Cryptographic Randomized Response Techniques

Andris Ambainis[1], Markus Jakobsson[2], and Helger Lipmaa[3]

[1] Institute of Mathematics and CS, University of Latvia, Raiņa bulv. 29
Rīga, LV-1459, Latvia,
ambainis@lanet.lv
RSA Laboratories, 174 Middlesex Turnpike, Bedford, MA 01730, USA
mjakobsson@rsasecurity.com
[2] Laboratory for Theoretical CS, Department of CS&E
Helsinki University of Technology, P.O.Box 5400, FIN-02015 HUT, Espoo, Finland
helger@tcs.hut.fi

Abstract. We develop cryptographically secure techniques to guarantee unconditional privacy for respondents to polls. Our constructions are efficient and practical, and are shown not to allow cheating respondents to affect the "tally" by more than their own vote — which will be given the exact same weight as that of other respondents. We demonstrate solutions to this problem based on both traditional cryptographic techniques and quantum cryptography.

Keywords: classical cryptography, oblivious transfer, polling, privacy, privacy-preserving data-mining, quantum cryptography, randomized response technique

1 Introduction

In some instances, privacy is a matter of keeping purchase information away from telemarketers, competitors, or other intruders. In other instances, privacy translates to security against traffic analysis, such as for web browsing; or to security of personal location information. In still other instances, which we study in this paper, privacy is a *precondition* to being able to obtain answers to important questions. Two concrete examples of instances of latter are *elections* and *surveys/polls*.

While the first of these examples is the one of the two that has received — by far — the most attention in the field of cryptography, there are important reasons to develop better privacy tools for polling. Surprisingly, the two examples (namely, elections and polls), while quite similar at a first sight, are very different in their requirements. Since it is typically the case that there is more funding available for providing privacy in elections than in surveys and polls, it follows that the tallying process in the former may involve more costly steps than that in the latter — whether the process is electronic (using, e.g., mix networks) or mechanic. Second, while in the case of the voting scheme, we have that users need to entrust their privacy with some set of authorities, it is often the case that there is less trust established between the parties in polls. Yet another reason to treat the two situations separately is that elections involve many more respondents than polls typically do, thereby allowing a unique opinion (e.g., vote) to be hidden among many more in the case of elections than in the case of polls. Finally, while elections require as exact tallying as is possible, *statistical truths* are both sufficient and desirable in polls. This allows the use of polling techniques that are very different from election techniques — in terms of their cost; how tallying is done; and how privacy is protected.

While not given much attention in cryptography, important work on polling has been done in statistics. In particular, the *randomized response technique* (RRT) was proposed by Warner [War65] in 1965, with the goal of being used in polls relating to sensitive issues, such as drug abuse, sexual preferences and shoplifting. The underlying idea behind Warner's proposal is for respondents to randomize each response according to a certain, and known, probability distribution. More precisely, they answer the question truthfully with some probability $p_{\mathsf{ct}} > 1/2$, while with a fixed and known probability $1 - p_{\mathsf{ct}}$ they lie. Thus, users can always claim that their answer — if it is of the "incriminating" type — was a lie. When evaluating all the answers of the poll, these lies become statistically insignificant given a large enough sample (where the size of the sample can be simply computed from the probability distribution governing lying.)

However, a pure RRT by itself is not well suited for all types of polls. E.g., it is believed that people are more likely to vote for somebody who leads the polls than somebody who is behind. Therefore, it could be politically valuable not to lie (as required by the protocol) in polls relating to ones political opinion, and therefore have one's "vote" assigned a greater weight. (This is the case since people with the opposite opinion — if honestly following the protocol — will sometimes cast a vote according to your opinion, but you would never cast a vote according to their opinion, assuming you are willing to cheat.) While the results of the poll remain meaningful if *everybody* cheats (i.e., tells the truth with a probability different from that specified by the protocol), this is *not* the case when only some people deviate from the desired behavior. Also, while one might say that the increased weight in the polls is gained at the price of the cheater's privacy, this is not necessarily the case if the cheater *claims* to have followed the protocol, and there is no evidence to the contrary.

To address the problem of cheating respondents in RRT, we propose the notion of *cryptographic randomized response technique* (CRRT), which is a modification of RRT that prevents cheating. We present three efficient protocols for CRRT; two of them using classic cryptographic methods (and being efficient for different values of p_{ct}), and one using quantum methods. Importantly, the quantum RRT protocol is implementable by using contemporary technology. We give rigorous proofs of security for one of the classical protocols and for the quantum protocol.

For all of our proposed solutions, the privacy of the respondent will be guaranteed information-theoretically (more precisely, statistically). This is appropriate to stimulate truthful feedback on topics that may affect the respondent for years, if not decades. All proposed solutions also *guarantee* that the respondents reply based on the desired probability distributions. Clearly, this requires that the respondent cannot determine the outcome of the protocol (as viewed by the interviewer) before the end of the protocol. Otherwise, he could simply halt the execution of the protocol to suppress answers in which the communicated opinion was a lie. We will therefore require protocols to offer privacy for the *interviewer* as well as for the respondent, meaning that the respondent cannot learn what the outcome of the protocol is, as seen by the interviewer. (One could relax this requirement slightly to allow the respondent to learn the outcome at the same time as the interviewer does, or afterward.)

While we believe that it is important to prevent the respondent from biasing the outcome by selective halting (corresponding to the protocol being *strongly secure*), we also describe simplified versions of our protocols in which this protection mechanism is not available. Such simplified versions (which we refer to as *weakly secure*) can still be useful in some situations. They may, for example, be used as the default scheme for a given application — where they would be replaced by their strongly secure relatives if too many interactions are halted prematurely. (The decision of when the shift would be performed should be based on standard statistical methods, and will not be covered herein.) The benefit of considering such dual modes is that the weakly secure versions typically are computationally less demanding than the strongly secure versions.

Finally, we also discuss cryptographic enhancements to two alternative RRT techniques. In the first, referred to as RRT-IQ, the respondent always gives the truthful answer to the question he is presented with. However, with a certain probability, he is presented with an Innocous Question instead of the intended question. A second alternative RRT technique is what is referred to as *polychotomous* RRT. In this version of RRT, the respondent is given more than two possible options per question.

In particular, our first protocol uses a novel protocol for information-theoretically secure *verifiable oblivious transfer* that enables easier zero-knowledge proofs on the properties of the transferred values. The described protocol may also be useful in other applications. We also note that our techniques have applications in the privacy-preserving data-mining, see Section 3.

Outline. We first review the details of the randomized response technique (Section 2), after which we review some related work in cryptography (Section 3). We then introduce the cryptographic building blocks of our protocols (Section 4). We then describe the functionality of our desired solution in terms of functional black boxes and protocol requirements (Section 5). In Section 6, we present our secure CRRT protocols. In Section 7 we describe cryptographic solutions to other variants of the standard RRT. The appendix contains additional information about the new oblivious transfer protocol and about the quantum RRT protocol.

2 Short Review of Randomized Response Technique

When polling on sensitive issues like sexual behavior or tax evasion, respondents often deny their stigmatizing behavior due to the natural concern about their privacy. In 1965, Warner [War65] proposed the Randomized Response Technique (RRT) for organization of polls where an unbiased estimator (UE) to the summatory information — the proportion of people belonging to a stigmatizing group A — can be recovered, while the privacy of every individual respondent is protected statistically. Since then, different variations of the RRT have been proposed in statistics, see [CM88] for a survey. These different variations provide, for example, smaller variance, smaller privacy breaches, optimality under different definitions of privacy, and ability to answer polychotomous questions. Next we will give a short overview of three types of RRT.

RRT-W. In Wagner's original method (RRT-W), the respondents provide a truthful answer to the question "Do you belong to a stigmatizing group A?" with a certain fixed and publicly known probability $p_{\mathsf{ct}} > 1/2$. With probability $1 - p_{\mathsf{ct}}$ they lie — i.e., answer the opposite question. Define π_A to be the true proportion of the population that belongs to A (or whose *type* is $t = 1$). Let p_{yes} be the proportion of "yes" responses in the poll. Clearly, in RRT-W the *a priori* probability of getting a "yes" response is $p_{\mathsf{yes}} = p_{\mathsf{ct}} \cdot \pi_A + (1 - p_{\mathsf{ct}})(1 - \pi_A)$. In the case of N players, L of which answer "yes", an UE of p_{yes} is $\widehat{p_{\mathsf{yes}}} = L/N$, the sample proportion of "yes" answers. From this, one can simply compute the unbiased estimator of π_A. This equals $\widehat{\pi_A} = \frac{\widehat{p_{\mathsf{yes}}}-(1-p_{\mathsf{ct}})}{2p_{\mathsf{ct}}-1} = \frac{p_{\mathsf{ct}}-1}{2p_{\mathsf{ct}}-1} + \frac{L}{N} \cdot \frac{1}{(2p_{\mathsf{ct}}-1)}$. Similarly, the variance $\mathrm{var}(\widehat{\pi_A})$ and its UE can be computed.

RRT-IQ. An alternative RRT is the *innocuous question method* (RRT-IQ), first analyzed in [GASH69]. When using RRT-IQ, the respondent answers the sensitive question with a probability p_{ct}, while with probability $1 - p_{\mathsf{ct}}$ to an unrelated and innocuous question, such as "Flip a coin. Did you get tails?". The RRT-IQ achieves the same goals as RRT-W but with less variance [CM88], which makes it more suitable for practical polling. Many other RRT-IQs are known, including some with unknown estimate of the the proportion of the population belonging to the innocuous group.

PRRT. The RRTs for dichotomous polling (where the answer is yes or no) can be generalized to *polychotomous RRT* (PRRT) where the respondent can belong to one of the m mutually exclusive groups $A_1, \ldots, A_m$, some of which are stigmatizing. A typical sensitive question of this kind is "When did you have your first child?", with answers "1 — while not married", "2 — within 9 months after the wedding" and "3 — more than 9 months after the wedding". In many cultures, the answer 1 is stigmatizing, the answer 3 is innocuous, while the answer 2 is somewhere inbetween. The interviewer wants to know an UE for the proportion π_i of people who belong to the group A_i, $i \in [1, m]$. There are many possible PRRTs [CM88, Chapter 3]. One of the simplest is the following technique PRRT-BD by Bourke and Dalenius [BD76]: first fix the probabilities p_{ct} and $p_1, \ldots, p_m$, such that $p_{\mathsf{ct}} + \sum_{i \in [1,m]} p_i = 1$. A respondent either reveals her true type $t \in [1, m]$ with probability p_{ct}, or answers $i \in [1, m]$ with probability p_i. To recover an UE of $\boldsymbol{\pi} := (\pi_1, \ldots, \pi_m)^T$, define $\boldsymbol{p} := (p_1, \ldots, p_m)^T$ and $\boldsymbol{p}_{\mathsf{ans}} = (p_{\mathsf{ans}_1}, \ldots, p_{\mathsf{ans}_m})^T$, where p_{ans_i} is the proportion of people who answer i. Then $\boldsymbol{p}_{\mathsf{ans}} = p_{\mathsf{ct}} \cdot \boldsymbol{\pi} + \boldsymbol{p}$, and hence $\widehat{\boldsymbol{\pi}} = p_{\mathsf{ct}}^{-1} \cdot (\widehat{\boldsymbol{p}_{\mathsf{ans}}} - \boldsymbol{p})$.

