting makes it both a natural question to explore, and raises very intriguing possibility of "creating" anonymity in a seemingly non-anonymous environment.

Anonymous transfer and whistleblowing. One central motivation for studying anonymous transfer is its relation to whistleblowing, where whistleblowers wish to leak confidential and oftentimes sensitive information, while operating in a potentially untrusted environment. The whistleblowers themselves usually risk being subjected to both harsh social, financial, and even legal consequences if caught [ABE14, Inz18, BBC22]. One natural mitigation for those risks is the use appropriate tools, typically cryptographic ones, to ensure anonymity of the leak. And indeed, a large body of work is devoted to build such tools.

One crucial aspect of these tools is the assumptions made on resources available to the whistleblower, which we would ideally like to minimize. From a practical perspective, it seems unreasonable to assume the general availability of, say, anonymous channels or online trusted parties to whistleblowers. In fact, even given the availability of such anonymous channels, their use alone could potentially be incriminating. From a more theoretical perspective, cryptographic solutions leveraging such assumptions could be seen as bootstrapping weaker forms of anonymity. Unfortunately, as far as we are aware, except the work of [ACM22], all prior work on whistleblowing assume the availability of an online form of trust, and thus do not seem to answer the initial question we consider. In contrast, [ACM22] asks the intriguing question of whether cryptography can *create* forms of anonymity in a more fundamental sense.

[1] For concreteness, the public discussion could occur over Facebook or Twitter, and users need to be logged in with their true identity.

[2] This departs from our informal setting, where a real discussion occurred, while we now assume that "real discussions" are uniformly random. Various works, including [HLv02, vH04, vHL05] show how to embed uniform randomness into real discussions. Concretely, it suffices to (randomly) encode uniformly random messages to the distribution representing the (non-necessarily uniform) communication pattern, in a way that the random messages can be decoded.

Prior work on anonymous transfer. Along with introducing anonymous transfer, [ACM22] gives both lower bounds, and, perhaps surprisingly, plausibility results on its feasibility. Let us introduce some notation. The *correctness error* $\varepsilon = \varepsilon(\lambda)$ of an anonymous transfer is the probability secret documents fail to be reconstructed, and the *anonymity* $\delta = \delta(\lambda)$ of an AT is the advantage a transcript of the AT provides towards identifying the sender.[3] An AT is in general interactive, and consists of $c = c(\lambda)$ rounds of interaction.

On the negative side, [ACM22] shows that no protocol can satisfy close to ideal forms of correctness and security, namely $\varepsilon, \delta = \text{negl}(\lambda)$, against all polynomial time adversaries. They supplement this lower bound with a plausibility result, by giving heuristic constructions of anonymous transfer with *fine-grained* security. This heuristic construction provides negligible correctness error, but weaker anonymity guarantees (namely $\delta \approx 1/c$, where c is the number of rounds), and only against a restricted class of *fine-grained* adversaries, who are allowed restricted to be at most $O(c)$ times more powerful than honest users, which are argued secure by relying on ideal obfuscation.

Still, the work of [ACM22] leaves open the possibility of building anonymous transfer with non-optimal correctness and security guarantees (e.g, $\delta \leq 1/c$) secure against arbitrary polynomial-time attacks.

Our results. In this work, we give improved lower bounds for anonymous transfer, largely ruling out potential improvements over the heuristic upper bound from [ACM22]. Throughout this exposition, we will consider the case of 2 participants, one sender and a non-sender; [ACM22] shows that lower bounds in that setting translates to lower bounds for any larger number of parties. Our main theorem shows that anonymous transfer with any non-trivial anonymity against general polynomial-time attackers is impossible, solving a conjecture explicitly stated in [ACM22].

Theorem 1.1 (Informal)**.** *For any 2-party anonymous transfer protocol for $\omega(\log \lambda)$-bit messages with correctness error ε, for all polynomial $\alpha = \alpha(\lambda)$, there exists a polynomial-time adversary that identifies the sender with probability at least $1 - \varepsilon - 1/\alpha$.*

Note that, the probability of identifying the sender is essentially optimal, as, with probability ε, the sender might act as a dummy party, and therefore this rules out any non-trivial constructions.

Our attack runs in polynomial-time, but where the polynomial is fairly large. This unfortunately does not match the run-time of allowed adversaries in the heuristic construction of [ACM22].

As a secondary result, we show that even in the setting of fine-grained adversaries whose run-time is essentially equivalent to that of the honest parties, we can identify senders with probability $1/c$ whenever the secret document can be reconstructed. This shows that, even in the fine-grained setting, one cannot improve on the quantitative anonymity guarantees achieved by the heuristic construction of [ACM22].

Theorem 1.2 (Informal)**.** *For any 2-party anonymous transfer protocol for $\omega(\log \lambda)$-bit messages, with correctness error ε, and having c-round of interaction, there exists a fine-grained adversary whose run-time matches that of the reconstruction procedure up to additive constant factors, that identifies the sender with probability at least $(1 - \varepsilon - \text{negl}(\lambda))/c$.*

Theorem 1.2 in particular rules out all fine-grained protocols with a polynomial number of rounds, if both δ and ε are negligible. For comparison, the lower bound of [ACM22] rules out very similar parameters, but where the run-time of the adversary is $m(\lambda) = \lambda \cdot c^g$ times larger than the one of the reconstruction procedure, for some arbitrary constant $g > 0$.

Related work on whistleblowing. Current solutions for anonymous messaging and anonymous whistleblowing include systems based on onion routing [DMS04], mix-nets [Cha03], and Dining Cryptographer networks or DC-nets [Cha88, CGBM15, APY20, ECZB21, NSSD21]. Additionally, there have been other applications developed that utilize new techniques inspired by the models mentioned previously [CGBM15, Ber16, CCD+20]. Each of these solutions, however, intrinsically assumes that there exists non-colluding

[3] In this work, we use the convention that an AT is *stronger* as ε, δ tend to 0; this is the opposite convention of [ACM22] where this held whenever ε, δ tend to 1.

honest servers that participate to ensure anonymity. [ACM22] is the first to introduce a model which does not rely on this assumption. Impossibility results could be interpreted as evidence that such an assumption *is* in fact necessary.

Open problems. The main open question left by [ACM22] and this work is the construction of fine-grained anonymous transfer matching their heuristic construction, but under standard assumptions.

Additionally, our attack in Theorem 1.1 runs in fairly large polynomial time, which does not tightly match the fine-grained security proved in the heuristic construction of [ACM22]. We leave for future work the possibility of improving the run-time of an attack matching the properties of Theorem 1.1.

2 Technical Overview

Anonymous transfer. Let us first recall some basic syntax and notations for anonymous transfer (henceforth AT), introduced in [ACM22]. In this overview, we will focus on 2-party anonymous transfer, which features a *sender*, a *dummy party* and an external *receiver*.[4][5] The sender takes as input a message μ to transfer. The sender and the dummy party exchange messages in synchronous rounds, with the restriction that the dummy party only sends random bits. An execution of the transfer spans over c rounds of interaction, where both parties send a message at each round. Given a full transcript, the external receiver can (attempt to) reconstruct the original message. We say that an AT has ε correctness error if the reconstruction procedure fails to recover μ with probability at most ε; and that it is δ-anonymous if no adversary has advantage greater than δ in identifying the sender amongst the two participating parties over a random guess, where the adversary can choose the message to be sent.[6] We refer to Section 3.1 for formal definitions.

In that setting, [ACM22] showed the following lower bound on AT.

Theorem 2.1 ([ACM22], paraphrased)**.** *Every (two-party, silent receiver) AT with ε-correctness and δ-anonymity against all polynomial-time adversary, and consisting of c rounds, satisfies $\delta \cdot c \geq \frac{1-\varepsilon}{2} - 1/m(\lambda)$ for all polynomial $m(\lambda)$.*

In particular, no AT can satisfy $\delta, \varepsilon = \mathrm{negl}(\lambda)$ (assuming $c = \mathrm{poly}(\lambda)$, which holds if participants are polynomial-time). More precisely, [ACM22] show, for all polynomial $m(\lambda)$, an attack with runtime $m(\lambda) \cdot \mathrm{poly}(\lambda)$ with advantage at least $\frac{1}{c} \cdot (\frac{1-\varepsilon}{2} - 1/m(\lambda))$.

The main limitation of Theorem 2.1 is that it does not rule out the existence of AT protocols with anonymity δ scaling inverse-polynomially with the number of rounds c, e.g. $\delta = 1/c$. In other words, the trade-off between correctness and security could potentially be improved by relying on a large amount of interaction. And indeed, [ACM22] does provide a plausibility result, where, assuming ideal obfuscation, there exists a *fine-grained* AT with $\delta \approx 1/c$, $\varepsilon = \mathrm{negl}(\lambda)$, so that anonymity does improve with the number of rounds. A secondary limitation is that, because the attack corresponding to Theorem 2.1 needs to call the honest algorithms a polynomial number of times (even though the polynomial can arbitrarily small), this potentially leaves room for "very fine-grained" protocols, where security would only hold against adversaries running in mild super-linear time compared to honest users.

Our main results are stronger generic attacks on anonymous transfer protocols.

A general blueprint for our attacks. The core idea behind all our attacks is a simple notion of *progress* associated to any (potentially partial) transcript of an AT. We do so by associating a real *value* $p \in [0,1]_{\mathbb{R}}$ to partial transcripts of an AT, as follows. We can complete any partial transcripts, replacing all missing messages by uniformly random ones, and attempt to recover the input message $\mu \leftarrow \{0,1\}^{\ell}$ from the sender.

[4] Anonymous transfer can also be defined with more than a single "dummy" party. We focus for simplicity on the 2-party case for this overview, and will show how to extend the attacks to the N-party case subsequently.

[5] We consider here "silent" receivers who do not send any messages — this is similarly known to be sufficient for lower bounds [ACM22].

[6] We remind the reader that [ACM22] takes different conventions than ours for ε and δ. With our notation, an AT satisfies stronger properties as ε and δ get smaller and closer to 0, and are ideally negligible in the security parameter.

For a partial transcript, we define $p \in [0,1]_{\mathbb{R}}$ to be the probability that a random completion of the transcript allows to recover μ.

The next step is to attribute partial evolutions of p, as the transcript gets longer, to parties in the protocol. Namely if, after party A sends the ith message in the transcript, the value of the protocol evolves from p_{i-1} to p_i, and we attribute to A some progress dependent on p_{i-1} and p_i. We then make the following observations: the empty transcript has value $p_0 = 1/2^\ell$ close to 0 (if μ is chosen uniformly at random), and full transcripts have (on expectation) value $p_{2c} = 1 - \varepsilon$ close to 1 by correctness. Our main leverage is that messages sent by the unaware, dummy participant in an AT do not significantly change the value of a partial transcript: this is because, in our random completion of transcripts, messages from the dummy party follow their real distribution. Furthermore, as long as the final value p_{2c} is significantly larger than the initial value p_0, then a significant amount of total progress has to be made *at some point*. Therefore the messages from the sender have to significantly bias the values of partial transcripts towards 1.

This results in the following blueprint for identifying the sender of the AT. We first estimate the contribution of each party towards total progress, namely, the evolution of the values p associated to partial transcripts where the last message was sent from that party.[7] Then, we argue that (1) the contribution of the dummy party is likely to be small overall and (2) the total contribution of both parties is fairly large (on expectation), from which we conclude that the party whose messages contributed the most to increasing the value p is likely to be the AT sender.

Covert cheating games. We abstract out this recipe as a natural game, that we call a covert cheating game. A covert cheating game played by two players A and B, who take $2c$ alternate turns moving a point, or current *state of the game*, on the real interval $[0,1]$. One player is designed to be a *bias inducer*, and the other a *neutral party*. The initial state is p_0, and the final state is p_{2c} is either 0 or 1. We say that a strategy has *success rate* $p_f > p_0$ if $\mathbb{E}[p_{2c}] \geq p_f$, regardless of the identity of the bias inducer. The neutral party is restricted to exclusively making randomized moves that do not affect the current state on expectation. The goal of a third player, the *observer* C, is to determine, given access to the states of the game, which player is the bias inducer. Our main technical contribution is to show generic observer strategies for this game. We refer to Definition 4.1 for a more detailed definition.