3 Related Cryptographic Work.

In [KANG99], Kikuchi et al. propose techniques with similar goals as ours. Seemingly unaware of the previous work on RRT, the authors reinvent this notion, and propose a protocol for performing the data exchange. However, their protocol is considerably less efficient than ours. Also, it does not offer strong security in our sense. This vulnerability to cheating makes their protocol unsuitable for their main application (voting), as well as polls where respondents may wish to bias their answer. Our protocols can be used in their framework.

Our work has a relation to work on biased coin flipping, where heads must come out with probability $p_{\mathsf{ct}} = \ell/n$. In our case, the coin can be biased by the first participant in several ways, where the choice of the distribution encodes the opinion of the respondent to the poll. More concretely, consider a coin where one outcome (say, 1) corresponds to "yes", and the other (0) to "no". Let us assume that the respondent should give his correct opinion with 75% probability. Then, if his opinion is "yes", the coin will have bias 0.75, while it will have bias 0.25 if his opinion is "no". However, our technique is not merely a generalization of biased coin flipping, as we also want our protocols to implement privacy. This is an issue that is not important in the context of ordinary biased coin flipping.

In order to guarantee that responses are made according to the intended distribution, we introduce a "blinding" requirement: we need our protocols to be constructed such that they do not leak the response to the respondent — at least not until the response has been delivered to the interviewer. From a bird's eye's view, this makes our protocols similar to those in [JY96], in which a party proves either language membership or language non-membership to a verifier, but without being able to determine which one. However, the similarities between our protocols and those in [JY96] do not run much deeper than that.

In contrast, there is a much closer relationship between our protocols and protocols for oblivious transfer [Rab81,EGL85]. While our goals are orthogonal to those of oblivious transfer, the techniques are hauntingly similar. In particular, one of our CRRT protocols uses a protocol for oblivious transfer as a building block. While in principle *any* such protocol can be used, it is clear that the properties of the building block will be inherited by the main protocol. Therefore, in order to provide unconditional guarantees of privacy for the respondents, we use a *verifiable* variant of the information theoretic protocol for oblivious transfer, namely that proposed by Naor and Pinkas [NP01b]. (An efficient protocol that offers computational security for the sender was proposed by Tzeng [Tze02].)

Cryptographic randomized response techniques are also related to oblivious function evaluation [Gol02], where one party has data μ, while another party needs to compute $f(\mu)$, without getting to know any additional information on μ, while the first party will not get to know f. Cryptographic RRTs can be seen as protocols for oblivious function evaluation of some specific *randomized* functions f.

Furthermore, our work is related to the work on Private Information Retrieval (PIR) — and even to privacy-preserving data-mining — in that the goal of our interviewer is to retrieve some element from the respondent, without the latter learning what was retrieved. More specifically, if some ℓ out of n elements represent the respondent's opinion, and the remaining $n - \ell$ elements represent the opposite opinion, then the interviewer will learn the respondent's opinion with probability ℓ/n if he retrieves a random element. Of course, in order to guarantee the interviewer that the elements are correctly formed, additional mechanisms are required.

In privacy-preserving data-mining a related data randomization approach has been proposed [AS00]: namely, the users input their data to the central database (e.g., a loyal customer inputs the name of the product he bought), and the database maintainer needs to do some statistical analysis on the database. However, the maintainer should not be able to recover individual items. Database randomization in the case when the maintainer is limited to the SUM function corresponds exactly to the RRT. For the same reasons as in the RRT, one should not be able to bias the data. Our protocols are also applicable in the privacy-preserving data-mining and hopefully even in the case when more elaborated randomizations [ESAG02] are applied.

4 Cryptographic Building Blocks

Assume that p is a large prime, and $q, q \mid (p-1)$, is another prime. Then $\mathbb{Z}_p$ has a unique subgroup G of order q. Let g and h be two generators of G, such that nobody knows their mutual discrete logarithms $\log_g h$ and $\log_h g$. We let k be the security parameter, in our setting we can take $k = q$. The key K consists of public parameters, $K := (g; h)$.

Pedersen's Commitment Scheme. In this scheme [Ped91], a message $\mu \in \mathbb{Z}_q$ is committed by drawing a random $\rho \leftarrow_R \mathbb{Z}_q$, and setting $\mathsf{C}_K(\mu;\rho) := g^\mu h^\rho$. The commitment can be opened by sending μ and ρ to the verifier. This scheme is *homomorphic*, i.e., $\mathsf{C}_K(\mu;\rho)\mathsf{C}_K(\mu';\rho') = \mathsf{C}_K(\mu+\mu';\rho+\rho')$. Since it is also perfectly hiding and computationally binding, it can be used as a building block in efficient zero-knowledge arguments, such as protocols for arguing the knowledge of plaintext μ.

Variant of Naor-Pinkas 1-out-of-n Oblivious Transfer. The oblivious transfer (OT) protocol by Naor and Pinkas [NP01b] guarantees information-theoretic privacy for the sender $\mathcal{R}$, and computational privacy for the chooser $\mathcal{I}$. Assume the sender $\mathcal{R}$ has a vector $\mu = (\mu_1, \dots, \mu_n) \in M^n$ for some set $M \subseteq \mathbb{Z}_q$. The chooser $\mathcal{I}$ has made a choice $\sigma \in [1, n]$. The Naor-Pinkas protocol works as follows:

1. $\mathcal{I}$ generates random $a, b \leftarrow \mathbb{Z}_q$ and sends $(A, B, C) \leftarrow (g^a, g^b, g^{ab-\sigma+1})$ to $\mathcal{R}$.
2. $\mathcal{R}$ performs the following, for $i \in [1, n]$: Generate random (r_i, s_i). Compute $w_i \leftarrow g^{r_i} A^{s_i}$, compute an encryption y_i of μ_i using $v_i \leftarrow B^{r_i}(C \cdot g^{i-1})^{s_i}$ as the key. Send (w_i, y_i) to $\mathcal{I}$.
3. $\mathcal{I}$ computes $w_\sigma^b (= v_\sigma)$ and decrypts y_σ using v_σ as the key, obtaining μ_σ.

(Both $\mathcal{R}$ and $\mathcal{I}$ halt if any received transcript is not correctly formatted.) Note that $w_i = g^{r_i + a s_i}$, while $v_i = B^{r_i}(C \cdot g^{i-1})^{s_i} = w_i^b \cdot g^{(i-\sigma)s_i}$. Thus, $v_\sigma = w_\sigma^b$, while for $i \neq \sigma$, v_i is a random element of G. Thus, in the third step $\mathcal{R}$ recovers v_σ, while obtaining no information about v_i for $i \neq \sigma$.

The Naor and Pinkas [NP01b] paper does not specify the encryption method, mentioning only that the encryption scheme must be semantically secure. We propose to use Pedersen's commitment scheme instead of an encryption scheme. Herein, we use $K = (g; h)$ as the parameters of the commitment scheme, and use v_i instead of r_i as the random coin, producing a commitment $y_i := \mathsf{C}_K(\mu_i; v_i)$. We denote this version of Naor-Pinkas protocol, where y_i is defined as $y_i = C_K(\mu_i, v_i)$, by $\binom{1}{n}$-$\mathsf{OT}_K(\mu;\sigma)$. (The full protocol is presented in Appendix A.)

The idea behind this unconventional trick is that as the result, the sender can argue in zero-knowledge for all $i \in [1, n]$ that the values μ_i satisfy some required conditions. (We call such an OT protocol *verifiable*.) The chooser cannot decrypt y_i without knowing v_i, and thus he cannot guess the value of μ_i for $i \neq \sigma$ (with probability higher than $|M|^2/q$, as we will show in Appendix A), even if he knows that μ_i is chosen from a fixed two-element set. (This constitutes the security of OT protocol in the *left-or-right* sense. See Appendix A.) On the other hand, $\mathcal{I}$ can "decrypt" y_σ with the "key" v_σ, given that the possible message space M is small enough for the exhaustive search on the set $\{g^x : x \in M\}$ to be practical. In the case of dichotomous RRT, $M = \{0, 1\}$.

Noninteractive Zero-Knowledge Arguments. We will use zero-knowledge arguments (and not proofs) of knowledge in our protocol, since they are (at the very least) statistically hiding and computationally convincing. This property is important in a setting where a verifier must not be able to extract additional information even if he is given infinite time.

Our first protocol uses only two very standard statistical zero-knowledge arguments. (The arguments for the second protocol are described in appendices.) The first one is an argument that a given value y_i (Pedersen-)commits to a Boolean value $\mu_i \in \{0, 1\}$. One can use standard disjunctive proofs [CDS94] for this. We denote the (possibly parallelized) argument that this holds for $i \in [1, n]$ by $\mathsf{AKEncBool}(y_1, \ldots, y_n)$. The second argument of knowledge, $\mathsf{AKLin}(y_1, \ldots, y_{n+1}; a, b)$, is an argument that the prover knows some set of values μ_i, for which y_i is a commitment of μ_i, and such that $\sum_{i \leq n} \mu_i + a\mu_{n+1} = b$. This argument of knowledge can be constructed from Pedersen's commitment scheme by computing $y \leftarrow \prod_{i \leq n} y_i \cdot y_{n+1}^a$ and then arguing that the result y is a commitment to b. Note that such an argument of knowledge is secure only when accompanied by zero-knowledge arguments of knowledge of the values μ_i; for this purpose, we employ $\mathsf{AKEncBool}(y_1, \ldots, y_{n+1})$ as described above.

5 Security Definitions

In this section, we will give the definition of a weakly and strongly secure cryptographic RRT (CRRT). The security definitions will be in accordance with the ones in secure two-party computation [Gol02]. We will also explain why these requirements are relevant in the case of CRRT.

Assume we have a concrete variant of RRT, like RRT-W or RRT-IQ. Let Φ_p be the function that implements the desired functionality. For example, in the case of RRT-W, $\Phi_{p_{\mathsf{ct}}}(x)$ is a randomized function that with probability p_{ct} returns x, and with probability $1 - p_{\mathsf{ct}}$ returns $1 - x$. The ideal-world CRRT protocol, has three parties, the interviewer $\mathcal{I}$, the respondent $\mathcal{R}$, and the trusted third party $\mathcal{T}$. $\mathcal{R}$ has her type, $t_{\mathcal{R}}$ as her private input, while $\mathcal{I}$ has no private input. Then, $\mathcal{R}$ communicates $t_{\mathcal{R}}$ to $\mathcal{T}$, who selects the value $r_{\mathcal{R}} \leftarrow \Phi_{p_{\mathsf{ct}}}(t_{\mathcal{R}})$ and sends $r_{\mathcal{R}}$ to $\mathcal{I}$. After that, the private output of $\mathcal{I}$ will be $\Phi_{p_{\mathsf{ct}}}(t_{\mathcal{R}})$, while $\mathcal{R}$ will have no private output. It is required that at the end of the protocol, the participants will have no information about the private inputs and outputs of their partners, except for what can be deduced from their own private inputs and outputs. In particular, $\mathcal{I}$ (resp. $\mathcal{R}$) has no information about the value of $t_{\mathcal{R}}$ (resp. $r_{\mathcal{R}}$), except what they can deduce from their private inputs and outputs.

In an ideal world, exactly the next three types of attacks are possible [Gol02, Section 2.1.2]: a party can (a) refuse to participate in the protocol; (b) substitute his private input to the trusted third party with a different value; or (c) abort the protocol prematurely. In our case, the attack (c) is irrelevant, since $\mathcal{R}$ has no output. (Attack (c) models the case when the first party halts the protocol after receiving his private output but before the second party has enough information to compute her output.) Therefore, in an ideal-world RRT protocol, we cannot protect against a participant, who (a) refuses to participate in polling (*non-participation attack*) or (b) claims that her type is $1 - t_{\mathcal{R}}$, where $t_{\mathcal{R}}$ is her real type (*absolute denial attack*). No other attacks should be possible. Note that neither (a) nor (b) is traditionally considered an attack in the context of polling or voting. The argument here is game-theoretic, and the solutions must be proposed by mechanism design, instead of cryptography: namely, a non-manipulable mechanism (e.g., the algorithm with which the election winner is determined from all the collected votes) must be designed so that answering against one's true type (or non-participation) would not give more beneficial results to the respondent than the truthful answer.

On the other hand, as we stated, no other attacks should be allowed. This requirement is very strict, so we will explain why it is necessary in the RRT's context. Clearly, one must protect the privacy of $\mathcal{R}$, since this is the primarily goal of a RRT. It is also necessary to protect the privacy of $\mathcal{I}$, although the reason here is more subtle. Namely, if $\mathcal{R}$ obtains any additional information about $r_{\mathcal{R}}$ before the end of the protocol (for example, if she suspects that $r_{\mathcal{R}} \neq t_{\mathcal{R}}$), she might halt the protocol. Such a behavior by a malicious respondent might cause a bias in the poll, as already explained. (Halting the protocol while having no information on $r_{\mathcal{R}}$ is equivalent to the non-participation attack.) The third requirement on the protocol, of course, is that $\mathcal{I}$ either halts or receives $\Phi_{p_{\mathrm{ct}}}(x)$, where x is the input submitted by the $\mathcal{R}$.

In a real-world implementation, we want to replace $\mathcal{T}$ by a cryptographic protocol $\Pi = (\mathcal{R}, \mathcal{I})$ between $\mathcal{R}$ and $\mathcal{I}$. This protocol $(\mathcal{R}, \mathcal{I})$ is assumed to be "indistinguishable" from the ideal-world protocol, that is, with a high probability, it should be secure against all attacks that do not involve attacks (a) or (b). "Secure" means that the privacy of $\mathcal{R}$ (resp. $\mathcal{I}$) must be protected, if $\mathcal{R}$ (resp. $\mathcal{I}$) follows the protocol, and that $\mathcal{I}$ either halts, or receives the value $\Phi_{p_{\mathrm{ct}}}(x)$, where x was the submitted value of $\mathcal{R}$. The security of the respondent should be information-theoretical, while the security of interviewer can be computational. That is, a secure CRRT-W protocol must have the next three properties (here, k is the security parameter):

Privacy of Respondent: Let $\mathcal{I}^*$ be an algorithm. After the end of the protocol execution $(\mathcal{R}, \mathcal{I}^*)$, $\mathcal{I}^*$ will have no more information on $t_{\mathcal{R}}$ than it would have had after the execution of the ideal world protocol. That is, assuming that $\mathsf{view}_{\mathcal{I}^*}$ is his view of the protocol $(\mathcal{R}, \mathcal{I}^*)$, define

$$\mathsf{Adv}_k^{\mathsf{pri-r}}(\mathcal{R}, \mathcal{I}^*) := |\Pr[\mathcal{I}^*(\mathsf{view}_{\mathcal{I}^*}, r_{\mathcal{R}}) = t_{\mathcal{R}}] - \Pr[t_{\mathcal{R}} | r_{\mathcal{R}}]| \enspace ,$$

where the probability is taken over the internal coin tosses of $\mathcal{I}^*$ and $\mathcal{R}$. We say that a CRRT protocol is *privacy-preserving for the respondent*, if $\mathsf{Adv}_k^{\mathsf{pri-r}}(\mathcal{R}, \mathcal{I}^*)$ is negligible (in k) for any unbounded adversary $\mathcal{I}^*$.

Privacy of Interviewer: Let $\mathcal{R}^*$ be an algorithm. Assume that $\mathcal{I}$ halts when $\mathcal{R}^*$ halts. After the end of the protocol execution $(\mathcal{R}^*, \mathcal{I})$, $\mathcal{R}^*$ will have no more information on $t_{\mathcal{R}}$ than it would have had after the execution of the ideal world protocol. That is, assuming that $\mathsf{view}_{\mathcal{R}^*}$ is her view of the protocol $(\mathcal{I}, \mathcal{R}^*)$, define

$$\mathsf{Adv}_k^{\mathsf{pri-i}}(\mathcal{R}^*, \mathcal{I}) := |\Pr[\mathcal{R}^*(\mathsf{view}_{\mathcal{R}^*}, t_{\mathcal{R}}) = r_{\mathcal{R}}] - \Pr[\mathcal{R}^*(t_{\mathcal{R}}) = r_{\mathcal{R}}]| \enspace ,$$

where the probability is taken over the internal coin tosses of $\mathcal{R}^*$ and $\mathcal{I}$. We say that a CRRT protocol is *privacy-preserving for the interviewer*, if for any adversary $\mathcal{R}^*$, if $\mathsf{Adv}_k^{\mathsf{pri-i}}(\mathcal{R}^*, \mathcal{I}) \leq \varepsilon$ and $\mathcal{R}^*$ takes τ steps of computation then $\varepsilon\tau$ is negligible (in k).

Correctness: Let $\mathcal{R}^*(x)$ be an algorithm with private input x to the protocol $(\mathcal{R}^*, \mathcal{I})$. Assume that $\mathcal{I}$ halts when $\mathcal{R}^*$ halts. We require that at the end of the protocol execution $(\mathcal{R}^*, \mathcal{I})$, $\mathcal{I}$ will either halt, or otherwise receive $\Phi_{p_{\mathsf{ct}}}(x)$ with high probability. That is, assuming that $\mathsf{view}_{\mathcal{I}}$ is $\mathcal{I}$'s view of the protocol $(\mathcal{R}^*, \mathcal{I})$, define

$$\mathsf{Adv}_k^{\mathsf{crct}}(\mathcal{R}^*, \mathcal{I}) := 1 - \Pr[\mathcal{I}(\mathsf{view}_{\mathcal{I}}) = \Phi_{p_{\mathsf{ct}}}(x) | \mathcal{I} \text{ does not halt}] \enspace ,$$

where the probability is taken over the internal coin tosses of $\mathcal{I}$ and $\mathcal{R}^*$. We say that a CRRT protocol is *correct*, if for any adversary $\mathcal{R}^*$, if $\mathsf{Adv}_{\mathcal{I}}^{\mathsf{crct}}(\mathcal{R}^*) = \varepsilon$ and $\mathcal{R}^*$ takes up to t steps of computation then $\varepsilon\tau$ is negligible (in k).

We call a cryptographic RRT (CRRT) protocol *weakly secure* if it is privacy-preserving for the respondent and correct. We call CRRT protocol *(strongly) secure* if it is weakly secure and it is privacy-preserving for the interviewer. While a secure CRRT protocol is preferable in many situations, there are settings where a weakly secure CRRT protocol suffices, such as where halting can be easily detected and punished, or means for state recovery prevent modifications between a first and second attempt of executing the protocol.

6 Cryptographic RRT

We will propose three different CRRT-W protocols. In the first two protocols, the common parameters are $p_{\mathsf{ct}} = \ell/n > 1/2$; generators g and h whose mutual discrete logs are unknown (at least by $\mathcal{R}$); and $K = (g; h)$. $\mathcal{R}$ has private input $t = t_{\mathcal{R}}$, and $\mathcal{I}$'s private output is $r_{\mathcal{R}}$.

CRRT Protocol Based on Oblivious Transfer. Our first implementation of RRT-W is described in Protocol 1. The arguments of knowledge can be efficiently constructed, see Sect. 4. Here, we can use $\mathsf{AKLin}(y_1, \ldots, y_{n+1}; 2\ell - n; \ell)$ since $\sum_{i \leq n} \mu_i + (2\ell - n)\mu_{n+1} = \ell$ independently of the value of t. All the steps in this protocol must be authenticated.

PRECOMPUTATION STEP:

1. $\mathcal{R}$ prepares n random bits $\mu_i \in \{0, 1\}$ for $i \in [1, n]$, such that $\sum \mu_i = \ell$ if $t = 1$ and $\sum \mu_i = n - \ell$ if $t = 0$. Additionally, she sets $\mu_{n+1} \leftarrow 1 - t$.
2. $\mathcal{I}$ chooses an index $\sigma \in [1, n]$.