We use this abstraction to capture the fact that our attacks use the AT in a specific black-box sense, namely, only to measure out values $p \in [0,1]_{\mathbb{R}}$, and using all the AT algorithms in a black-box manner. Overall, our abstraction of attacks on ATs as strategies in a game captures a natural family of black-box distinguishing algorithms, which we believe capture most reasonable attacks on AT.[8] Indeed, it is not clear how to leverage any non-black-box use of honest user algorithms, as they could potentially be obfuscated (and indeed, the plausibility result of [ACM22] does rely on obfuscated programs to be run by honest users). We believe this game to be natural enough to see other applications in the future.

In the rest of the technical overview, we focus on describing generic attacks in the language of covert cheating games.

A generic "free-lunch" attack. We describe our first attack on the game introduced above, which corresponds to a proof sketch of Theorem 1.2. Our attack is very simple, and only leverages the fact that, on expectation over a random move, moves done by the bias inducer bias the outcome by an additive term $(p_f - p_0)/c$, while moves from the neutral party do not add any bias. Suppose the game consists of c rounds (each consisting of one move from each party A, B), and that party A makes the first move, so that A makes the odd moves $2k+1$, and B makes the even moves $2k$. Our strategy is to pick a random move $k \leftarrow [c]$ from A, whose kth move makes the game evolve from state p_{2k} to p_{2k+1}. We simply output "A is the bias inducer" with probability p_{2k+1} (and B with probability $1 - p_{2k+1}$).

[7] In an AT, rounds are by default synchronous; for the sake of this general blueprint, any arbitrary sequentialization of the messages would be meaningful.

[8] More precisely, strategies are black-box in the AT algorithms, but need to consider full transcripts in a slightly non-black-box way (namely, by separating messages and considering random continuations).

The main idea is that if A is the neutral party, then on expectation $p_{2k+1} = p_{2k}$, and thus our strategy outputs A with probability p_k. On the other hand, if A is the bias inducer, our strategy outputs A with probability p_{2k+1}.[9] Because B is then a neutral party, B's total expected contribution is 0, namely $\mathbb{E}_k[p_{2k+2} - p_{2k+1}] = 0$, so that the advantage of our algorithm towards determining A is:

$$\mathbb{E}_k[p_{2k+1} - p_{2k}] = \mathbb{E}_k[p_{2k+1} - p_{2k} + \underbrace{(p_{2k+2} - p_{2k+1})}_{0}] = (p_f - p_0)/c.$$

The cost of our attack is the cost of obtaining a single sample with probability p_{2k+1}. Going back to AT, this corresponds to the cost of running the honest users' algorithms *once* (namely, attempting to reconstruct the message of a random completion of a randomly chosen partial transcript with last message from A). We conclude no AT can provide security with parameters from Theorem 1.2, in *any* fine-grained setting (as long as adversaries are allowed to be in the same complexity class as honest users).

A generic attack with large advantage. We now describe a slightly more involved attack that achieves stronger advantage, at the cost of running in larger polynomial time. The main inspiration behind this new attack comes from taking a closer look on the restriction that the neutral party's moves do not change the game state on expectation. We observe that this is a more stringent restriction if the current game state p is close to 0. For concreteness, if the current state of the game is $p = 1/2$, then the neutral party could potentially move the state to $p' = 0$ or $p' = 1$ with probability $1/2$ each, inducing a large change of the value of p. However, starting at $p \gtrapprox 0$, Markov's inequality ensures that p' cannot be too large.

This motivates us to consider a different quantification of progress where *additive* progress close to 0 is weighed more significantly than *additive* progress at large constants (e.g. $1/2$). We do so by considering a *multiplicative* form of progress associated to moves and players. Namely, if the ith move of the game transforms the game state from p_{i-1} to p_i, then we define the multiplicative progress associated with the move as

$$r_i = \frac{p_i}{p_{i-1}}.$$[10]

The total progress associated with a player would then be the product of the progress associated with its moves.

Our blueprint still applies in this context. The total progress of all the moves combined is

$$\prod_{i \in [2c]} r_i = \prod_{i \in [2c]} \frac{p_i}{p_{i-1}} = \frac{p_f}{p_0},$$[11]

and so one of the players (on expectation) needs to have progress at least $\sqrt{p_f/p_0}$. Furthermore, one can show that the restriction on neutral party's moves implies that the product of the r_i associated to the neutral party is 1 on expectation. Namely, denoting N the set of indices corresponding to moves made by the neutral party: $\mathbb{E}\left[\prod_N r_i\right] = 1$. Markov's inequality then gives:

$$\Pr\left[\prod_N r_i \geq \sqrt{\frac{p_f}{p_0}}\right] \leq \sqrt{\frac{p_0}{p_f}}.$$

[9] Technically, the quantities p_k when A is the neutral party and p_k when A is the bias inducer are not necessarily related. But without loss of generality, the strategies used by the bias inducer and the neutral party are independent of their identity as A or B, in which case the quantities p_{2k} *are* equal.

[10] One technically needs to be careful handling cases where $p_i = 0$ for some i. We largely ignore this technicality in this overview. For concreteness, it will be enough to output a random guess if this happens, and observe that, for games resulting from an AT, this happens with probability at most $1 - p_f$, and therefore does not affect our advantage too much. We refer to Section 4.3 for more details.

[11] Actually, the total progress is only guaranteed to be p_f/p_0 *on expectation*, which induces several technical issues. We will assume the progress is always equal to p_f/p_0 for the sake of this overview, and we refer to Section 4.2 for more details on the issues and a solution.

Overall, this shows that with good probability $1 - \sqrt{p_0/p_f}$, the sender has a large total contribution, and the dummy party has a small contribution, so that an attacker can identify them with at least such a probability.

We are unfortunately not done yet, because observers do not have direct access to the real values $p \in [0, 1]$: they are only given the ability to sample coins with probability p (going back to AT, recall that this is done by sampling a random completion of a transcript and testing whether the reconstructed message matched the sender's message). This is problematic: from the perspective of a polynomial-time observer, the values $p = \mathrm{negl}(\lambda)$ and $p = 0$ are indistinguishable, given only sampling access. How can we then ensure that the ratios $r_i = p_i/p_{i-1}$ are even well-defined (that is, that $p_{i-1} \neq 0$)?

We solve this issue by conditioning our product to be over moves $i \geq i^*$, such that for all $i > i^*, p_i \geq \tau$ for some small accuracy threshold $p_0 < \tau < p_f$ (think $\tau = 1/\mathrm{poly}(\lambda)$), and where we set the convention $p_{i^*} = \tau$. Now the ratios are well-defined, and the total contribution is now p_f/τ. It remains to argue that the product corresponding to the neutral party is small. While we might have biased the distribution of the neutral party by conditioning on the product starting at i^*, we argue by a union bound that, with sufficiently high probability $1 - c\sqrt{\tau/p_f}$, *all* "suffix-products" from the dummy party are small (namely, smaller than $\sqrt{p_f/\tau}$)

Summing up, our final observer strategy estimates all the p_i up to some sufficiently good precision (using Chernoff) so that the product of the $r_i = p_i/p_{i-1}$ is ensured to be accurate, as long as all the terms p_i that appear in the product are large enough compared to our threshold τ. We refer to Section 4.4 for more formal details.

Taking a step back, the major strength of Theorem 1.1 is that the advantage of the associated attack is independent of the number of rounds: only its running time scales with the number of rounds (in order to ensure sufficient precision with Chernoff bounds). This is in our eyes a quantitative justification that multiplicative progress is better suited to identify bias in a covert cheating game.

Extending the Lower Bound to N parties. Last, we sketch how to extend our attack from Theorem 1.1 to the N-party setting, which consists of a sender interacting with $N - 1$ dummy parties. Our first step is to observe that our attacks described above directly translate to *targeted-predicting attacks*, which correctly identify the sender given the promise that the sender is either party $i \in [N]$ or $j \in [N]$ where $i \neq j$ are arbitrary but fixed for the targeted predictor. This follows from [ACM22], which builds a 2-party AT from any N-party AT, while preserving targeted-predicting security.[12] In other words, given the promise that the sender is either party i or party j, we can correctly identify the sender with the same guarantees as in Theorem 1.1 (or even Theorem 1.2).

However, we ideally wish to obtain general predicting attacks that do not rely on any additional information to correctly output the identity of the sender. We generically upgrade any targeted-predicting attack to a standard predicting attack, while preserving the advantage δ, as follows. The attack simply runs the targeted-predicting attack on all pairs of distinct indices $\{(i, j) \,|\, i, j \in [N], i \neq j\}$, and outputs as the sender the party i^* that got designated as the sender in all the runs (i^*, j), $j \neq i^*$.[13] Now, if i^* is the sender of the N-party AT, an union bound implies that the probability that all the internal runs $(i^*, j), j \neq i^*$ of the targeted-predicting attack correctly point out to i^* as the sender is at least $\delta' \geq 1 - N(1 - \delta)$. Starting with the attack from Theorem 1.1 with $\alpha' = N \cdot \alpha$,[14] we obtain the same lower bound as Theorem 1.1 in the N-party setting, at the cost of a $\mathrm{poly}(N)$ overhead in the runtime of our attack.[15]

[12] This is done by considering all the messages sent by parties $k \neq i, j$ as part of the CRS of the new 2-party protocol.

[13] Note that there is at most one such index. If no such index exist, our attack, say, outputs party 1.

[14] This corresponds to setting $\delta = 1 - 1/\alpha'$, conditioned on executions where the message can be correctly reconstructed. We refer to Section 5.3 for more details.

[15] The overhead arises from both the $O(N^2)$ calls to the internal distinguisher, and the runtime of the internal distinguisher itself which is $\mathrm{poly}(\alpha') = \mathrm{poly}(N) \cdot \alpha$.

3 Preliminaries and Definitions

Notations. When X is a distribution, or a random variable following this distribution, we let $x \leftarrow X$ denote the process of sampling x according to the distribution X. If X is a set, we let $x \leftarrow X$ denote sampling x uniformly at random from X; if Alg is a randomized algorithm, we denote by $x \leftarrow \mathsf{Alg}$ the process of sampling an output of Alg using uniformly random coins. We use the notation $[k]$ to denote the set $\{1, \ldots, k\}$ where $k \in \mathbb{N}$, and $[0,1]_{\mathbb{R}}$ to denote the real interval $\{x \in \mathbb{R} \,|\, 0 \leq x \leq 1\}$. We denote by $\mathrm{negl}(\lambda)$ functions f such that $f(\lambda) = 1/\lambda^{\omega(1)}$.

Chernoff Bound. We will use the following (multiplicative) form of Chernoff-Hoeffding inequality.

Lemma 3.1 (Multiplicative Chernoff.)**.** *Suppose $X_1, \cdots, X_n$ are independent Bernouilli variables with common mean p. Then, for all $t > 0$, we have:*

$$\Pr\left[\sum_{i=1}^{n} X_i \notin [(1-t)\cdot np, (1+t)\cdot np)\right] \leq 2e^{-2t^2p^2n}.$$

3.1 Anonymous Transfer

We recall here the notion anonymous transfer, introduced in [ACM22]. Throughout most of this work, we focus the two-party setting, involving a sender, a dummy non-sender and a (silent) receiver.[16]

Definition 3.2 ((Two-Party, Silent-Receiver) Anonymous Transfer); adapted from [ACM22])**.** *A two-party anonymous transfer (AT) Π^{ℓ}_{AT}, with correctness error $\varepsilon \in [0,1]_{\mathbb{R}}$, anonymity $\delta \in [0,1]_{\mathbb{R}}$, consisting of $c \in \mathbb{N}$ rounds and message length $\ell \in \mathbb{N}$ (all possibly functions of λ), is a tuple of PPT algorithms* $(\mathsf{Setup}, \mathsf{Transfer}, \mathsf{Reconstruct})$ *with the following specifications:*

- $\mathsf{Setup}(1^{\lambda})$ *takes as input a unary encoding of the security parameter λ and outputs a common reference string* crs*.*
- $\mathsf{Transfer}(\mathsf{crs}, b, \mu)$ *takes as input a common reference string* crs*, the index of the sender $b \in \{0,1\}$, the message to be transferred $\mu \in \{0,1\}^{\ell}$, and outputs a transcript π. Transcripts π consists of c rounds of interaction between the sender and the dummy party, where the dummy party (with index $1-b$) sends uniform and independent messages at each round, and the each message from the sender depends on the partial transcript so far, with a next message function implicitly defined by* $\mathsf{Transfer}(\mathsf{crs}, b, \mu)$*.*

  By default, we assume that the receiver does not send any messages (namely, the receiver is silent*).*[17]
- $\mathsf{Reconstruct}(\mathsf{crs}, \pi)$ *takes as input a common reference string* crs*, a transcript π and outputs a message $\mu' \in \{0,1\}^{\ell}$.*

  By default, we assume that $\mathsf{Reconstruct}$ *is deterministic.*[18]

We require that the following properties are satisfied.