INTERACTIVE STEP:

1. $\mathcal{I}$ and $\mathcal{R}$ follow $\binom{1}{n}$-$\mathsf{OT}_K(g^{\mu_1}, \ldots, g^{\mu_n}; \sigma)$. $\mathcal{I}$ obtains g^{μ_σ}, and computes μ_σ from that.
2. $\mathcal{R}$ sends to $\mathcal{I}$ noninteractive zero-knowledge arguments $\mathsf{AKEncBool}(y_1, \ldots, y_{n+1})$, and $\mathsf{AKLin}(y_1, \ldots, y_{n+1}; 2\ell - n; \ell)$.
3. $\mathcal{I}$ verifies the arguments, and halts if the verification fails.

Protocol 1: A secure CRRT-W protocol based on oblivious transfer

If we take the number of bits that must be committed as the efficiency measure (communication complexity of the protocol), then our protocol has complexity $O(n)$.

In the polling application, one can most probably assume that $n \leq 5$. The security proofs of this protocol follow directly from the properties of underlying primitives. As a direct corollary from Theorem 2, we get that Protocol 1 is privacy-preserving for respondent ($\mathsf{Adv}_k^{\mathsf{pri-r}}(\mathcal{R}, \mathcal{I}^*) \leq 2/q + O(1/q)$, where the constant comes in from the use of statistically-hiding zero-knowledge arguments). It is privacy preserving for interviewer, given the Decisional Diffie-Hellman (DDH) assumption. The correctness of this protocol follows from the properties of the zero-knowledge arguments used under the DDH assumption.

In a simplified weakly secure protocol based on the same idea, $\mathcal{R}$ commits to all μ_i by computing and publishing $y_i \leftarrow \mathsf{C}_K(\mu_i; \rho_i)$. Next, $\mathcal{R}$ argues that $\mathsf{AKEncBool}(y_1, \dots, y_{n+1})$, and $\mathsf{AKLin}(y_1, \dots, y_{n+1}; 2\ell - n; \ell)$. After that, $\mathcal{I}$ sends σ to $\mathcal{R}$, who then reveals μ_σ and ρ_σ. Upon obtaining these, $\mathcal{I}$ verifies the correctness of the previous corresponding commitment, outputting μ_σ.

CRRT from Coin-Flipping. Protocol 2 depicts a secure CRRT-W protocol with communication complexity $\Theta(d \log_2 n)$, where $d := \lceil 1/(1 - p_{\mathsf{ct}}) \rceil$, and $p_{\mathsf{ct}} = \ell/n$ as previously. While in the common RRT application one can usually assume that n is relatively small, this second protocol is useful in some specific game-theoretic applications where for the best outcome, p_{ct} must have a very specific value. The idea behind this protocol is that at least one of the integers $\mu + \nu + i\ell \mod n$ must be in interval $[0, \ell - 1]$, and at least one of them must be in interval $[\ell, n-1]$. Hence, $\mathcal{I}$ gets necessary proofs for both the 0 and the 1 answer, which is sufficient for his goal. For his choice to be accepted, he must accompany the corresponding r with $\mathcal{R}$-s signature on his commitment on σ.

PRECOMPUTATION STEP:

1. $\mathcal{R}$ chooses a random $\mu \leftarrow_R [0, n-1]$.
2. $\mathcal{I}$ chooses random $\nu \leftarrow_R [0, n-1]$ and $\sigma \leftarrow_R [0, d-1]$.

INTERACTIVE STEP:

1. $\mathcal{R}$ commits to t and μ, and sends the commitments to $\mathcal{I}$.
2. $\mathcal{I}$ commits to σ, by setting $y \leftarrow \mathsf{C}_K(\sigma; \rho)$ for some random ρ. He sends ν and y to $\mathcal{R}$, together with a zero-knowledge argument that y is a commitment of some $i \in [0, d-1]$.
3. $\mathcal{R}$ verifies the argument. She computes values μ'_i, for $i \in [0, d-1]$, such that $\mu'_i = t \iff (\mu + \nu + i\ell \mod n) < \ell$. She signs y, and sends her signature together with $\{\mu'_i\}$ and the next zero-knowledge argument for every $i \in [0, d-1]$: $[\mu'_i = t \iff (\mu + \nu + i\ell \mod n) < \ell]$.
4. After that, $\mathcal{I}$ sets $r_{\mathcal{R}} \leftarrow \mu'_\sigma$. He will accompany this with $\mathcal{R}$-s signature on the commitment, so that both $\mathcal{R}$ and third parties can verify it.

Protocol 2: A secure CRRT-W protocol based on coin-flipping

A weakly secure version of this protocol is especially efficient. There, one should set $d \leftarrow 1$, and omit the steps in Protocol 2 that depend on σ being greater than 1. (E.g., there is no need to commit to σ anymore.) Thus, such a protocol would have communication complexity $\Theta(\log_2 n)$. Now, $p_{\mathsf{ct}} > 1/2$ (otherwise one could just do a bit-flip on the answers), and hence $d > 2$. On the other hand, the privacy of respondents

PRECOMPUTATION STEP:

1. $\mathcal{I}$ chooses random $u_0 \leftarrow_R [0,1]$, $u_1 \leftarrow_R [0,1]$. He generates quantum states $|\psi_0\rangle = \sqrt{p_{\mathsf{ct}}}|u_0\rangle + \sqrt{1-p_{\mathsf{ct}}}|1-u_0\rangle$, $|\psi_1\rangle = \sqrt{p_{\mathsf{ct}}}|u_1\rangle + \sqrt{1-p_{\mathsf{ct}}}|1-u_1\rangle$.
2. $\mathcal{R}$ chooses a random $i \leftarrow_R [0,1]$.

INTERACTIVE STEP:

1. $\mathcal{I}$ sends $|\psi_0\rangle$ and $|\psi_1\rangle$ to $\mathcal{R}$.
2. $\mathcal{R}$ sends i to $\mathcal{I}$.
3. $\mathcal{I}$ sends u_i to $\mathcal{R}$.
4. $\mathcal{R}$ measures the state $|\psi_i\rangle$ in the basis $|\psi_{u_i}\rangle = \sqrt{p_{\mathsf{ct}}}|u_i\rangle + \sqrt{1-p_{\mathsf{ct}}}|1-u_i\rangle$, $|\psi_{u_i}^{\perp}\rangle = \sqrt{1-p_{\mathsf{ct}}}|u_i\rangle - \sqrt{p_{\mathsf{ct}}}|1-u_i\rangle$ and halts if the result is not $|\psi_{u_i}\rangle$.
5. If the verification is passed, $\mathcal{R}$ performs the transformation $|0\rangle \to |t\rangle$, $|1\rangle \to |1-t\rangle$ on the state $|\psi_{1-i}\rangle$ and sends it back to $\mathcal{I}$.
6. $\mathcal{I}$ measures the state in the basis $|0\rangle, |1\rangle$, gets outcome s. $\mathcal{I}$ outputs $r \leftarrow u_i \oplus s$.

Protocol 3: A quantum CRRT-W protocol.

is in danger if say $p_{\mathsf{ct}} \geq 3/4$. Thus, we may assume that $d \in [3,4]$. Therefore, Protocol 2 will be more communication-efficient than Protocol 1 as soon as $n/\log_2 n > 4 \geq d$, or $n \geq 16$. The weakly secure version will be *always* more communication-efficient.

This protocol is especially efficient if the used commitment scheme is an integer commitment scheme [FO99,DF02]. In this case, to argue that $(\mu + \nu + i\ell \mod n) < \ell$ one only must do the next two simple steps: first, argue that $\mu + \nu + i\ell = z + en$ for some z, e, and then, argue that $z \in [0, \ell - 1]$. This can be done efficiently by using the range proofs from [Bou00,Lip01]. One can also use Pedersen's scheme, but this would result in more complicated arguments.

Quantum-Cryptographic RRT. We also present a *quantum CRRT protocol* (see Protocol 3) that allows for a value p_{ct} that does not have to be a rational number, and which provides a relaxed form of information-theoretic security to *both* parties. While not secure by our previous definitions, it provides meaningfully low bounds on the probabilities of success for a cheater. Namely, (a) if dishonest, $\mathcal{R}$ cannot make his vote count as more than $\sqrt{2}$ votes: if $p_{\mathsf{ct}} = \frac{1}{2} + \varepsilon$, then $p_{\mathsf{adv}} \leq \frac{1}{2} + \sqrt{2}\varepsilon$ (we also show a slightly better bound with a more complicated expression for p_{adv}, cf. Appendix B). (b) if dishonest strategy allows $\mathcal{I}$ to learn t with probability $p_{\mathsf{ct}} + \varepsilon$, it also leads to $\mathcal{I}$ being caught cheating with probability at least $\frac{2p_{\mathsf{ct}}-1}{2}\varepsilon$. This form of security (information-theoretic security with relaxed definitions) is common for quantum protocols for tasks like bit commitment [ATVY00] or coin flipping [Amb01,SR02]. The security guarantees of our quantum protocol compare quite well to ones achieved for those tasks. A desirable property of this quantum protocol is that it can be implemented by using contemporary technology, since it only involves transmitting and measuring single qubits, and no maintaining of coherent multi-qubit states.

To show the main ideas behind quantum protocol, we now show how to analyze a simplified version of protocol 3. The security proof for the full protocol is quite complicates and is given in appendix B. We also refer to appendix B for definitions of quantum states and operations on them.

The simplified version of Protocol 3 is:

1. $\mathcal{I}$ chooses a random $u \leftarrow_R [0,1]$, prepares a quantum bit in the state $|\psi_u\rangle = \sqrt{p_{\mathsf{ct}}}|u\rangle + \sqrt{1-p_{\mathsf{ct}}}|1-u\rangle$ and sends it to $\mathcal{R}$.
2. $\mathcal{R}$ performs a bit flip if her type $t = 1$, and sends the quantum bit back to $\mathcal{I}$.
3. $\mathcal{I}$ measures the state in the computational basis $|0\rangle$, $|1\rangle$, gets answer s. The answer is $r = u \oplus s$.

If both parties are honest, the state returned by respondent is unchanged: $\sqrt{p_{\mathsf{ct}}}|u\rangle + \sqrt{1-p_{\mathsf{ct}}}|1-u\rangle$ if $t = 0$ and $\sqrt{p_{\mathsf{ct}}}|1-u\rangle + \sqrt{1-p_{\mathsf{ct}}}|u\rangle$ if $t = 1$. Measuring this state gives the correct answer with probability $1 - p_{\mathsf{ct}}$. Next, we show that respondent is unable to misuse this protocol.

Theorem 1. *For any respondent's strategy $\mathcal{R}^*$, the probability of honest interviewer $\mathcal{I}$ getting $r = 1$ is between $1-p_{\mathsf{ct}}$ and p_{ct}. Therefore, the previous protocol is both correct and privacy-preserving for the interviewer.*

Proof. We show that the probability of $r = 1$ is at most p_{ct}. The other direction is similar. We first modify the (simplified) protocol by making $\mathcal{R}^*$ to measure the state and send the measured result to $\mathcal{I}$, this does not change the result of the honest protocol since the measurement remains the same. Also, any cheating strategy for $\mathcal{R}^*$ in the original protocol can be used in the new protocol as well. So, it is sufficient to bound the probability of $r = 1$ in the new protocol.

Now, the answer is $r = 1$ if $\mathcal{I}$ sent $|\psi_i\rangle$ and $\mathcal{R}^*$ sends back j, with $i = j$. Thus, we have the setting of Fact 1 (see Appendix B.1). The rest is a calculation: to determine the angle β between $|\psi_0\rangle$ and $|\psi_1\rangle$, it suffices to determine the inner product which is $\sin\beta = 2\sqrt{p_{\mathsf{ct}}(1-p_{\mathsf{ct}})}$. Therefore, $\cos\beta = \sqrt{1-\sin^2\beta} = 2p_{\mathsf{ct}} - 1$ and $\frac{1}{2} + \frac{\cos\beta}{2} = p_{\mathsf{ct}}$. □

On the other hand, when using this simplified version, a dishonest interviewer $\mathcal{I}^*$ can always learn t with probability 1. Namely, it suffices to send the state $|0\rangle$. If $t = 0$, $\mathcal{R}$ sends $|0\rangle$ back unchanged. If $t = 1$, $\mathcal{R}$ applies a bit flip. The state becomes $|1\rangle$. $\mathcal{I}$ can then distinguish $|0\rangle$ from $|1\rangle$ with certainty by a measurement in the computational basis.

Note that this is similar to a classical "protocol", where $\mathcal{I}$ first generates a random u and sends a bit i that is equal to u with probability p_{ct} and $1-u$ with probability $1-p_{\mathsf{ct}}$. $\mathcal{R}$ then flips the bit if $t = 1$ and sends it back unchanged if $t = 0$. The interviewer XORs it with u, getting t with probability p_{ct} and $1 - t$ with probability $1 - p_{\mathsf{ct}}$. In this "protocol", $\mathcal{R}$ can never cheat. However, $\mathcal{I}^*$ can learn t with probability 1 by just remembering i and XORing the answer with i instead of u. In the classical world, this flaw is fatal because $\mathcal{I}$ cannot prove that he has generated i from the correct probability distribution and has not kept a copy of i for himself. In the quantum case, $\mathcal{I}$ can prove to $\mathcal{R}$ that he has correctly prepared the quantum state. Then, we get Protocol 3 with $\mathcal{I}$ sending two states $|\psi_{u_0}\rangle$ and $|\psi_{u_1}\rangle$, one of which is verified and the other is used for transmitting t. (See Appendix B for detailed analysis of this protocol.)

7 Protocols for Other RRTs and Extensions

Protocol for Cryptographic RRT-IQ. Recall that in one version of RRT-IQ, the respondent would reply with his true opinion $t_{\mathcal{R}}$ with a rational probability $p_{\mathsf{ct}} = \ell/n$, while he would otherwise flip a coin and answer whether it came up tails. Like for CRRT-W, it is important to guarantee the use of correct distributions. Protocol 1 can be easily changed to work for this version of RRT-IQ. Instead of n random bits, $\mathcal{R}$ prepares $2n$ random bits μ_i, so that $\sum \mu_i = n + \ell$ if $t_\mathcal{R} = 1$, and $\sum \mu_i = n - \ell$ if $t_\mathcal{R} = 0$. She also prepares a checksum bit $\mu_{2n+1} = 1 - t_\mathcal{R}$. The rest of the protocol is principally the same as in Protocol 1, with n changed to $2n$, and $\mathcal{R}$ arguing that $\mathsf{AKLin}(y_1, \dots, y_{2n+1}; 2\ell; 2n - \ell)$.

Protocol for Cryptographic PRRT-BD. The next protocol is a modification of Protocol 1 as well. Let p_i be such that $p_{\mathsf{ct}} + \sum_{i \in [1,m]} p_i = 1$, and assume that every respondent has a type $t_\mathcal{R} \in [1, m]$. Assume $p_{\mathsf{ct}} = \ell/n$, $p_i = \ell_i/n$ and that $p_i = 0$ if $i \notin [1, m]$. Assume $D \geq \max(\ell, \ell_1, \dots, \ell_m) + 1$. The respondent prepares n numbers D^{μ_i}, such that $\sharp\{i : \mu_i = t_\mathcal{R}\} = \ell_{t_\mathcal{R}} + \ell$, and $\sharp\{i : \mu_i = j\} = \ell_j$, if $j \neq t_\mathcal{R}$. Then the interviewer and respondent will execute a variant of OT with choice σ, during which the interviewer only gets to know the value μ_σ. Then the respondent argues that the sum of all commitments is a commitment to the value $\sum \ell_i D^{\mu_i} + \ell D^j$, for some $j \in [1, m]$, by using range-proofs in exponents [LAN02]. (A more efficient proof methodology is available when D is a prime [LAN02], given that one uses an integer commitment scheme.) Additionally, she argues that every single commitment corresponds to a value D^i for $i \in [1, m]$, also using range-proofs of exponents [LAN02]. After the OT step, the interviewer gets g^{μ_σ}, and recovers μ_σ from it efficiently. (Note that $m \leq 10$ is typical in the context of polling.)

Extensions to Hierarchies of Interviewers. One can consider a hierarchy of interviewers, reporting to some central authority. If there is a trust relationship between these two types of parties, no changes to our protocol would be required. However, if the central authority would like to be able to avoid having to trust interviewers, the following modifications could be performed. First, each respondent would have to authenticate the transcript he generates, whether with a standard signature scheme, a group signature scheme, etc. Second, and in order to prevent collusions between interviewers and respondents, the interviewers must not be allowed to know the choice σ made in a particular interview. Thus, the triple (A, B, C) normally generated by the interviewer during the Naor-Pinkas OT protocol would instead have to be generated by the central authority, and kept secret by the same. More efficient versions of *proxy* OT satisfying our other requirements are beneficial for this application [NP01a].

Acknowledgments

We would like to thank Jouni K. Seppänen for introducing us to the RRT, and for fruitful discussions on the topic. We would like to thank Benny Pinkas for comments.

References

[Amb01] Andris Ambainis. A New Protocol and Lower Bounds for Quantum Coin Flipping. In *Proceedings of the Thirty-Third Annual ACM Symposium on the Theory of Computing*, pages 134–142, Heraklion, Crete, Greece, July 6–8 2001. ACM Press.

[AS00] Rakesh Agrawal and Ramakrishnan Srikant. Privacy-Preserving Data Mining. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 439–450, Dallas, TX, USA, May 2000.

[ATVY00] Dorit Aharonov, Amnon Ta-Shma, Umesh V. Vazirani, and Andrew Chi-Chih Yao. Quantum Bit Escrow. In *Proceedings of the Thirty-Second Annual ACM Symposium on the Theory of Computing*, pages 705–714, Portland, Oregon, USA, May 21–23 2000. ACM Press.

[BD76] Patrick D. Bourke and Tore Dalenius. Some New Ideas in the Realm of Randomized Inquiries. *International Statistics Review*, 44:219–221, 1976.

[Bou00] Fabrice Boudot. Efficient Proofs that a Committed Number Lies in an Interval. In Bart Preneel, editor, *Advances in Cryptology — EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 431–444, Bruges, Belgium, May 14–18 2000. Springer-Verlag. ISBN 3-540-67517-5.

[CDS94] Ronald Cramer, Ivan Damgård, and Berry Schoenmakers. Proofs of Partial Knowledge and Simplified Design of Witness Hiding Protocols. In Yvo G. Desmedt, editor, *Advances in Cryptology—CRYPTO '94*, volume 839 of *Lecture Notes in Computer Science*, pages 174–187, Santa Barbara, USA, August 21–25 1994. Springer-Verlag.

[CM88] Arijit Chaudhuri and Rahul Mukerjee. *Randomized Response: Theory and Techniques*, volume 95 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., 1988. ISBN: 0824777859.

[DF02] Ivan Damgård and Eiichiro Fujisaki. An Integer Commitment Scheme Based on Groups with Hidden Order. In Yuliang Zheng, editor, *Advances on Cryptology — ASIACRYPT 2002*, volume 2501 of *Lecture Notes in Computer Science*, pages 125–142, Queenstown, New Zealand, December 1–5 2002. Springer-Verlag.

[EGL85] Shimon Even, Oded Goldreich, and Abraham Lempel. A Randomized Protocol for Signing Contracts. *Communications of the ACM*, 28(6):637–647, June 1985.

[ESAG02] Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke. Privacy Preserving Mining of Association Rules. In *Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining*, pages 217–228, Edmonton, Canada, July23–26 2002. ACM.

[FO99] Eiichiro Fujisaki and Tatsuaki Okamoto. Statistical Zero-Knowledge Protocols to Prove Modular Polynomial Relations. *IEICE Transaction of Fundamentals of Electronic Communications and Computer Science*, E82-A(1):81–92, January 1999.

[GASH69] Bernard G. Greenberg, Abdel-Latif A. Abul-Ela, Walt R. Simmons, and Daniel G. Horvitz. The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association*, 64(326):520–539, June 1969.

[Gol02] Oded Goldreich. Secure Multi-Party Computation. Final (Incomplete) Draft, October 27 2002.

[JY96] Markus Jakobsson and Moti Yung. Proving Without Knowing: On Oblivious, Agnostic and Blindfolded Provers. In Neal Koblitz, editor, *Advances in Cryptology—CRYPTO '96*, volume 1109 of *Lecture Notes in Computer Science*, pages 186–200, Santa Barbara, California, USA, August 18–22 1996. Springer-Verlag.

[KANG99] Hiroaki Kikuchi, Jin Akiyama, Gisaku Nakamura, and Howard Gobioff. Stochastic Voting Protocol To Protect Voters Privacy. In *1999 IEEE Workshop on Internet Applications*, pages 103–111, July 26–27 1999.