ε-Correctness. *An anonymous transfer Π^{ℓ}_{AT} has correctness error ε if, for all large enough security parameter λ, index $b \in \{0,1\}$, message length $\ell \in \mathrm{poly}(\lambda)$, and all message $\mu \in \{0,1\}^{\ell}$, we have:*

$$\Pr\left[\begin{array}{ll} \mathsf{crs} \leftarrow \mathsf{Setup}(1^{\lambda}) & \\ \pi \leftarrow \mathsf{Transfer}(\mathsf{crs}, b, \mu) & : \mu' \neq \mu \\ \mu' \leftarrow \mathsf{Reconstruct}(\mathsf{crs}, \pi) & \end{array}\right] \leq \varepsilon.$$

[16] The work of [ACM22] more generally considers a setting with N parties, namely a sender and $N-1$ dummy parties. Our work focuses on the two-party case, but our main result extend to the N-party case: see Remark 3.6 and Section 5.3 for more details.

[17] This is without loss of generality; see Remark 3.4.

[18] This is without loss of generality; see Remark 3.5

δ-Anonymity. *An anonymous transfer Π^{ℓ}_{AT} is δ-anonymous if, for all PPT algorithm D, all large enough security parameter λ, message length $\ell \in \mathrm{poly}(\lambda)$, and all message $\mu \in \{0,1\}^{\ell}$,*

$$\left| \begin{array}{l} \Pr[\pi^{(0)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, 0, m) : D(\pi^{(0)}) = 1] \\ - \Pr[\pi^{(1)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, 1, m) : D(\pi^{(1)}) = 1] \end{array} \right| \leq \delta, \tag{1}$$

where the probability is over the randomness of Setup, Transfer, *and the internal randomness of D.*

We alternatively say that Π^{ℓ}_{AT} is δ-anonymous with respect to a class of adversaries $\mathcal{C}$, *if Eq. (1) holds instead for all distinguishers $D \in \mathcal{C}$.*

Remark 3.3 (Comparison with [ACM22])**.** Our notation and definitions are slightly different but equivalent from the ones from [ACM22]. With our conventions, ε denotes a correctness error, and δ denotes a (bound on a) distinguishing advantage, and therefore an AT has stronger correctness (resp. stronger anonymity) guarantees as ε, δ decrease. This is the opposite of [ACM22], where guarantees gets better as their parameters ε, δ tend to 1.

Our definition of correctness error is over the random choice of $\mathsf{crs} \leftarrow \mathsf{Setup}$, while it is worst case over crs in [ACM22]. Because this defines a larger class of protocols, ruling out the definition above makes our lower bounds stronger.

Our definition of δ-anonymity is formulated specifically for the two-party case, and is worded differently from theirs, which states, up to the mapping $\delta \mapsto 1 - \delta$ discussed above:

$$\left| \Pr_{b \leftarrow \{0,1\}} \left[\pi^{(b)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, b, m) : D(\pi^{(b)}) = b \right] - \frac{1}{2} \right| \leq \frac{\delta}{2}. \tag{2}$$

However one can easily show that both definitions correspond to the same value δ .

Remark 3.4 (Silent Receivers)**.** As noted in [ACM22], without loss of generality, the receiver in an anonymous transfer can be made silent, namely, does not send any messages in the protocol execution. This is because its random tape can be hard-coded in the CRS.

Remark 3.5 (Deterministic reconstruction)**.** We observe that Reconstruct can be assumed to be deterministic without loss of generality; this is because random coins for Reconstruct can be sampled and included in the common reference string crs.

Remark 3.6 (AT with larger number of parties)**.** [ACM22] more generally defines anonymous transfer with a larger number of participants $N \in \mathbb{N}$. We refer to [ACM22, Definition 3] for a formal definition.[19] The main difference (in the silent receiver case), is that δ is defined as an advantage over random guessing among the N participants. Namely, Eq. (2) becomes:

$$\left| \Pr_{k \leftarrow [N]} \left[\pi^{(k)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, k, m) : D(\pi^{(k)}) = k \right] - \frac{1}{N} \right| \leq \delta \cdot \frac{N-1}{N}.$$

In particular, while the indistinguishability-based definition in Eq. (1) and the predicting-based definition in Eq. (2) are equivalent in the two-party setting, it is not immediately clear that this holds in the N-party setting. Looking ahead, in order to extend our results from the 2-party to the N-party setting, our main observation is to show that this equivalence in fact holds, up to some mild loss in the parameters. We refer to Section 5.3 for more details.

4 Identifying Covert Cheaters

In this section, we introduce our abstraction of the *covert cheating games*, and then show generic strategies for the game.

[19] We remind the reader that the quantity δ in [ACM22] corresponds to $1 - \delta$ for us. Additionally, in this version of the definition, we do not include the receiver in the count for the number of parties.

4.1 Covert Cheating Game

We define a covert cheating game as follows.

Definition 4.1 (Covert Cheating Game)**.** *Let $c \in \mathbb{N}$, $p_0 \in]0,1[_{\mathbb{R}}$ be parameters.*

- ***Setup, players and roles.*** *A* covert cheating game *is played by two (randomized) players A and B, who can agree on a strategy in advance. They play against an* observer *C. During setup, one party is (randomly) designated as the* bias inducer *while the other is designated as the* neutral party*.*

- ***Execution and states of a game.*** *An execution of the game consists of players A and B take alternate moves making moves in the game, with the convention that player A makes the first move. The game consists of c rounds (that is, $2c$ total moves, c moves from A and c moves from B), where $c \in \mathbb{N}$ is a parameter of the game. At any point during the game, the current* state *of a game is represented by a real number $p \in [0,1]_{\mathbb{R}}$. The final state of the game is a bit $p_{2c} \in \{0,1\}$ (where one can consider 1 as a winning outcome for the players A, B, and 0 as a losing outcome).*

  For $k \in [c]$, if A is the bias inducer, we will use either of the notations $X_{2k-1} = X^{(A)}_{2k-1}$ (resp. $X_{2k} = X^{(B)}_{2k}$), to denote the random variable associated to the state resulting from A (resp. B) making its kth move. In other words, the superscript in $X^{(A)}_{2k-1}$ (resp. $X^{(B)}_{2k}$) is a redundant notation to make remind the reader that A (resp. B) made the $(2k-1)$st (resp. $(2k)$th) move of the game.

  Similarly, for $k \in [c]$, if B is the bias inducer, we will use either of the notations $Y_{2k-1} = Y^{(A)}_{2k-1}$ (resp. $Y_{2k} = Y^{(B)}_{2k}$), to denote the random variable associated to the state resulting from A (resp. B) making its kth move. Again, the redundant superscript is used to make the player associated to the move explicit.

  The initial state of the game is defined as $p_0 \in]0,1[_{\mathbb{R}}$, where p_0 is a parameter of the game. In other words, $X_0 = Y_0 = p_0$.

  We say that a strategy for A and B has success rate *p_f if $\mathbb{E}[X_{2c}] \geq p_f$ and $\mathbb{E}[Y_{2c}] \geq p_f$.*

- ***Rules on moves.*** *The* neutral party *is restricted to making moves that do not change the state of the game on expectation, namely, the moves behave as martingales with respect to the current game state. More formally, with our notation, for all $k \in [c]$, we have:*[20]

$$\mathbb{E}[X^{(B)}_{2k} | X_{2k-1}, \cdots, X_0] = X_{2k-1}. \tag{3}$$

$$\mathbb{E}[Y^{(A)}_{2k-1} | Y_{2k}, \cdots, Y_0] = Y_{2k-2}. \tag{4}$$

  where the first equation (resp. second equation) corresponds to A (resp. B) being the bias inducer.

- ***Objective of the game.*** *The goal of the game is, for the bias inducer, to be covert with respect to the observer C, while maintaining a high success rate p_f (namely, a high probability of ending up at 1 in the final state). The observer C has access to intermediate states of the execution $p_i \leftarrow X_i$ (if A is the bias inducer, or $p_i \leftarrow Y_i$ otherwise) via a (distribution of) oracles $\mathcal{O}$. In each oracle $\mathcal{O}$ is hard-coded a sequence of $2c$ states of the game $p_i, i \leq 2c$ induced by an execution of the game. We will respectively denote by $\mathcal{O}^{(A)}$ (resp. $\mathcal{O}^{(B)}$) (the distribution of) oracles corresponding to when A (resp. B) is designated as the bias inducer. We consider the following variants of the oracles $\mathcal{O} \in \{\mathcal{O}^{(A)}, \mathcal{O}^{(B)}\}$.*

  - *Sampling access. We say that the observer C gets* sampling access *to game states $p_i \in [0,1]_{\mathbb{R}}$, if oracles $\mathcal{O}$ are probabilistic oracles such that, for all $i \in [2c]$, $\mathcal{O}(i) = 1$ with probability p_i, and $\mathcal{O}(i) = 0$ with probability $1 - p_i$, where the randomness is uniformly and independently sampled at each oracle call. This is our default notion of access.*

[20] We technically are also conditioning the expectations X, Y on all the prior moves instead, but are omitting them for ease of notation. See Remark 4.2.

– *Direct access.* *We say that an observer gets* direct access *to game states $p_i \in [0,1]_{\mathbb{R}}$, if oracles $\mathcal{O}$ are defined as $\mathcal{O}_{\mathsf{direct}}(i) = p_i \in [0,1]_{\mathbb{R}}$ for all $i \in [2c]$.*

We say that the bias inducer successfully δ-fools a class $\mathcal{C}$ of observers with respect to sampling access if for every algorithm $C \in \mathcal{C}$, we have:

$$\left| \Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] \right| \leq \delta,$$

where $C^{\mathcal{O}^{(\mathcal{X})}}$, where $\mathcal{X} \in \{A, B\}$, denotes the experiment of sampling $\mathcal{O} \leftarrow \mathcal{O}^{(\mathcal{X})}$ (which is defined as sampling a random execution of the game when $\mathcal{X}$ is the bias inducer, yielding states $p_i, i \leq 2c$, and defining $\mathcal{O}$ with respect to $\{p_i\}$), and giving C oracle access to $\mathcal{O}$.

We say that the bias inducer successfully δ-fools a class $\mathcal{C}$ of observers with respect to direct access if the observer C gets oracle access to $\mathcal{O}_{\mathsf{direct}}$ instead.

- ***(Optional Property): Symmetricity.*** *We say that a strategy for players A and B is* symmetric *if:*

$$\forall k \in [c], \mathbb{E}\left[X_{2k}\right] = \mathbb{E}\left[Y_{2k}\right], \tag{5}$$

that is, the state of the game is (on expectation) independent of the identity of the bias inducer, whenever the bias inducer and the neutral party made an identical number of moves (which happens after an even number of total moves).

- ***(Optional Property): Absorption.*** *We say that a strategy is* absorbent *(implicitly, with respect to states 0 and 1) if, for all $i \in [2c]$ and bit $b \in \{0,1\}$:*

$$\{X_i = b\} \Longrightarrow \{\forall j \geq i, X_j = b\}. \tag{6}$$

Remark 4.2 (Implicit Conditioning on Prior Moves.)**.** We allow the strategies from A and B to be adaptive, namely, to depend on prior moves. As a result, all the expectations on random variables X_i, Y_i are technically always considered conditioned on all the prior moves. For ease of notation, we will not make this conditioning explicit, and will always implicitly consider the conditional version of expectations for these variables (and resulting variables defined as a function of X_i, Y_i).

4.2 Attack 1: Free-Lunch Attack with Weak Distinguishing Guarantees

We show here that there exists a very efficient generic observer strategy given sampling access to game states with small but non-negligible distinguishing advantage. Namely:

Theorem 4.3 (Free-Lunch Distinguisher)**.** *For any covert cheating game , consisting of $2c$ total moves, with starting state p_0 and satisfying symmetricity (see Definition 4.1, Eq. (5)), and any strategy for that game with success rate $p_f > p_0$, there exists an observer strategy C^* that determines the identity of the bias inducer with advantage $\delta = \frac{p_f - p_0}{c}$, by making a* single *call to the sampling oracle $\mathcal{O}$.*

In other words, the strategy does not δ-fool the class of observers making a single sampling oracle call.

Proof. We build our observer strategy as follows:

Observer C^*:

- Pick a random $k \leftarrow [c]$. Output $\mathcal{O}(2k-1) \in \{0,1\}$.

In other words, C^* picks a random move from A, and outputs 1 with probability the state of the game after A's kth move. Let us analyze the advantage of C^*.

Suppose A is the bias inducer. Then:

$$\mathbb{E}_{k \leftarrow [c]}\left[\mathcal{O}^{(A)}(2k-1)\right] = \mathbb{E}\left[X_{2k-1}\right],$$

and we furthermore have by Eq. (3) that for all $k \in [c]$:

$$\mathbb{E}\left[X^{(B)}_{2k} - X^{(A)}_{2k-1}\right] = 0, \text{[21]} \tag{7}$$

namely, B's moves do not change X on expectation.

Suppose now that B is the bias inducer. Then, Eq. (4) gives that for all $k \in [c]$:

$$\mathbb{E}\left[Y^{(A)}_{2k-1} - Y^{(B)}_{2k-2}\right] = 0, \tag{8}$$

namely, A's moves do not change Y on expectation. This gives:

$$\begin{aligned}\mathbb{E}_{k\leftarrow[c]}\left[\mathcal{O}^{(B)}(2k-1)\right] &= \mathbb{E}\left[Y_{2k-1}\right] \\ &= \mathbb{E}\left[Y_{2k-2}\right] \\ &= \mathbb{E}\left[X_{2k-2}\right],\end{aligned}$$

where the second equality comes from Eq. (8), and the last equality follows by symmetry if $k > 1$ (Eq. (5)), or as $X_0 = Y_0 = p_0$ if $k = 1$.

Overall, we obtain that the advantage of C^* is, by telescoping:

$$\begin{aligned}&\left|\Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right]\right| \\ &\geq \Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] \\ &= \mathbb{E}_{k\xleftarrow{\$}[c]}\left[X^{(A)}_{2k-1} - Y^{(A)}_{2k-1}\right] \\ &= \mathbb{E}_{k\xleftarrow{\$}[c]}\left[X^{(A)}_{2k-1} - X^{(B)}_{2k-2}\right] \\ &= \sum_{k\in[c]} \frac{\mathbb{E}\left[X^{(A)}_{2k-1} - X^{(B)}_{2k-2}\right]}{c} + \underbrace{\frac{\mathbb{E}\left[X^{(B)}_{2k} - X^{(A)}_{2k-1}\right]}{c}}_{=0 \text{ (Eq. (7))}} \\ &= \frac{\mathbb{E}\left[X_{2c} - X_0\right]}{c} \\ &= \frac{p_f - p_0}{c}\end{aligned}$$

which concludes the proof. □

Remark 4.4 (Correct predictions)**.** Our attack provides a slightly better guarantee than stated in Theorem 4.3: it correctly outputs the identity of the bias inducer (say by associating output 1 to A being the bias inducer), as opposed to simply distinguishing them. In other words, we have:

$$\Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] = \frac{p_f - p_0}{c}.$$

4.3 Attack 2.1: A Strong Attack given Direct Access to States

Next, we describe a generic attack with large advantage, given direct access to game states. We refer to the technical overview (Section 2) for an intuition of the attack. Compared to the exposition in the technical overview, the main difference is that we have to deal with games where the end state is not consistently p_f, but rather 0 or 1 with expectation p_f. This does lead to technical complications. Indeed, one crucial argument in our proof is that the (multiplicative) contribution of the neutral party is 1 on expectation, which

[21] Recall that each expectation is implicitly conditioned on prior moves (Remark 4.2).

allows us to call Markov's inequality. However, switching to the expectation statement, conditioning on the end state being 1 might skew the contribution of the neutral party, which might prevent us from concluding. Instead, we carefully define several useful events, which allows us to compute the advantage of our strategy without ever conditioning on successful runs. More precisely, we now prove the slightly stronger statement that for all winning executions (that is, such that $p_{2c} = 1$), we only fail to identify the sender with small probability $\sqrt{p_0}$.[22]

Last, there is a minor technicality in how to handle denominators being equal to 0 (again, where we do not wish to condition on denominators not being equal to 0), which we solve by requiring a stronger, but natural "absorption" property of the covert cheating game (Definition 4.1, Eq. (6)).

Theorem 4.5 (Strong Distinguisher given Direct Access)**.** *For any covert cheating game satisfying absorption (Definition 4.1, Eq. (6)), consisting of $2c$ total moves, with starting state $p_0 > 0$, and any strategy for that game with success rate $p_f > 2\sqrt{p_0}$, there exists an observer strategy C^* that determines the identity of the bias inducer with advantage at least $p_f - 2\sqrt{p_0}$ given $2c$ oracle calls to the direct access oracle $\mathcal{O}_{\mathsf{direct}}$ (Definition 4.1).*

Proof. We describe the observer strategy.

Observer C^*:

- Compute for all $i \in [2c]$: $p_i = \mathcal{O}_{\mathsf{direct}}(i)$. If $p_{2c} = 0$, output a random bit $\beta \leftarrow \{0,1\}$.
  Otherwise, compute:
$$t^{(A)} = \prod_{k=1}^{c} \frac{p_{2k-1}}{p_{2k-2}},$$
$$t^{(B)} = \prod_{k=1}^{c} \frac{p_{2k}}{p_{2k-1}},$$
  with the convention that $t^{(A)} = 1$ (resp. $t^{(B)} = 1$) if $p_{2k-2} = 0$ for some $k \in [c]$ (resp. if $p_{2k-1} = 0$ for some $k \in [c]$).
- Output 1 (that we associate to outputting "A") if $t^{(A)} \geq \sqrt{\frac{1}{p_0}}$. Otherwise, if $t^{(B)} \geq \sqrt{\frac{1}{p_0}}$, output 0 (that we associate to outputting "B"). Otherwise, output $\perp$.[23]

Let us analyze the advantage of C^*.

Case 1: A is the bias inducer. Suppose A is the bias inducer. We define the following events:

$$\mathsf{CORRECT}_A \coloneqq \{X_{2c} = 1\};$$
$$\mathsf{LARGE}_A^{(A)} \coloneqq \left\{t^{(A)} \geq \sqrt{\frac{1}{p_0}}\right\};$$
$$\mathsf{SMALL}_A^{(B)} \coloneqq \left\{t^{(B)} < \sqrt{\frac{1}{p_0}}\right\};$$
$$\mathsf{GOOD}_A \coloneqq \mathsf{CORRECT}_A \wedge \mathsf{LARGE}_A^{(A)} \wedge \mathsf{SMALL}_A^{(B)}.$$

Note that if GOOD_A occurs, our algorithm is correct when A is the bias inducer. We argue that GOOD_A occurs with high probability. We start by analyzing the contribution $t^{(B)}$ of B.

[22] This is a stronger statement in the sense that observers can test whether an execution is winning, and therefore can output an arbitrary bit $p_{2c} \neq 1$.

[23] Technically, $\perp$ can be replaced by any arbitrary output, e.g. 0; but considering this output separately is in our eyes conceptually cleaner.

Lemma 4.6. *We have:*

$$\Pr[\mathsf{SMALL}_A^{(B)}] \geq 1 - \sqrt{p_0}.$$

Proof. For $k \in [c]$, let us define the partial product of ratios associated to B:

$$P_k^{(B)} = \prod_{j=1}^{k} \frac{X_{2j}^{(B)}}{X_{2j-1}^{(A)}},$$

with the convention that $P_k^{(B)} = 1$ if $p_{2j-1} = 0$ for some $j \in [k]$, and observe that:

$$t^{(B)} \leftarrow P_c^{(B)}.$$

First, observe that

$$\mathbb{E}[P_1^{(B)}] = \frac{\mathbb{E}[X_1^{(B)}]}{p_0} = 1,$$

by Eq. (4).

Let $k \in \{2, \cdots, c\}$; suppose that $\mathbb{E}[P_{k-1}^{(B)}] = 1$. We have:

$$\begin{aligned}
\mathbb{E}[P_k^{(B)}] &= \mathbb{E}_{Y_0,\cdots Y_{2k-1}}\left[\mathbb{E}[P_k^{(B)}|Y_0,\cdots,Y_{2k-1}]\right] \\
&= \mathbb{E}_{p_0,\cdots p_{2k-1}}\left[\mathbb{E}\left[P_{k-1}^{(B)} \cdot \frac{X_{2k}^{(B)}}{X_{2k-1}^{(A)}}\middle| Y_0 = p_0,\cdots,Y_{2k-1} = p_{2k-1}\right]\right] \\
&= \mathbb{E}_{p_0,\cdots p_{2k-1}}\left[\mathbb{E}\left[P_{k-1}^{(B)} \cdot \frac{X_{2k}^{(B)}}{p_{2k-1}}\middle| Y_0 = p_0,\cdots,Y_{2k-1} = p_{2k-1}\right]\right] \\
&= \mathbb{E}_{p_0,\cdots p_{2k-1}}\left[t_{k-1}^{(B)} \cdot \frac{\mathbb{E}\left[X_{2k}^{(B)}\right]}{p_{2k-1}}\middle| Y_0 = p_0,\cdots,Y_{2k-1} = p_{2k-1}\right] \\
&= \mathbb{E}_{p_0,\cdots p_{2k-1}}\left[t_{k-1}^{(B)}\middle| Y_0 = p_0,\cdots,Y_{2k-1} = p_{2k-1}\right] \\
&= P_{k-1}^{(B)},
\end{aligned}$$

where we define $t_{k-1}^{(B)} = t_{k-1}^{(B)}(p_0,\cdots,p_{2k})$ as: $t_{k-1}^{(B)} = \prod_{j=1}^{k-1} \frac{p_{2j}}{p_{2j-1}}$, and where the second to last equality follows by Eq. (4), and with the convention that a fraction with denominator 0 is equal to 1. Therefore, for all $k \in [c]$ (and in particular $k = c-1$), we have:

$$\mathbb{E}[P_k^{(B)}] = 1. \tag{9}$$

Markov's inequality thus gives:

$$\Pr[\neg\mathsf{SMALL}_A^{(B)}] = \Pr\left[P_k^{(B)} \geq \sqrt{\frac{1}{p_0}}\right] \leq \sqrt{p_0},$$

which concludes the proof of Lemma 4.6. □

Next, by definition of success rate and p_f, we have $\Pr[\mathsf{CORRECT}_A] \geq p_f$. Thus:

$$\Pr\left[\mathsf{CORRECT}_A \wedge \mathsf{SMALL}_A^{(B)}\right] \geq \Pr\left[\mathsf{CORRECT}_A\right] - \Pr\left[\neg\mathsf{SMALL}_A^{(B)}\right] \geq p_f - \sqrt{p_0}.$$

Last, we observe that $\mathsf{CORRECT}_A \wedge \mathsf{SMALL}_A^{(B)}$ implies $\mathsf{CORRECT}_A \wedge \mathsf{LARGE}_A^{(A)} \wedge \mathsf{SMALL}_A^{(B)}$. Indeed, suppose $\mathsf{CORRECT}_A$ occurs. By absorption of the game Eq. (6), none of the terms used in a denominator equal 0

(otherwise the final state would be 0). Furthermore, whenever $\mathsf{CORRECT}_A$ occurs, we have by a telescoping product:

$$t^{(A)} \cdot t^{(B)} = 1,$$

and therefore, $t^{(B)} < \sqrt{1/p_0}$ (given by $\mathsf{SMALL}_A^{(B)}$) implies that $t^{(A)} \geq \sqrt{1/p_0}$, namely that $\mathsf{LARGE}_A^{(A)}$ occurs.