[LAN02] Helger Lipmaa, N. Asokan, and Valtteri Niemi. Secure Vickrey Auctions without Threshold Trust. In Matt Blaze, editor, *Financial Cryptography — Sixth International Conference*, volume 2357 of *Lecture Notes in Computer Science*, pages 87–101, Southhampton Beach, Bermuda, March 11–14 2002. Springer-Verlag.

[Lip01] Helger Lipmaa. Statistical Zero-Knowledge Proofs from Diophantine Equations. Cryptology ePrint Archive, Report 2001/086, November 20 2001. http://eprint.iacr.org/.

[NC00] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[NP01a] Moni Naor and Benny Pinkas. Distributed Oblivious Transfer. In Tatsuaki Okamoto, editor, *Advances on Cryptology — ASIACRYPT 2001*, volume 1976 of *Lecture Notes in Computer Science*, pages 205–219, Kyoto, Japan, 3–7 December 2001. Springer-Verlag. ISBN 3-540-41404-5.

[NP01b] Moni Naor and Benny Pinkas. Efficient Oblivious Transfer Protocols. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 448–457, Washington, DC, USA, January 7–9 2001.

[Ped91] Torben P. Pedersen. Non-Interactive and Information-Theoretic Secure Verifiable Secret Sharing. In J. Feigenbaum, editor, *Advances in Cryptology—CRYPTO '91*, volume 576 of *Lecture Notes in Computer Science*, pages 129–140, Santa Barbara, California, USA, August 11–15 1991. Springer-Verlag, 1992.

[Rab81] Michael Rabin. How to exchange secrets by oblivious transfer. Technical Report TR-81, Aiken Computation Laboratory, Harvard University, 1981.

[SR02] Robert Spekkens and Terry Rudolph. A Quantum Protocol for Cheat-Sensitive Weak Coin Flipping. *Physical Review Letters*, 89:227901, 2002.

[Tze02] Wen-Guey Tzeng. Efficient 1-Out-n Oblivious Transfer Schemes. In David Naccache and Pascal Paillier, editors, *Public Key Cryptography '2002*, volume 2274 of *Lecture Notes in Computer Science*, pages 159–171, Paris, France, February12–14 2002. Springer-Verlag.

[War65] Stanley L. Warner. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309):63–69, March 1965.

A Security of Modified Oblivious Transfer Protocol

From our oblivious transfer protocol $\binom{1}{n}\text{-}\mathsf{OT}_K(\mu;\sigma)$ we will require that it must be secure in the next sense. The attack scenario consists of the following game. The chooser $\mathcal{I}^*$ chooses σ and two different vectors, $\mu[1] = (\mu[1]_1, \dots, \mu[1]_n) \in M^n$ and $\mu[2] = (\mu[1]_1, \dots, \mu[1]_n) \in M^n$, such that $\mu[1]_\sigma = \mu[2]_\sigma$. Denote an $\mathcal{I}^*$ that has made such choices by $\mathcal{I}^*(\mu[1], \mu[2])$. He submits both tuples to the responder, who flips a fair coin $b \leftarrow_R [1, 2]$. After that, the chooser and the responder execute the protocol $\binom{1}{n}\text{-}\mathsf{OT}_K(\mu[b];\sigma)$. After receiving $\mu[b]_\sigma$, $\mathcal{I}^*$ guesses the value of b. Let $\mathsf{Adv}_k^{\mathsf{lor}}(\mathcal{I}^*, \mathcal{R})$ be the probability that $\mathcal{I}^*$ guesses the correct b, where probability is taken over the internal coin tosses of $\mathcal{I}^*$ and $\mathcal{R}$. We say that the oblivious transfer protocol is ε-secure in the *left-or-right* sense, if for any unbounded algorithm $\mathcal{I}^*$, $\mathsf{Adv}_k^{\mathsf{lor}}(\mathcal{I}^*, \mathcal{R}) \le \varepsilon$.

Recall that the proposed variant of the Naor-Pinkas protocol works as follows:

1. $\mathcal{I}$ generates random $a, b \leftarrow \mathbb{Z}_q$ and sends $(A, B, C) \leftarrow (g^a, g^b, g^{ab-\sigma+1})$ to $\mathcal{R}$.
2. $\mathcal{R}$ performs the following, for $i \in [1, n]$: Generate random (r_i, s_i). Compute $w_i \leftarrow g^{r_i} A^{s_i}$, compute an encryption $y_i \leftarrow g^{\mu_i} h^{v_i}$, where $v_i \leftarrow B^{r_i} (C \cdot g^{i-1})^{s_i}$. Send (w_i, y_i) to $\mathcal{I}$.
3. $\mathcal{I}$ computes $w_\sigma^b (= v_\sigma)$ and recovers $g^{\mu_\sigma} \leftarrow y_\sigma / h^{w_\sigma^b}$.

Theorem 2. *Let* $\binom{1}{n}$-$\mathsf{OT}_K(\cdot;\cdot)$ *be the described oblivious transfer protocol. (a) If a malicious* $\mathcal{R}^*$ *can guess the value of* σ *with advantage* ε*, then he can solve the Decisional Diffie Hellman (DDH) problem with the same probability and in approximately the same time. (v) This protocol is* $(m-d)(m-1)/q \leq m(m-1)/q$*-secure in the left-or-right sense, where* $d := q \mod m$ *and* $m := |M|$.

Proof (Sketch.). (a) Assume that $\mathcal{R}^*$ can guess σ with probability ε, given her view $(A, B, C) = (g^a, g^b, g^{ab-\sigma+1})$. But then she can solve the DDH problem (given (g^a, g^b, g^c) for random a and b, decide whether $c = ab$ or not) with probability ε: given an input (g^a, g^b, g^c), she just computes such a σ, for which $c = ab - \sigma + 1$. After that, she only has to check whether $\sigma = 1$ or not.

(b) W.l.o.g., assume that $\sigma = 1$. Define $\nu[j]$ to be a vector, for which which $\nu[j]_i = \mu[1]_i$ if $i > j$, and $\nu[j]_i = \mu[2]_i$ if $i \leq j$. Thus $\nu[1] = \mu[1]$ (since $\mu[1]_1 = \mu[2]_1$), while $\nu[n] = \mu[2]$, and for all j, $\nu[j-1]$ and $\nu[j]$ differ only in the jth element $\nu[j]_j \neq \nu[j+1]_j$. Thus, our goal is to show that $\mathcal{I}^*(\nu[1], \nu[n]) \leq m(m-1)/q$. For this we will prove that $\mathcal{I}^*(\nu[j-1], \nu[j]) \leq (m-d)/q \leq m/q$ for every $j \in [2, n]$ and then use the triangle equality to establish that $\mathsf{Adv}_k^{\mathsf{lor}}(\mathcal{I}^*(\mu[1], \mu[2]), \mathcal{R}) \leq \sum_{i=2^n} \mathsf{Adv}_k^{\mathsf{lor}}(\mathcal{I}^*(\nu[j-1], \nu[j]), \mathcal{R})$.

Now, fix a $j \in [2, n]$. After the protocol execution $(\mathcal{I}^*, \mathcal{R})$, $\mathcal{R}$ flipping the coin $b \leftarrow_R [1, 2]$, $\mathcal{I}^*$ must guess the value of b, based on his private input $(\mu[1], \mu[2])$, his private output $\mu[b]_1$, and the protocol view. Since $\nu[j-1]_i = \nu[j]_i$ for $i \neq j$, this is equivalent to guessing whether $\nu[j-2+b]_j = \nu[j-1]_j$ or $\nu[j-2+b]_j = \nu[j]_b$. Clearly, his success is maximized here when $\nu[j-1]_j \neq \nu[j]_j$. Next, $\mathcal{I}^*$'s view consists of $(A, B, C; \{(w_j, y_j)\})$, where $(w_j, y_j) \leftarrow (g^{r_j} A^{s_j}, g^{\mu_j} h^{B^{r_j} \cdot (C \cdot g^{j-1})^{s_j}})$ for A, B and C chosen by himself. Since $\mathcal{I}^*$ is unbounded, he can find the value of $\alpha \neq 0$, and therefore he knows that $(w_j, y_j) = (g^{r_j + a s_j}, g^{\mu_j + \alpha B^{r_j} (C \cdot g^{j-1})^{s_j}})$. Since r_j and s_j are randomly chosen by a honest $\mathcal{R}$, then the elements w_j look completely random to $\mathcal{I}^*$, and do not help in guessing the value of μ_j. He also cannot use any information in (w_j, y_j), $j \neq j$, since these values do not depend on μ_j.

Thus, to guess the value $\nu[j-2+b]_j$, he must find a bias in the value $aB^{r_j}(Cg^{j-1})^{s_j} = \alpha g^{b r_j + (ab+j-\sigma) s_j} \mod q$. Note that $x := \alpha g^{b r_j + (ab+j-\sigma) s_j}$ is a random element of $\mathbb{Z}_p^*$ due to the choice of r_j and s_j, unless $b = ab + j - \sigma = 0$. The latter will automatically hold if $i = \sigma$, but only with a negligible probability otherwise. Thus, we can assume that x is chosen randomly from $\mathbb{Z}_p^*$. Guessing $\mu_j \in \mathbb{Z}_m$ from y_j is equivalent to guessing the value $(x \mod q) \mod m$. Denote $e := \lfloor q/m \rfloor$. Since $q \mid (p-1)$ then $x \mod q$ is a random element of $\mathbb{Z}_q$, and $\sharp\{x : x \mod q \mod m = j\} \in e + c$, where $c \in \{0, 1\}$ is 1 iff $j < d$. Thus the best strategy of $\mathcal{I}^*$ is to guess that x is equivalent to some element $j < d$, and equivalently, that $\nu[j-2+b]_j \mod m \geq d$. He will achieve this by choosing exactly one of the two element $\nu[j-1]_b$ and $\nu[j-1]_b$ to have residue modulo m that is less than d. Then he will succeed with probability $e/q + 1/q$ which gives him an advantage $e/q + 1/q - 1/m = (m-d)/q \leq m/q$ over random guessing the bit b. □

Security in the left-or-right sense is both necessary and sufficient for our purposes. Namely, in the RRT-W protocol (Sect. 6), the interviewer $\mathcal{I}^*$ knows that the input is — up to the permutation of indices — one of the two values. For small n, the number of permutations is small, and thus with a high probability $\mathcal{I}^*$ can guess that μ is one of the two, known for him, Boolean vectors. Without security in the left-or-right sense, he would be able to guess which of the two vectors is currently used, and thus to find the type of the respondent. On the other hand, if the oblivious transfer protocol is secure in the left-or-right sense, $\mathcal{I}^*$ cannot predict the Hamming weight $w_h(\mu) = \sharp\{i : \mu_1 = 1\}$ of $\mathcal{R}$'s input.