Overall, this ensures:

$$\Pr[\mathsf{GOOD}_A] \geq \Pr[\mathsf{CORRECT}_A \wedge \mathsf{SMALL}_A^{(B)}] \geq p_f - \sqrt{p_0},$$

and therefore C^* will be correct with probability at least $p_f - \sqrt{p_0}$ when A is the bias inducer when $\mathsf{CORRECT}_A$ occurs, and correct with probability $1/2$ when $\neg\mathsf{CORRECT}_A$ occurs (which occurs with probability $1 - p_f$ by definition of p_f). In other words, when A is the bias inducer, C^* outputs 1 with probability at least $p_f - \sqrt{p_0} + (1 - p_f)/2$.

Case 2: B is the bias inducer. Suppose now B is the bias inducer. We can similarly define:

$$\begin{aligned}
\mathsf{CORRECT}_B &:= \{Y_{2c} = 1\}; \\
\mathsf{LARGE}_B^{(B)} &:= \left\{ t^{(B)} \geq \sqrt{\frac{1}{p_0}} \right\}; \\
\mathsf{SMALL}_B^{(A)} &:= \left\{ t^{(A)} < \sqrt{\frac{1}{p_0}} \right\}; \\
\mathsf{GOOD}_B &:= \mathsf{LARGE}_A^{(A)} \wedge \mathsf{SMALL}_A^{(B)}.
\end{aligned}$$

An almost identical analysis (using random variables X instead of Y, and shifting the indices appropriately) shows that

$$\Pr[\mathsf{GOOD}_B] \geq p_f - \sqrt{p_0},$$

and therefore C^* will be correct with probability at least $p_f - \sqrt{p_0} + (1 - p_f)/2$ when B is the bias inducer.

Wrapping up. Overall, the advantage of C^* is

$$\begin{aligned}
\left| \Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] \right| &\geq \Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] \\
&\geq 2(p_f - \sqrt{p_0}) - 1 + (1 - p_f) = p_f - 2\sqrt{p_0},
\end{aligned}$$

which concludes the proof. □

4.4 Attack 2.2: A Strong Attack given Sampling Access to States

Next, we port our attack from Section 4.3 to the much weaker sampling setting. Overall, this new attack

- works in the much weaker sampling setting, and does not require the game to satisfy absorption (Eq. (6)),
- but has slightly weaker advantage $\approx p_f - 1/\mathrm{poly}(\lambda)$ while requiring p_0 to be fairly small (the advantage holding for any $\mathrm{poly}(\lambda)$ of our choice, as long as p_0 is small enough), and has a quite larger polynomial sample complexity $q \approx c^6 \cdot \mathrm{poly}(\lambda)$ with respect to the sampling oracle.[24]

Our new analysis is more involved, as to carefully estimate the multiplicative progress of the players despite having imperfect access to the game states p_i. The main problem arises when the state of the game becomes (say, exponentially close or even equal to) 0. Indeed, such states are indistinguishable from the

[24] Looking ahead, this large sample complexity is a result of the techniques we use in our analysis which require us compute precise estimations for each game state.

state being actually 0 from the view of a polynomial-time observer with only sampling access to the state. However, they cannot be treated using an absorption argument (Eq. (6)), like in Theorem 4.5: this is because Eq. (6) only holds for the two states 0 and 1. We solve this by thresholdizing the (partial) products, and only considering "suffix-products" (that is, over indices $i \geq i^*$ for some index i^*) when all the probabilities handled are large enough (say $\gtrsim 1/c^2$). We refer to Section 2 for an intuition for the attack.

One difference with the overview in Section 2 is, again, that the strategy from the players A, B do not necessarily finish at state $p_{2c} \geq p_f$; this guarantee only holds on expectation. We solve this issue similarly to Theorem 4.5, by defining several useful events, and argue that products associated to the neutral party are small with high probability without (significantly) conditioning. And similarly to Theorem 4.5, we prove a slightly stronger result: for all winning executions of the game (that is, such that $p_{2c} = 1$), we only fail to identify the sender with probability $\approx 1/\mathrm{poly}(\lambda)$.

Overall, we present an attack with guarantees comparable with the ones from Theorem 4.5. Even if the analysis is quite tedious and notation-heavy, it is still very similar in spirit to the one of Theorem 4.5.

Theorem 4.7 (Strong Distinguisher given Sampling Access)**.** *Let $\alpha(\lambda) \geq 1$ be a polynomial. For all covert cheating game satisfying $p_0 \leq O\left(\frac{1}{c^2\alpha^2}\right)$, and any strategy with success rate $p_f \geq 1/\alpha(\lambda)$, there exists an observer C^* that determines the identity of the bias inducer with advantage at least $p_f - 1/\alpha - \mathrm{negl}(\lambda)$.*

Furthermore, the observer makes $c^6\alpha^4\omega(\log^2\lambda)$ calls to the sampling oracle $\mathcal{O}$.

In particular, if $p_0 = \mathrm{negl}(\lambda)$, there is, for all polynomial α, an observer strategy with advantage at least $p_f - 1/\alpha - \mathrm{negl}(\lambda)$, with a query complexity of $c^6\alpha^4\omega(\log^2\lambda)$.

Proof. Suppose the covert cheating game satisfies the constraints on p_0 and p_f; observe that in particular $p_0 \leq p_f$.

We describe our attack, which uses the following parameters:

- $\tau = \tau(\lambda, c) \in [0,1]_{\mathbb{R}}$, a threshold precision for our estimation procedure. We will use $\tau = 1/(64c^2 \cdot \alpha^2(\lambda))$ where α is specified in Theorem 4.7.
- $t = t(\lambda, c) \in [0,1]_{\mathbb{R}}$, a multiplicative approximation factor for our estimation. We will use $t = 1/2c$.
- $s = \mathrm{poly}(\lambda, c, \tau, t)$, a number of repetitions for our estimation. We will set $s = c^6 \cdot \mathrm{poly}(\lambda)$, so that $s = \frac{\log c \cdot \omega(\log\lambda)}{\tau^2 t^2}$.

Observer C^*:

1. (Estimation of p_i's):
   - Set $\widetilde{p_0} = p_0$.
   - For $i = 1$ to $2c$:
     - For $j = 1$ to s, sample $b_j \leftarrow \mathcal{O}(i)$.
     - Compute $\widetilde{p_i} = \frac{1}{s}\sum_{j=1}^{s} b_j$.
   - If $\widetilde{p_{2c}} \leq 1 - \tau$, output a random bit $\beta \leftarrow \{0,1\}$.[25]
   - Otherwise, let i^* be the largest index in $[0, 2c]$ such that $\widetilde{p_i} \leq \tau$ (which exists as we set $p_0 = \widetilde{p_0} \leq \tau$).
2. (Estimation of partial numerator and denominator): Compute

$$\widetilde{t^{(A)}}_{\mathsf{num}} = \prod_{k \mid 2k-2 \geq i^*}^{c} \widetilde{p_{2k-1}}. \qquad \widetilde{t^{(B)}}_{\mathsf{num}} = \frac{1}{\widetilde{p_{2c}}} \prod_{k \mid 2k-1 \geq i^*}^{c} \widetilde{p_{2k}}$$

$$\widetilde{t^{(A)}}_{\mathsf{denom}} = K^{(A)} \cdot \prod_{k \mid 2k-1 \geq i^*}^{c} \widetilde{p_{2k-2}}, \qquad \widetilde{t^{(B)}}_{\mathsf{denom}} = K^{(B)} \cdot \prod_{k \mid 2k \geq i^*}^{c} \widetilde{p_{2k-1}},$$

[25] The choice of the specific output doesn't matter for the sake of the analysis, as long as is the same distribution whenever A or B is the bias inducer.

where $K^{(A)} = \begin{cases} 1 \text{ if } i^* \text{ is odd} \\ \tau \text{ if } i^* \text{ is even,} \end{cases}$ and $K^{(B)} = \begin{cases} \tau \text{ if } i^* \text{ is odd} \\ 1 \text{ if } i^* \text{ is even.} \end{cases}$

In other words, this computes partial products starting at i^* with the convention $\widetilde{p_{i^*}} = \tau$ (which only appears in one denominator, according to the parity of i^*), and $\widetilde{p_{2c}} = 1$.

3. Output: Output 1 (that we associate to outputting "A") if

$$\widetilde{t^{(A)}} := \frac{\widetilde{t^{(A)}_{\mathsf{num}}}}{\widetilde{t^{(A)}_{\mathsf{denom}}}} \geq \sqrt{\frac{1}{\tau}}.$$

Otherwise, output 0 (that we associate to outputting "B") if

$$\widetilde{t^{(B)}} := \frac{\widetilde{t^{(B)}_{\mathsf{num}}}}{\widetilde{t^{(B)}_{\mathsf{denom}}}} \geq \sqrt{\frac{1}{\tau}}.$$

Otherwise, output $\perp$.[26]

Let us analyze the advantage of C^*.

Case 1. **A is the bias inducer.** We define the following events, similar to the proof of Theorem 4.5, adapted to the approximate setting:

$$\begin{aligned} \mathsf{CORRECT}_A &:= \{\widetilde{p_{2c}} \geq 1 - t\}; \\ \mathsf{LARGE}_A^{(A)} &:= \left\{\widetilde{t^{(A)}} \geq \sqrt{\frac{1}{\tau}}\right\}; \\ \mathsf{SMALL}_A^{(B)} &:= \left\{\widetilde{t^{(B)}} < \sqrt{\frac{1}{\tau}}\right\}; \\ \mathsf{GOOD}_A &:= \mathsf{CORRECT}_A \wedge \mathsf{LARGE}_A^{(A)} \wedge \mathsf{SMALL}_A^{(B)}. \end{aligned}$$

We furthermore define the following auxiliary events related to the accuracy of the estimation procedure:

$$\begin{aligned} \mathsf{BAD}_0 &:= \{\forall i \text{ s.t. } p_i \geq \tau, \widetilde{p_i} \notin [(1-t)\,p_i, (1+t)\,p_i]\}; \\ \mathsf{BAD}_1 &:= \{p_{i^*} \leq 2\tau\}. \end{aligned}$$

We first argue that these auxiliary events only hold with negligible probability.

Lemma 4.8. *We have:* $\Pr[\neg\mathsf{BAD}_0 \wedge \neg\mathsf{BAD}_1] = \mathrm{negl}(\lambda)$.

Proof. This follows from routine Chernoff bounds. Define:

$$\mathsf{BAD}_2 := \{\exists i > i^*, p_i \leq \tau/2\}.$$

Combining Chernoff (Lemma 3.1 with $t = 1$) with an union bound over the at most $2c$ indices i gives $\Pr[\mathsf{BAD}_2] \leq 2c \cdot e^{-8s\tau^2}$. Whenever BAD_2 does not occur, we have $\Pr[\mathsf{BAD}_0 \wedge \neg\mathsf{BAD}_2] \leq 4c \cdot e^{-2\tau^2 t^2 s}$ by another combination of Chernoff and an union bound, which overall yields:

$$\Pr[\mathsf{BAD}_0] \leq 6c \cdot e^{-8\tau^2 t^2 s} \leq \mathrm{negl}(\lambda),$$

as long as $\tau^2 t^2 s \geq \log(c)\omega(\log\lambda)$, which holds by our setting of s.

Similarly, a Chernoff bound with $t = 2$ gives $\Pr[\mathsf{BAD}_1] \leq e^{-8\tau^2 s}$, which is negligible as long as $\tau^2 s \geq \omega(\log\lambda)$. □

[26] Again, $\perp$ can be replaced by any arbitrary output, e.g. 0; but considering this output separately is in our eyes conceptually cleaner.

We want to prove two main claims, namely:

(1) Whenever GOOD_A occurs, C^* correctly outputs 1.

(2) GOOD_A occurs with sufficiently high probability (Claim 4.9).

Claim (1) follows, similarly to the case in the proof of Theorem 4.5, from the claim that $\mathsf{CORRECT}_A \wedge \mathsf{SMALL}_A^{(B)}$ holding implies $\mathsf{CORRECT}_A \wedge \mathsf{LARGE}_A^{(A)}$
$\wedge\, \mathsf{SMALL}_A^{(B)}$ holds except with negligible probability. Indeed, whenever
$\mathsf{CORRECT}_A$ and BAD_0 occur, we have

$$\widetilde{t^{(A)}} \cdot \widetilde{t^{(B)}} = 1,$$

and therefore $\widetilde{t^{(B)}} < \frac{1}{2} \cdot \sqrt{1/\tau}$ (given by $\mathsf{SMALL}_A^{(B)}$) implies that $t^{(A)} \geq 2\sqrt{1/\tau}$, namely that $\mathsf{LARGE}_A^{(A)}$ occurs, and Lemma 4.8 concludes the claim.

It therefore suffices to prove (2).

Claim 4.9. *We have:*

$$\Pr[\mathsf{GOOD}_A] \geq p_f - 4c \cdot \sqrt{\tau} - \mathrm{negl}(\lambda).$$

We first show a few intermediate lemmas.

Proof of Claim 4.9. We proceed similarly as in Section 4.3. We start by showing that:

$$\Pr[\mathsf{SMALL}_A^{(B)}] \geq \Pr[\mathsf{SMALL}_A^{(B)} \wedge \neg\mathsf{BAD}_0 \wedge \neg\mathsf{BAD}_1] \geq 1 - c \cdot \sqrt{\tau} - \mathrm{negl}(\lambda). \tag{10}$$

By a similar analysis to Section 4.3, using Eq. (4), we have that for any fixed $i \in [2c]$:

$$\Pr\left[\prod_{k|2k-1\geq i}^{c} \frac{p_{2k}^{(B)}}{p_{2k-1}^{(A)}} \geq \frac{1}{4} \cdot \sqrt{\frac{1}{\tau}}\right] \leq 4\sqrt{\tau}.$$

A union bound over $i = 2k-1 \in [2c]$, (there are c different such products), then gives:

$$\Pr\left[\exists i \in [2c], \quad \prod_{k|2k-1\geq i}^{c} \frac{p_{2k}^{(B)}}{p_{2k-1}^{(a)}} \geq \frac{1}{4} \cdot \sqrt{\frac{1}{\tau}}\right] \leq 4c \cdot \sqrt{\tau}.$$

Furthermore, whenever $\neg\mathsf{BAD}_1$ occurs, we have $p_i^* \leq 2\tau$, that is $1/p_i^* \geq 2/\tau$, so that, using Lemma 4.8:

$$\Pr\left[\frac{1}{2} \cdot \prod_{k|2k-1\geq i^*}^{c} \frac{p_{2k}'^{(B)}}{p_{2k-1}'^{(A)}} \geq \frac{1}{4} \cdot \sqrt{\frac{1}{\tau}}\right] \leq 4c \cdot \sqrt{\tau} + \mathrm{negl}(\lambda),$$

where p' are defined as $p'_{i^*} = \tau$, and $p'_i = p_i$ for all $i \neq i^*$.

Last, whenever $\neg\mathsf{BAD}_0$ additionally occurs, we have:

$$\begin{aligned}
\frac{\widetilde{t_{\mathsf{num}}^{(B)}}}{\widetilde{t_{\mathsf{denom}}^{(B)}}} &= \frac{1}{K^{(B)}} \cdot \frac{1}{\widetilde{p_{2c}}} \cdot \frac{\prod_{k\,|\,2k-1\geq i^*}^{c} \widetilde{p_{2k}}}{\prod_{k\,|\,2k\geq i^*}^{c} \widetilde{p_{2k-1}}} \\
&\leq \left(\frac{1+t}{1-t}\right)^c \prod_{k\,|\,2k-1\geq i^*}^{c} \frac{p_{2k}'^{(B)}}{p_{2k-1}'^{(A)}} \\
&\leq 2 \cdot \prod_{k|2k-1\geq i^*}^{c} \frac{p_{2k}'^{(B)}}{p_{2k-1}'^{(A)}},
\end{aligned}$$

whenever $t \leq 1/2c$. Therefore:

$$\Pr[\neg \mathsf{SMALL}_A^{(B)}] = \Pr\left[\frac{\widetilde{t_{\mathsf{num}}^{(B)}}}{\widetilde{t_{\mathsf{denom}}^{(B)}}} \geq \sqrt{\frac{1}{\tau}}\right] \leq 4c \cdot \sqrt{\tau} + \mathrm{negl}(\lambda).$$

Next, we have that if $\neg \mathsf{BAD}_0$ holds, then $\Pr[\mathsf{CORRECT}_A] \geq p_f$ (by definition of p_f), and therefore $\Pr[\mathsf{CORRECT}_A] \geq p_f - \mathrm{negl}(\lambda)$, and thus

$$\begin{aligned}
&\Pr\left[\mathsf{CORRECT}_A \wedge \mathsf{SMALL}_A^{(B)} \wedge \mathsf{LARGE}_A^{(B)}\right] \\
&\geq \Pr\left[\mathsf{CORRECT}_A \wedge \mathsf{SMALL}_A^{(B)}\right] - \mathrm{negl}(\lambda) \\
&\geq \Pr\left[\mathsf{CORRECT}_A\right] - \Pr\left[\neg \mathsf{SMALL}_A^{(B)}\right] - \mathrm{negl}(\lambda) \\
&\geq p_f - 4c\sqrt{\tau} - \mathrm{negl}(\lambda),
\end{aligned}$$

which concludes the proof of Claim 4.9. □

Overall, if A is the bias inducer, given C^* outputs 1 with probability $1/2$ whenever $\neg \mathsf{CORRECT}_A$ occurs, we have:

$$\Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] \geq p_f - 4c\sqrt{\tau} + \frac{1 - p_f}{2} - \mathrm{negl}(\lambda).$$

Case 2. B is the bias inducer. Similarly to Section 4.3, we define and analyze the analogues of the events when B is the bias inducer, and conclude that in this case, C^* outputs 0 with probability at least $p_f - 4c\sqrt{\tau} + \frac{1-p_f}{2} - \mathrm{negl}(\lambda)$.

Wrapping up. Overall, the advantage of C^* is

$$\begin{aligned}
&\left|\Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right]\right| \\
&\geq \Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] \\
&\geq p_f - 4c\sqrt{\tau} + \frac{1 - p_f}{2} - 1 + (p_f - 4c\sqrt{\tau} + \frac{1 - p_f}{2}) - \mathrm{negl}(\lambda) \\
&= p_f - 8c\sqrt{\tau} - \mathrm{negl}(\lambda),
\end{aligned} \tag{11}$$

and plugging in the parameters in the beginning of the proof gives $8c\sqrt{\tau} = 1/\alpha$, which concludes the proof. □

Remark 4.10 (Correct predictions)**.** Again, our attack provides a slightly better guarantee than stated in Theorem 4.7: it correctly outputs the identity of the bias inducer (say by associating output 1 to A being the bias inducer), as opposed to simply distinguishing them. In other words, we have:

$$\Pr\left[C^{\mathcal{O}^{(A)}} = 1\right] - \Pr\left[C^{\mathcal{O}^{(B)}} = 1\right] = p_f - 8c\sqrt{\tau} - \mathrm{negl}(\lambda).$$

Looking ahead, we will crucially use this fact to extend our result to the many-party case.

Remark 4.11 (Cost of the Attack, and Fine-grained Guarantees)**.** The sampling complexity of our strategy C^* is a large, but fixed polynomial $c^6 \cdot \alpha^4 \cdot \omega(\log^2 \lambda)$. Concretely, in the setting where $p_f \geq K$ for a constant K, and $p_0 = \mathrm{negl}(\lambda)$, we obtain attack with *constant advantage* (or even advantage $1 - 1/\mathrm{poly}$ if $p_f = 1 - 1/\mathrm{poly}$) which has a fixed overhead sampling cost as a function of c.

In other words, our attack rules out combinations of games and strategies that δ-fool *fine-grained observers* with sample complexity $m(c)$, if m is allowed to be a large enough polynomial.[27]

[27] Here, we implicitly take the convention that, because players make c moves, they have complexity at least c. This is informal, and there is a mismatch: we are comparing sample complexity of C^* against standard complexity of A and B. The translation to AT lower bounds will make this statement more precise.

5 Lower Bounds on Anonymous Transfer

In this section, we tie the attacks on covert cheating games in Section 4 to impossibility results for anonymous transfer, thus obtaining Theorem 5.2 and Theorem 5.5. Last, we show how to extend Theorem 5.2 to the N-party setting in Section 5.3.

5.1 Reducing Anonymous Transfer to Covert Cheating Games

Theorem 5.1. *Let Π^{ℓ}_{AT} be a two-party anonymous transfer protocol, with correctness error $\varepsilon \in [0,1]_{\mathbb{R}}$, anonymity $\delta \in [0,1]_{\mathbb{R}}$ with respect to a class $\mathcal{C}$ of adversaries, consisting of $c \in \mathbb{N}$ rounds and message length $\ell \in \mathbb{N}$ (all possibly functions of λ) and satisfying deterministic reconstruction (which is without loss of generality, see Remark 3.5).*

Then there exists a covert cheating game, along with player strategy, where the game consists of c rounds, the initial state of the game is $2^{-\ell}$, the expected final state is $p_f = 1 - \varepsilon$ and the player strategy δ-fools observers in $\mathcal{C}$.

Moreover, the covert cheating game satisfies absorption (Definition 4.1, Eq. (6)), and is symmetric if Π^{ℓ}_{AT} is symmetric (Definition A.2).

Proof. Let $\Pi^{\ell}_{AT} = (\mathsf{Setup}, \mathsf{Transfer}, \mathsf{Reconstruct})$ be an AT with the notation of Theorem 5.1. We define our game as follows.

- **Players and roles.** The players of the game are the participants of the AT. The bias inducer is the sender of the AT using a uniformly random message $\mu \leftarrow \{0,1\}^{\ell}$, the neutral party is the dummy party of the AT, and observers are distinguishers.

- **Execution and states.** Moves in the covert cheating game are messages sent in the AT. In other words, a full execution of the game is a full AT transcript. Because moves in the covert cheating game are sequential, we sequentialize the messages of the AT by consider player A to move first within the round. This induces an order of messages, indexed by $i \in [2c]$.

  Let us fix an execution of the game, that is a full AT transcript $\pi \leftarrow \mathsf{Transfer}(\mathsf{crs}, b, \mu)$, where $\mathsf{crs} \leftarrow \mathsf{Setup}(1^{\lambda})$ and $\mu \leftarrow \{0,1\}^{\ell}$. The associated states of the game p_i, where $i \in [2c]$, are defined as follows. Let $\pi[i]$ denote the partial transcript consisting of the first i messages of the protocol $\mathsf{Transfer}$ (with the sequential order from above). Let $\overline{\pi[i]}$ denote the distribution of *randomly completed* partial transcripts, where $\pi[i]$ is completed with $2c - i$ uniformly sampled random message to obtain a full transcript. We then define:

$$p_i = p(\mathsf{crs}, \pi[i]) \coloneqq \Pr\left[\mu' \leftarrow \mathsf{Reconstruct}(\mathsf{crs}, \overline{\pi[i]}) : \mu' = \mu\right],$$

  where $\mu \leftarrow \{0,1\}^{\ell}$ is the input to the AT sender. The probability is over the randomness of the random completion (recall that $\mathsf{Reconstruct}$ is deterministic).

  The initial state of the game is $p_0 = 1/2^{\ell}$, over the sole randomness of $\mu \leftarrow \{0,1\}^{\ell}$.

  Π^{ℓ}_{AT} having correctness error ε implies that the resulting covert cheating strategies have success rate $p_f = 1 - \varepsilon$. Furthermore, the final state satisfies $p_{2c} \in \{0,1\}$ by determinism of $\mathsf{Reconstruct}$ and definition of p_{2c} (as there is no randomness in $\mathsf{Reconstruct}(\mathsf{crs}, \overline{\pi})$).

- **Restriction on the neutral party.** We argue that Eqs. (3) and (4) hold. This is because in an AT, dummy messages are sampled uniformly at random, and are therefore identically distributed as its counterpart obtained from random completion. More formally, supposing A is the bias inducer/sender, we have for all $k \in [c]$ that the completions $\overline{(\pi[2k-1] \| \mathsf{msg})}$ where msg is a random AT protocol

message, and $\overline{\pi[2k-1]}$ are identically distributed by definition of completion, so that

$$\begin{aligned}\mathbb{E}[X_{2k}^{(B)}|X_{2k-1},\cdots,X_0] &= \mathbb{E}_{\mathsf{msg}}\left[\Pr\left[\mu' \leftarrow \mathsf{Reconstruct}(\mathsf{crs}, \overline{(\pi[2k-1]\|\mathsf{msg})}) : \mu' = \mu\right]\right] \\ &= \Pr\left[\mu' \leftarrow \mathsf{Reconstruct}(\mathsf{crs}, \overline{\pi[2k-1]}) : \mu' = \mu\right] \\ &= X_{2k-1},\end{aligned}$$

and similarly when B is the bias inducer/sender.