B Detailed Quantum CRRT

B.1 Background on Quantum Information

In this section, we describe the basic notions of quantum information needed to understand the quantum protocol and the analysis of its simplified version in section 6.

For a more detailed introduction to quantum information, we refer to book by Nielsen and Chuang [NC00]. A *qubit* is the basic unit of quantum information, similar to a bit in the conventional (classical) computing. A qubit has two basis states that are denoted by $|0\rangle$ and $|1\rangle$. A general state of a qubit is $\alpha|0\rangle + \beta|1\rangle$, with α, β being complex numbers with $|\alpha|^2 + |\beta|^2 = 1$.

We can perform two types of operations on quantum bits: unitary transformations and measurements. The simplest *measurement* of of a qubit $\alpha|0\rangle + \beta|1\rangle$ is in the *computational basis* that gives the result 0 with probability $|\alpha|^2$ and 1 with probability $|\beta|^2$. The state of the qubit then becomes $|0\rangle$ or $|1\rangle$. Therefore, repeating the measurement gives the same outcome. As long as we only consider this one type of measurement, the state $\alpha|0\rangle + \beta|1\rangle$ behaves similarly to a probabilistic state that has been prepared as 0 with probability $|\alpha|^2$ and 1 with probability $|\beta|^2$. This analogy disappears, though, when we consider other transformations. *A unitary transformation* is a linear transformation on the two-dimensional space of all $\alpha|0\rangle + \beta|1\rangle$ that preserves the vector norm. Two examples of unitary transformations are the identity $I(\alpha|0\rangle + \beta|1\rangle) = \alpha|0\rangle + \beta|1\rangle$ and the bit flip $X(\alpha|0\rangle + \beta|1\rangle) = \alpha|1\rangle + \beta|0\rangle$. A *general von Neumann measurement* on a qubit $|\Psi\rangle$ is specified by two orthogonal vectors $|\Phi_0\rangle$ and $|\Phi_1\rangle$. The outcome is either 0 or 1; the probability of outcome i is equal to the squared inner product of $|\Psi\rangle$ and $|\Phi_i\rangle$. The state of the qubit becomes $|\Phi_i\rangle$. This measurement can be reduced to the measurement in the computational basis as follows. We take a unitary U that maps $|\Phi_0\rangle$ to $|0\rangle$ and $|\Phi_1\rangle$ to $|1\rangle$. We apply U to the state $|\Psi\rangle$ that we want to measure. Then, we measure the resulting state in the computational basis and apply U^{-1}.

Distinguishability. Assume someone prepares two states $|\Phi_0\rangle$ and $|\Phi_1\rangle$, flips a fair coin $i \leftarrow_R [0, 1]$, and sends $|\Phi_i\rangle$ it to us. We would like to guess i by measuring the state. We measure our success by the probability that our guess $j \in \{0, 1\}$ coincides with i. If $|\Phi_0\rangle$ and $|\Phi_1\rangle$ are orthogonal, a von Neumann measurement in $|\Phi_0\rangle$, $|\Phi_1\rangle$ basis tells i with certainty. For non-orthogonal states, no measurement gives i with certainty.

Fact 1 *[NC00] The maximum success probability with what we can distinguish $|\Phi_0\rangle$ from $|\Phi_1\rangle$ is $\frac{1}{2} + \frac{\sin\beta}{2}$, β being the angle between $|\Phi_0\rangle$ and $|\Phi_1\rangle$.*

The above definitions are sufficient to understand the protocol and the analysis of simplified version in section 6. For the full security proof, more advanced notions like *density matrices* are needed, which are described in Sect. B.2.

B.2 Density Matrices

To prove the security of protocol 3, we need the more advanced formalism of *density matrices*. We interpret $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ as a column vector $(\alpha, \beta)^T$. Let $\langle\psi|$ denote a row vector $(\alpha^*\beta^*)$, with $*$ being the complex conjugation operator. Then, the density matrix of $|\psi\rangle$ is

$$|\psi\rangle\langle\psi| = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}(\alpha^*\beta^*) = \begin{pmatrix}\alpha\alpha^* & \alpha\beta^*\\ \beta\alpha^* & \beta\beta^*\end{pmatrix} \quad .$$

Next, assume that we generate a classical random variable that is i with probability p_i and then prepare a quantum state $|\psi_i\rangle$ dependent on i. This creates a *mixed* quantum state. It can be also described by a density matrix $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$. If we measure a mixed state with a density matrix ρ in a basis $|\Phi_0\rangle$, $|\Phi_1\rangle$, the probability of getting outcome i is $\langle\Phi_i|\rho|\Phi_i\rangle$ (i.e., we multiply the density matrix with the row vector $\langle\Phi_i|$ on the left and the column vector $|\Phi_i\rangle$ on the right and get a number which is the probability). The following is a counterpart of Fact 1 for mixed states.

Fact 2 *[NC00] The maximum success probability with which we can distinguish ρ_0 from ρ_1 is $\frac{1}{2} + \frac{\|\rho_0-\rho_1\|_t}{4}$, where $\|A\|_t$ is the trace norm of A (the trace (sum of diagonal entries) of matrix $\sqrt{A^\top A}$).*

B.3 Security Proofs for Protocol 3

Security against Malicious Interviewer.

Theorem 3. *If a strategy for dishonest $\mathcal{I}^*$ leads to being caught with probability at most ε, $\mathcal{I}^*$ can learn r correctly with probability at most $p_{\mathsf{ct}} + \frac{2}{2p_{\mathsf{ct}}-1}\varepsilon$.*

The security of this type (cheating is possible but not without risk of being detected) is common to many quantum protocols, for example quantum bit commitment [ATVY00] or coin flipping [SR02]. We note that our security guarantee is stronger than one achieved in [ATVY00]. Namely, in the bit commitment protocol of [ATVY00], a dishonest party can successfully cheat with probability ε so that the probability of being detected is just $O(\varepsilon^2)$.

Proof (Theorem 3). Assume that we are given a strategy for dishonest $\mathcal{I}^*$. First, notice that if we reverse the roles of $|0\rangle$ and $|1\rangle$ everywhere in this strategy, both the probability of passing the test and the probability of learning t correctly remain the same. Therefore, we can assume that the protocol is symmetric w.r.t. switching $|0\rangle$ and $|1\rangle$.

Consider the state of the first quantum bit sent by $\mathcal{I}^*$. In the general case, $\mathcal{I}^*$ can send probabilistic combinations of various quantum states. Therefore, the first quantum bit can be in a mixed state with some density matrix

$$\rho = \begin{pmatrix} a & \alpha+\beta i\\ \alpha-\beta i & b\end{pmatrix} \quad .$$

Since the strategy is symmetric w.r.t. switching $|0\rangle$ and $|1\rangle$, ρ must be also symmetric in the same sense, implying that $a = b = 1/2$ and $\beta = 0$. Thus,

$$\rho = \begin{pmatrix} 1/2 & \alpha \\ \alpha & 1/2 \end{pmatrix} \quad .$$

If $\mathcal{I}$ is honest, $\alpha = \sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})}$. Theorem 3 follows from the following two lemmas.

Lemma 1. *The probability of $\mathcal{I}^*$ failing the test if the first quantum bit is chosen for verification is at least* $(\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} - \alpha)\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})}$.

Lemma 2. *The probability of $\mathcal{I}^*$ learning t correctly if the first bit is used for protocol and the second bit used for verification is at most* $\frac{1}{2} + \frac{\sqrt{1-4\alpha^2}}{2}$.

We will for a moment assume the validity of these theorems (their proofs are given slightly later), and will now continue with the proof of the theorem.

Let ε be the probability with which $\mathcal{I}^*$ allows to be caught. By Lemma 1, $(\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} - \alpha)\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} \leq \varepsilon$. Therefore, $\alpha \geq \sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} - \frac{\varepsilon}{\sqrt{p_{\mathsf{ct}}(1-p_{\mathsf{ct}})}}$. By substituting that into Lemma 2, we get $\frac{1}{2} + \frac{\sqrt{1-4\alpha^2}}{2} \leq \frac{1}{2} + \frac{\sqrt{1-4p_{\mathsf{ct}}(1-p_{\mathsf{ct}})+8\varepsilon}}{2}$. If $\mathcal{I}$ is honest, the probability that $r = t$ is $\frac{1}{2} + \frac{\sqrt{1-4p_{\mathsf{ct}}(1-p_{\mathsf{ct}})}}{2}$. The extra advantage gained by $\mathcal{I}^*$ is at most $\frac{\sqrt{1-4p_{\mathsf{ct}}(1-p_{\mathsf{ct}})+8\varepsilon}}{2} - \frac{\sqrt{1-4p_{\mathsf{ct}}(1-p_{\mathsf{ct}})}}{2} \leq \frac{2\varepsilon}{2p_{\mathsf{ct}}-1}$ (assuming that $p_{\mathsf{ct}} > 1/2$). □

Proof (Lemma 1). When the first bit is chosen for verification, $\mathcal{I}^*$ either claims that it is $|\psi_0\rangle$ or $|\psi_1\rangle$. By symmetry, the probability of each of those is 1/2. We partition $\rho = \frac{1}{2}\rho_0 + \frac{1}{2}\rho_1$, with ρ_i being the part for which $\mathcal{I}^*$ claims that the state is $|\psi_i\rangle$. Let

$$\rho_0 = \begin{pmatrix} a' & \alpha' \\ \alpha' & b' \end{pmatrix} \quad .$$

By symmetry, ρ_1 should be the same with $|0\rangle$ and $|1\rangle$ reversed:

$$\rho_1 = \begin{pmatrix} b' & \alpha' \\ \alpha' & a' \end{pmatrix} \quad .$$

Since $\rho = \frac{1}{2}\rho_0 + \frac{1}{2}\rho_1$, $a' + b' = 1$ and $\alpha' = \alpha$. Therefore, we have

$$\rho_0 = \begin{pmatrix} a' & \alpha \\ \alpha & 1 - a' \end{pmatrix} \quad .$$

The probability of this state passing verification as $|\psi_0\rangle$ is

$$\begin{aligned} \langle \Psi_0 | \rho_0 | \Psi_0 \rangle = & \left(\sqrt{p_{\mathsf{ct}}}\ \sqrt{1 - p_{\mathsf{ct}}}\right) \begin{pmatrix} a' & \alpha \\ \alpha & 1 - a' \end{pmatrix} \begin{pmatrix} \sqrt{p_{\mathsf{ct}}} \\ \sqrt{1 - p_{\mathsf{ct}}} \end{pmatrix} \\ = & a' p_{\mathsf{ct}} + (1 - a')(1 - p_{\mathsf{ct}}) + 2\alpha\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} \\ \leq & p_{\mathsf{ct}}^2 + (1 - p_{\mathsf{ct}})^2 + 2\alpha\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} \\ = & (p_{\mathsf{ct}} + (1 - p_{\mathsf{ct}}))^2 - (\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} - \alpha)\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} \\ = & 1 - (\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} - \alpha)\sqrt{p_{\mathsf{ct}}(1 - p_{\mathsf{ct}})} \quad . \end{aligned}$$

□

Proof (Lemma 2). We assume that the second qubit has been prepared perfectly and its verification always succeeds. (If $\mathcal{I}^*$ cheated in preparing the second qubit as well, this only decreases the probability of success for $\mathcal{I}^*$ and the claim that we prove remains valid.)