- **Observers and security.** Given an AT transcript, we implement a sampling oracle as follows. On input i, sample $\overline{\pi[i]}$ and compute $\mu' \leftarrow \mathsf{Reconstruct}(\mathsf{crs}, \overline{\pi[i]})$ Output 1 if $\mu' = \mu$, and 0 otherwise. By definition, this procedure tosses a coin with probability p_i.

  Overall, if an observer strategy distinguishes $\mathcal{O}^{(\mathcal{A})}$ from $\mathcal{O}^{(\mathcal{B})}$ in time t, with q sampling oracle queries and advantage δ, then there exists a distinguisher for the AT running in time $t + q \cdot (n + \rho(c))$ with advantage δ, where n is the complexity of computing $\mathsf{Reconstruct}$ and $\rho(c)$ is the complexity of sampling c uniformly random protocol messages.

- **Absorption.** Because completions are sampled uniformly random from the whole message space of the protocol, by definition of p_i, $p_i = 1$ implies that all completions of $\pi[i]$ recover μ, which implies that all possible continuations of $\pi[i]$ satisfy $p = 1$. Similarly, $p_i = 0$ implies that all completions of $\pi[i]$ fail to recover μ, so that all continuations of $\pi[i]$ satisfy $p = 0$.

- **Symmetricity.** Suppose the AT is symmetric (Definition A.2), and let $k \in [c]$. Then (1) by symmetry of $\mathsf{Reconstruct}$, $\mathsf{Reconstruct}(\mathsf{crs}, \overline{\pi[2k]})$ is identically distributed as $\mathsf{Reconstruct}(\mathsf{crs}, \overline{\mathsf{Mirror}(\pi[2k])})$, where Mirror flips the identities of the participants in the transcript and (2) by symmetry of $\mathsf{Transfer}$, the unordered set $(\mathsf{dummy}^{(A)}, \mathsf{msg}^{(B)})$ is identically distributed as $(\mathsf{dummy}^{(B)}, \mathsf{msg}^{(A)})$. We can therefore replace all the consecutive pairs of messages $(2j-1, 2j)$ from $\{\mathsf{dummy}_{2j-1}^{(A)}, \mathsf{msg}_{2j}^{(B)}\}$ to $\{\mathsf{msg}_{2j-1}^{(A)}, \mathsf{dummy}_{2j}^{(B)}\}$, for all $j \leq k$, without changing the distribution of the outcome of $\mathsf{Reconstruct}$. Doing so $2k$ times gives:

$$\mathbb{E}[X_{2k}] = \mathbb{E}[Y_{2k}].$$

□

5.2 Lower Bounds on Anonymous Transfer

We first rule out the existence of AT with non-trivial correctness error ε and anonymity δ, that are secure against arbitrary polynomial-time adversaries. We do so by combining Theorem 4.7 with Theorem 5.1, which gives the following:

Theorem 5.2. *Suppose Π_{AT}^{ℓ} is a (two-party, silent receiver) anonymous transfer satisfying deterministic reconstruction, and with $\ell \geq \omega(\log \lambda)$-bit messages, with correctness error ε, and δ-anonymous against all polynomial-time adversaries. Then, for all polynomial $\alpha = \alpha(\lambda)$:*

$$\delta \geq 1 - \varepsilon - 1/\alpha(\lambda).$$

We observe that the relation between δ and ε is almost tight (up to $1/\mathrm{poly}(\lambda)$ factors), namely matches a trivial construction, (Claim A.1).

Remark 5.3 (Ruling out other versions of AT)**.** Thanks to the transformations in Section 3, Theorem 5.2 also rules out other versions of AT, including (all combinations of) the following: AT with non-silent receiver, AT with randomized reconstruction, AT with a large number N of parties (by considering $\delta' = (N-1) \cdot \delta$).

Remark 5.4 (Ruling out strong fine-grained results.)**.** In fact, denoting $n = n(\lambda)$ the running time of $\mathsf{Reconstruct}$, the attack obtained by combining Theorem 4.7 with Theorem 5.1 runs in time $m(\lambda) = n \cdot c^6 \cdot \omega(\log^2(\lambda))$, and therefore Theorem 5.2 further rules out schemes that are secure against adversaries running in fixed polynomial overhead over honest users $m \leq n^7$. In other words, fine-grained results for non-trivial parameters will at most provide security against adversaries running in time m.

Next, we rule out the existence of fine-grained AT, but for a smaller set of parameters. We do so by combining Theorem 4.3 with Theorem 5.1. Note that Theorem 4.3 requires the AT to be symmetric; this is without loss of generality by Claim A.3. This overall gives the following:

Theorem 5.5. *There are no fine-grained AT with ℓ-bit messages, correctness error ε, and anonymity δ, such that:*

$$\delta \cdot c \geq 1 - \varepsilon - 1/2^{\ell}.$$

More precisely, denoting $n = n(\lambda)$ the maximum runtime of $\mathsf{Transfer}, \mathsf{Reconstruct}$, and $\rho(c)$ is the cost of sampling c uniformly random protocol messages, combining Theorem 4.3 with Theorem 5.1 gives an attack with complexity $n(\lambda) + \rho(c) \leq 2n(\lambda)$.

5.3 Extension to Anonymous Transfer with Many Parties

In this section, we show that Theorem 5.2 extends to rule out anonymous transfer with any polynomial number N of parties.[28] More precisely, we prove the following result.

Theorem 5.6. *Let $N = N(\lambda)$ be any polynomial. Suppose Π^{ℓ}_{AT} is an N-party (silent receiver) anonymous transfer satisfying deterministic reconstruction, with $\ell \geq \omega(\log \lambda)$-bit messages, with correctness error ε, and δ-anonymous against all polynomial-time adversaries. Then, for all polynomial $\alpha = \alpha(\lambda)$:*

$$\delta \geq 1 - \varepsilon - 1/\alpha(\lambda).$$

Our proof follows two main steps. In a first step, we observe in Claim 5.9 that the attack underlying Theorem 5.2 directly extends to an attack in the N-party case, given the promise that the sender's identity is restricted to two parties.[29] Then, in Lemma 5.10, we show that such an attack generically implies a standard attack on anonymity (without the promise mentioned above).[30]

We formalize the first step by introducing an alternate definition of anonymity, which is weaker than the original one of [ACM22].

Definition 5.7 (Targeted-predicting anonymity)**.** *We say that an N-party anonymous transfer Π^{ℓ}_{AT} is δ_T-targeted-predicting anonymous, if there exists $\{i, j\} \in [N]$ with $i \neq j$ such that, for all PPT algorithm A,*[31] *all large enough security parameter λ, message length $\ell \in \mathrm{poly}(\lambda)$, and all message $\mu \in \{0,1\}^{\ell}$:*

$$\Pr[\pi^{(i)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, i, m) : A(\pi^{(i)}, \{i, j\}) = i] - \Pr[\pi^{(j)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, j, m) : A(\pi^{(j)}, \{i, j\}) = j] \leq \delta_T, \tag{12}$$

where the probability is over the randomness of Setup, $\mathsf{Transfer}$, *and the internal randomness of A.*

Intuitively, Definition 5.7 ensures that no polynomial-time adversary can infer who the sender is, even given the promise that the sender is either party i or party j.[32] In other words, Definition 5.7 is conceptually

[28] Looking ahead, doing so comes at a mild loss in the resulting anonymity δ. While this loss is mild starting from Theorem 5.2 yielding the main result of the section, it is quite significant when starting from Theorem 5.5, in which case the anonymity guarantees we obtain are similar to the ones of [ACM22]. We therefore focus on Theorem 5.2 in this section.

[29] In fact, the same holds for the attack of Theorem 5.5.

[30] This step results in a loss of advantage that the attack of Theorem 5.5 does not handle well — it would result in a similar final statement than the N-party lower bound proven in [ACM22].

[31] In terms of syntax, A, on input $(\pi, \{i, j\})$, outputs either i or j.

[32] For technical simplicity, we consider a weak notion of anonymity which requires indistinguishability for *some* pair of parties $\{i, j\}$.

similar to indistinguishability-based notion of Definition 3.2, Eq. (1),[33] while the original predicting-based notion of Remark 3.6 from [ACM22] is the analogue of Definition 3.2, Eq. (2).

[ACM22] furthermore showed that any N-party AT can be transformed into a 2-party AT. This is done by hard-coding the randomness of extra dummy parties into the CRS. We observe that this transformation directly preserves targeted-predicting anonymity.[34]

Lemma 5.8 ([ACM22], Section 4.3, rephrased)**.** *Any N-party anonymous transfer with correctness error ε and targeted-predcting-based anonymity δ_T can be transformed into a 2-party protocol with correctness error $\varepsilon' = \varepsilon$ and anonymity δ_T. Furthermore, the number of rounds and message length are preserved, and, moreover, if the original AT is targeted-predicting secure against a class $\mathcal{C}$ of adversaries, then the 2-party protocol is also secure against $\mathcal{C}$.*

Given that our attacks underlying Theorem 5.2 (resp. Theorem 5.5) correctly predict the identity of the senders with high probability (Remark 4.10, resp. Remark 4.4), Theorem 5.2 (resp. Theorem 5.5) directly extend to the targeted-predicting notion above, by giving an explicit attack for all pairs $\{i, j\}$. We now state the claim resulting from Theorem 5.2.

Claim 5.9 (Lower bound for targeted-predicting AT)**.** *Suppose Π^{ℓ}_{AT} is an N-party (silent receiver) anonymous transfer satisfying deterministic reconstruction, and with $\ell \geq \omega(\log \lambda)$-bit messages, with correctness error ε, and δ_T-targeted-predicting anonymous against all polynomial-time adversaries. Then, for all polynomial $\alpha = \alpha(\lambda)$:*

$$\delta_T \geq 1 - \varepsilon - 1/\alpha(\lambda).$$

Next, we show how standard anonymity implies targeted-predicting anonymity (albeit with some loss in the parameters).

Lemma 5.10 (Standard anonymity implies targeted-predicting anonymity)**.** *Suppose Π^{ℓ}_{AT} is an N-party (silent receiver) anonymous transfer satisfying δ-anonymity against all polynomial-time adversaries (Remark 3.6).*

Then, Π^{ℓ}_{AT} is δ_T-targeted-predicting anonymous whenever $\delta \geq 1 - N \cdot (1 - \delta_T)/2$, against all polynomial-time adversaries.

Proof. Let A be a PPT attack against the targeted-predicting anonymity of Π^{ℓ}_{AT}, that is, A satisfies for all $\{i, j\} \in \mathbb{N}$ such that $i \neq j$:

$$\Pr[\pi^{(i)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, i, m) : A(\pi^{(i)}, \{i, j\}) = i] - \Pr[\pi^{(j)} \leftarrow \mathsf{Transfer}(\mathsf{crs}, j, m) : A(\pi^{(j)}, \{i, j\}) = j] \geq \delta_T.$$

We build a (standard) predictor P as follows. Given a transcript π, P runs $A(\pi, \{i, j\})$ for all $i, j \in \mathbb{N}, i \neq j$. If there exists some index i^* such that $A(\pi, \{i^*, j\}) = i^*$ for every $j \neq i^*$, P outputs i^*. Otherwise, P outputs say party 1 as the sender.

We argue that P correctly outputs the sender with probability at least $1 - (N - 1)(1 - \delta_T)$. Suppose that i^* is the sender. By δ_T-targeted-predicting anonymity, we have for all $j \neq i^*$:

$$\Pr[A(\pi, \{i^*, j\}) = i^*] - \Pr[A(\pi, \{i^*, j\}) = j] \geq \delta_T,$$

and thus $\Pr[A(\pi, \{i^*, j\}) = i^*] \geq 1/2 + \delta_T/2$. An union bound then gives that

$$\Pr[\forall j \neq i^*, A(\pi, \{i^*, j\}) = i^*] \geq 1 - (N - 1)(1/2 - \delta_T/2).$$

Noting that P outputs i^* whenever the above occurs (regardless of the calls to A not involving i^*), we have

$$\Pr[P(\pi) = i^*] = \Pr[\forall j \neq i^*, A(\pi, \{i^*, j\}) = i^*] \geq 1 - (N - 1)(1/2 - \delta_T/2)$$

[33] The main difference between these notions is that we require targeted-predicting attacks to correctly output the sender, as opposed to only distinguish the settings, namely, Eq. (12) doesn't have any absolute values.

[34] The original work of [ACM22] showed that any N-party AT with standard (predicting-based) anonymity δ can be transformed into a 2-party protocol with anonymity $\delta' \leq N \cdot \delta$.

and so, calling δ the advantage of P in breaking anonymity (Remark 3.6):

$$\begin{aligned}\delta &\geq \frac{N}{N-1}\left[\Pr[P(\pi)=i^*]-\frac{1}{N}\right]\\ &\geq 1-N\cdot\frac{1-\delta_T}{2},\end{aligned}$$

which proves the lemma. □

To conclude the proof of Theorem 5.6, we note that the attack proving Theorem 5.2 has advantage $1-1/\alpha(\lambda)$ conditioned on reconstruction succeeding on its input transcript π, and therefore so are the attacks corresponding to Claim 5.9. Combining Claim 5.9 with $\alpha' = N\cdot\alpha/2$ and Lemma 5.10 concludes the proof.

Remark 5.11 (Overhead of the extension)**.** The cost of our attack for Theorem 5.6 has a large polynomial overhead in N in its runtime compared to the 2-party case in Theorem 5.2. This is both due to the predictor making $O(N^2)$ calls to the 2-party attack, and setting a smaller distinguishing error $\alpha' = N\cdot\alpha/2$ for the 2-party attack itself (due to the loss in advantage), which incurs an additional $O(N^4)$ overhead.

Remark 5.12 (Towards a milder overhead in N)**.** We succinctly describe here another, slightly more involved reduction, from the N-party case to the 2-case with run-time overhead $O(N^4\cdot\log N)$ instead of the $O(N^6)$ from above.

The reduction, given a N-party anonymous transfer, builds a 2-party AT as follows. It samples as part of the CRS a random index $k\leftarrow[N]$. Then, party 0 in the 2-player AT acts as parties $E_0 \coloneqq \{k,\cdots,k+\lfloor N/2\rfloor \bmod N\}$ in the N-player AT, and party 1 acts as parties $E_1 \coloneqq \{k+\lfloor N/2\rfloor+1 \bmod N, k+(N-1) \bmod N\}$. The sender b in the 2-party AT chooses a random index $i^*\leftarrow E_b$ he owns, and acts as the sender for the N-party AT at index i^*, and emulates dummy parties for $E_b\setminus\{i^*\}$. The dummy party acts as dummy parties for all parties in E_{1-b}.

Correctness follows directly from the correctness of the N-party AT. For security, we start with our attack on 2-party AT, which breaks anonymity *for all* k with probability $\delta_T\geq 1-\varepsilon-1/\alpha$. Given a transcript, our reduction searches for i^* using binary search over k. An union bound over the N possible values of k gives that all the internal calls to the 2-party attack succeed with probability at least $1-\varepsilon-N(1-\delta_T)/2$,[35] and thus our attack has the same probability of success as in the previous "duel-based" approach. However, this reduction now only calls the 2-party distinguisher $\lceil\log N\rceil$ times as opposed to N^2 times.

Acknowledgements. We thank the reviewers for useful comments, especially ones on the N-party setting. Daniel Wichs was supported in part by the National Science Foundation under NSF CNS-1750795, CNS-2055510, and the JP Morgan faculty research award.

References

[ABE14] Susan Andrews, Bryan Burrough, and Sarah Ellison. The snowden saga. *Vanity Fair*, 2014.

[ACM22] Thomas Agrikola, Geoffroy Couteau, and Sven Maier. Anonymous whistleblowing over authenticated channels. In Eike Kiltz and Vinod Vaikuntanathan, editors, *TCC 2022, Part II*, volume 13748 of *LNCS*, pages 685–714. Springer, Heidelberg, November 2022.

[APY20] Ittai Abraham, Benny Pinkas, and Avishay Yanai. Blinder - scalable, robust anonymous committed broadcast. In Jay Ligatti, Xinming Ou, Jonathan Katz, and Giovanni Vigna, editors, *ACM CCS 2020*, pages 1233–1252. ACM Press, November 2020.

[BBC22] Alexei navalny: Russia's jailed vociferous putin critic. *British Broadcasting Corporation*, 2022.

[35] Actually, an union bound over $\lceil N/2\rceil$ values suffices, because the reduction can be adapted to only call the 2-party attack on $k\leq\lceil N/2\rceil$.

[Ber16] Charles Berret. 2016.

[CCD+20] Katriel Cohn-Gordon, Cas Cremers, Benjamin Dowling, Luke Garratt, and Douglas Stebila. A formal security analysis of the signal messaging protocol. *Journal of Cryptology*, 33(4):1914–1983, October 2020.

[CGBM15] Henry Corrigan-Gibbs, Dan Boneh, and David Mazières. Riposte: An anonymous messaging system handling millions of users. In *2015 IEEE Symposium on Security and Privacy*, pages 321–338, 2015.

[Cha88] David Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1(1):65–75, January 1988.

[Cha03] David Chaum. *Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms*, pages 211–219. Springer US, 2003.

[DMS04] Roger Dingledine, Nick Mathewson, and Paul F. Syverson. Tor: The second-generation onion router. In Matt Blaze, editor, *USENIX Security 2004*, pages 303–320. USENIX Association, August 2004.

[ECZB21] Saba Eskandarian, Henry Corrigan-Gibbs, Matei Zaharia, and Dan Boneh. Express: Lowering the cost of metadata-hiding communication with cryptographic privacy. In Michael Bailey and Rachel Greenstadt, editors, *USENIX Security 2021*, pages 1775–1792. USENIX Association, August 2021.

[HLv02] Nicholas J. Hopper, John Langford, and Luis von Ahn. Provably secure steganography. In Moti Yung, editor, *CRYPTO 2002*, volume 2442 of *LNCS*, pages 77–92. Springer, Heidelberg, August 2002.

[Inz18] Bastien Inzaurralde. The cybersecurity 202: Leak charges against treasury official show encrypted apps only as secure as you make them. *The Washinton Post*, 2018.

[NSSD21] Zachary Newman, Sacha Servan-Schreiber, and Srinivas Devadas. Spectrum: High-bandwidth anonymous broadcast with malicious security. Cryptology ePrint Archive, Report 2021/325, 2021. https://eprint.iacr.org/2021/325.

[vH04] Luis von Ahn and Nicholas J. Hopper. Public-key steganography. In Christian Cachin and Jan Camenisch, editors, *EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 323–341. Springer, Heidelberg, May 2004.

[vHL05] Luis von Ahn, Nicholas J. Hopper, and John Langford. Covert two-party computation. In Harold N. Gabow and Ronald Fagin, editors, *37th ACM STOC*, pages 513–522. ACM Press, May 2005.

A Additional Constructions and Transformations for Anonymous Transfer

We present here a construction of a weak form of AT in Appendix A.1, and show in Appendix A.2 a transformation from any AT into one such that its next message functions are symmetric.

A.1 A "Trivial" Anonymous Transfer

We show here a simple family of constructions of anonymous transfer (in the two-party, silent receiver case) that achieves, given $\delta \in [0,1]_{\mathbb{R}}$, anonymity δ and correctness $\varepsilon \geq 1 - \delta - \mathrm{negl}(\lambda)$.

Claim A.1. *For all $\delta \in [0,1]_{\mathbb{R}}$, there exists a single-round anonymous transfer protocol satisfying δ-anonymity and correctness error $\varepsilon = 1 - \delta - \mathrm{negl}(\lambda)$.*

Proof. We define our protocol as follows. The CRS consists of a random bitstring $k \leftarrow \{0,1\}^\lambda$.

Given an input μ, the sender flips a coin b that outputs 1 with probability δ. If $b = 1$, the sender sends the message $(\mu \| k)$, and otherwise sends a random message. The reconstruction algorithm parses both messages as $(\mu_0 \| k_0)$ and $(\mu_1 \| k_1)$. If $k_\beta = k$ for some β, it outputs μ_β (where β is chosen arbitrarily if this holds for both 0 and 1). Otherwise it outputs a random bit.

The protocol is δ-anonymous: this is because, the distribution of sender messages is different from the one of the dummy party with probability δ (namely, when $b = 1$). Correctness follows as the probability for the dummy party β that $k_\beta = k$ is negligible in λ.

□

A.2 Symmetric Anonymous Transfer

In this section, we show that without loss of generality for the sake of lower bounds, protocol specifications for the sender do not depend on its identity.

Definition A.2. *We say that an anonymous transfer is* symmetric *if the next message function of the sender, implicitly defined by* $\mathsf{Transfer}(\mathsf{crs}, b, \mu)$ *where b is the sender, does not depend on b, and if* $\mathsf{Reconstruct}$ *does not depend on the identities of the participants.*

Claim A.3. *Any 2-party anonymous transfer with polynomial-time algorithms, with correctness error ε and anonymity δ can be transformed into a* symmetric *anonymous transfer with correctness error $\varepsilon - \mathrm{negl}(\lambda)$ and anonymity δ. Furthermore, the number of rounds and message lengths are preserved.*

Proof. Let $(\mathsf{Setup}, \mathsf{Transfer}, \mathsf{Reconstruct})$ be an anonymous transfer with correctness error ε and anonymity δ. We define $(\mathsf{Setup}', \mathsf{Transfer}', \mathsf{Reconstruct}')$ as follows. Let us call P_0, P_1 the participants of the transfer.

A CRS from Setup' consists of a CRS from Setup along a one-time information theoretic MAC key k.[36] The next-message function of the new sender, on input μ, works as follows. Sample $b \leftarrow \{0,1\}$, and use the next-message function defined by $\mathsf{Transfer}(\mathsf{crs}, b, (\mu, \mathsf{MAC}(k, \mu))$, where the message to be sent is $(\mu, \mathsf{MAC}(k, \mu))$; this results in a transcript π. $\mathsf{Reconstruct}'(\mathsf{crs}, \pi)$ calls $\mathsf{Reconstruct}$ two times: once by parsing π normally, obtaining a message (μ_0, t_0), and once parsing π after flipping the identities of P_0, P_1, obtaining a message (μ_1, t_1). It uses the MAC key k to verify the MAC tags $\mathsf{Verify}(k, \mu_0, t_0)$ and $\mathsf{Verify}(k, \mu_1, t_1)$, and outputs any μ_b such that verification output accept, and any arbitrary message otherwise.

First, the resulting protocol is symmetric by construction.

We argue anonymity as follows. The distributions of the transcripts induced by P_0 sampling $b = 0$ and P_1 sampling $b = 1$ are δ-indistinguishable by anonymity of the base AT; and similarly distributions of the transcripts induced by P_0 sampling $b = 1$ and P_1 sampling $b = 0$ are also δ-indistinguishable by anonymity of the base AT, but with identities reversed. So the new AT satisfies δ-anonymity.

Correctness follows by perfect correctness and unforgeability of the one-time MAC, and correctness of the base AT. Namely, because $\mathsf{Transfer}(\mathsf{crs}, b, (\mu, \mathsf{MAC}(k, \mu))$ and (μ_0, t_0), (μ_1, t_1) can be computed in polynomial time given $\mathsf{MAC}(k, \mu)$, the probability that some (μ_b, t_b) satisfies $\mu_b \neq \mu$ and $\mathsf{Verify}(k, \mu_b, t_b) = \mathsf{accept}$ is negligible by one-time security of the MAC. In other words, with overwhelming probability, only tags for message μ verify and are potentially output by $\mathsf{Reconstruct}$. Then, correctness of the base AT ensures that if the sender P_β flips a bit b, then flipping identities if $b \neq \beta$ (and doing nothing otherwise) before

[36] Looking ahead, the CRS is perhaps surprisingly reusable. This is because we use MAC security for *correctness*; our adversary for the MAC will be the base AT protocol, which only ever sees a single tag.

calling $\mathsf{Reconstruct}$ recovers $(\mu, \mathsf{MAC}(k, \mu))$ with probability $1 - \varepsilon$. Overall, the new AT has correctness error $\varepsilon - \mathrm{negl}(\lambda)$.

□