After the test is passed on the second qubit, $\mathcal{R}$ has the first qubit in the mixed state ρ. The mixed state ρ is the same as one obtained by taking $\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ with probability 2α and $|0\rangle$, $|1\rangle$ with probabilities $\frac{1}{2} - \alpha$ each. Therefore, the joint state of $\mathcal{I}^*$ and $\mathcal{R}$ is equivalent to $|\psi_{(\mathcal{R},\mathcal{I}^*)}\rangle = \sqrt{\frac{1}{2} - \alpha}|0\rangle_{\mathcal{I}^*}|0\rangle_{\mathcal{R}} + \sqrt{\frac{1}{2} - \alpha}|1\rangle_{\mathcal{I}^*}|1\rangle_{\mathcal{R}} + \sqrt{2\alpha}|2\rangle_{\mathcal{I}^*}(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle)_{\mathcal{R}}$. If $\mathcal{R}$'s secret bit $t = 0$, he just sends his part back to $\mathcal{I}^*$. After that, $\mathcal{I}^*$ possesses the entire state $|\psi_{(\mathcal{R},\mathcal{I}^*)}\rangle$. Otherwise, $\mathcal{R}$ flips the qubit before sending back and $\mathcal{I}^*$ gets $|\psi'_{(\mathcal{R},\mathcal{I}^*)}\rangle = \sqrt{\frac{1}{2} - \alpha}|0\rangle_{\mathcal{I}^*}|1\rangle_{\mathcal{R}} + \sqrt{\frac{1}{2} - \alpha}|1\rangle_{\mathcal{I}^*}|0\rangle_{\mathcal{R}} + \sqrt{2\alpha}|2\rangle_{\mathcal{I}^*}(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle)_{\mathcal{R}}$. Now, the question is how well can $\mathcal{I}^*$ distinguish these two states. By Fact 1, the best probability with which he can get t is $\frac{1}{2} + \frac{\sin\beta}{2} = \frac{1}{2} + \frac{\sqrt{1-\cos^2\beta}}{2}$ where β is the angle between the two states. $\cos\beta$ is equal to the inner product of $|\psi_{(\mathcal{R},\mathcal{I}^*)}\rangle$ and $|\psi'_{(\mathcal{R},\mathcal{I}^*)}\rangle$ which is 2α (because the first two components of $|\psi_{(\mathcal{R},\mathcal{I}^*)}\rangle$ are orthogonal to the first two components of $|\psi'_{(\mathcal{R},\mathcal{I}^*)}\rangle$ but the third component is equal). □

Security against Malicious Respondent.

Theorem 4. *Let* $p_{\mathsf{ct}} < \frac{1}{2} + \frac{\sqrt{3}}{4} = 0.933\ldots$. *If* $\mathcal{I}$ *is honest,* $\mathcal{R}^*$ *cannot achieve* $t = 0$ *(or* $t = 1$*) with probability more than* $p_{\mathsf{adv}} \leq \frac{1}{2} + \sqrt{\sqrt{4p_{\mathsf{ct}} - 4p_{\mathsf{ct}}^2} - (4p_{\mathsf{ct}} - 4p_{\mathsf{ct}}^2)}$.

The probability p_{adv} remains less than 1 for all $p_{\mathsf{ct}} < 0.933\ldots$. Thus, our protocol offers nontrivial security guarantees for all $p_{\mathsf{ct}} < 0.933\ldots$. Since the expression for p_{adv} is quite complicated, we also present a simple but less precise bound. Let $p_{\mathsf{ct}} = \frac{1}{2} + \epsilon$. Then, $p_{\mathsf{adv}} \leq \frac{1}{2} + \sqrt{2}\epsilon$. Informally, this means that no $\mathcal{R}^*$ can make his vote count as more than $\sqrt{2}$ votes. This gives a non-trivial bound on p_{adv} for $p_{\mathsf{ct}} < \frac{1}{2} + \frac{1}{2\sqrt{2}} = 0.853\ldots$.

If $0.853\ldots \leq p_{\mathsf{ct}} \leq 0.933\ldots$, then $\frac{1}{2} + \sqrt{2}\epsilon \geq 1$ but $p_{\mathsf{adv}} < 1$ which can be seen by evaluating the expression of theorem 4 directly.

Proof. There are four possible states that a responder can receive from an honest $\mathcal{I}$: $|\psi_0\rangle|\psi_0\rangle$, $|\psi_0\rangle|\psi_1\rangle$, $|\psi_1\rangle|\psi_0\rangle$, $|\psi_1\rangle|\psi_1\rangle$. An honest responder then randomly requests to verify one of two quantum bits. A dishonest $\mathcal{R}^*$ can measure the state and then decide to verify one of two bits based on the result of the measurement so that his chances of guessing the other state are maximized. Without loss of generality, $\mathcal{R}^*$'s measurement has two outcomes: 0 and 1 and the index i that is sent back to $\mathcal{I}$ is equal to the outcome of the measurement. Then, we have

$$|\psi_{u_0}\psi_{u_1}\rangle = \alpha_{u_0u_1}|0\rangle|\psi'_{u_0u_1}\rangle + \beta_{u_0u_1}|1\rangle|\psi''_{u_0u_1}\rangle,$$

where the first qubit is the one being measured and $|\psi'_{u_0u_1}\rangle$ ($|\psi''_{u_0u_1}\rangle$) is the rest of the quantum state that remains with $\mathcal{I}$ after the measurement. By symmetry, we can assume that $\alpha_{u_0u_1} = \beta_{u_0u_1} = \frac{1}{\sqrt{2}}$.

Similarly to the simplified protocol in Sect. 6, the probability of $\mathcal{R}^*$ fixing $r = 0$ (or $r = 1$) is equal to the probability that he correctly guesses u_{1-i}. We bound this probability. For brevity, assume that $\mathcal{R}^*$ has requested u_1 from $\mathcal{I}$ and received $u_1 = 0$. Then, if $u_0 = 0$, his remaining state is $|\psi'_{00}\rangle$ and, if $u_0 = 1$, his remaining state is $|\psi'_{10}\rangle$. The probability with which he can guess u_0 is, by Fact 1, at most $p_{\mathsf{adv}} = \frac{1}{2} + \frac{\sin\beta'}{2}$ where β' is the angle between $|\psi'_{00}\rangle$ and $|\psi'_{10}\rangle$. Remember that, by analysis of Sect. 6, the probability of $r = t$ in the honest case is described by similar expression $p_{\mathsf{ct}} = \frac{1}{2} + \frac{\sin\beta}{2}$ where β is the angle between $|\psi_0\rangle$ and $|\psi_1\rangle$.

Next, we express β' by β. Remember that $\langle\psi|\psi'\rangle$ denotes the inner product between $|\psi\rangle$ and $|\psi'\rangle$. The inner product $\langle\psi_0|\psi_1\rangle$ is equal to $\cos\beta$. The inner product between $|\psi_0\rangle|\psi_0\rangle$ and $|\psi_1\rangle|\psi_0\rangle$ is the same $\cos\beta$ because the second qubit is in the same state in both cases. This inner product is also equal to $\frac{1}{2}\langle\psi'_{00}|\psi'_{10}\rangle + \frac{1}{2}\langle\psi''_{00}|\psi''_{10}\rangle$. The first part is $\cos\beta'$, the second part is at most 1. Therefore, $\frac{1}{2}(\cos\beta' + 1) \geq \cos\beta$ and $\cos\beta' \geq 2\cos\beta - 1$. We have $\sin\beta' = \sqrt{1 - \cos^2\beta'} \leq \sqrt{4(\cos\beta - \cos^2\beta)}$ and $p_{\mathsf{adv}} \leq \frac{1}{2} + \frac{\sin\beta'}{2} \leq \frac{1}{2} + \sqrt{\cos\beta - \cos^2\beta}$. Remember that in the honest protocol, the probability that $r = t$ is $p_{\mathsf{ct}} = \frac{1}{2} + \frac{\sin\beta}{2}$. Therefore, $\sin\beta = 2p_{\mathsf{ct}} - 1$, $\cos\beta = \sqrt{1 - \sin^2\beta} = \sqrt{4p_{\mathsf{ct}} - 4p_{\mathsf{ct}}^2}$ and, by substituting this into $p_{\mathsf{adv}} \leq \frac{1}{2} + \sqrt{\cos\beta - \cos^2\beta}$, we get the theorem. □

To show the $p_{\mathsf{adv}} \leq \frac{1}{2} + \sqrt{2}\epsilon$ upper bound, it suffices to show $\sqrt{\cos\beta - \cos^2\beta} \leq \sqrt{2}\epsilon$. Since $\epsilon = \frac{\sin\beta}{2}$, this follows from

$$\frac{\sqrt{\cos\beta - \cos^2\beta}}{(\sin\beta)/2} = \frac{2\sqrt{\cos\beta - \cos^2\beta}}{\sqrt{1 - \cos^2\beta}} = \frac{2\sqrt{\cos\beta}}{\sqrt{1 + \cos\beta}} \leq \frac{2\sqrt{\cos\beta}}{\sqrt{2\cos\beta}} = \sqrt{2}$$