Randomized voting

Gregory S. Warrington
warrings@wfu.edu
April 25, 2007

1. Introduction

In theory, an election is a simple process: People express their preferences. The winner is the candidate preferred by the majority.

In practice, an election is a amalgam of many choices and procedures, each of which can affect the outcome. These aspects run from the gamut from "Who is eligible to vote?" to "How does a voter physically express her preferences?" to "What mechanisms protect against fraud?" Mathematics is the ideal language for exploring some of the thorniest of such issues.

There are two fundamental stages to voting. In the first stage, voters express their preferences. In the second stage, these cumulative preferences are processed in order to determine the winner(s). Historically, much of the mathematical literature on voting has been concerned with the second stage (see, for example, [10]). In these studies, mathematicians have considered the framework that arises from the basic axioms of voting theory. Their work has shown that there are significant limitations to the robustness of voting systems.

More recently, mathematicians and computer scientists have focused their attentions on the first stage. The interest here arises from the two competing criteria of secrecy and verifiability. Roughly, secrecy means that there should be no way for anyone to tell how a specific individual voted. Secrecy is desirable as it guards against coercion of voters. Verifiability amounts to a mechanism for checking that the election was not subject to any fraud or errors. Either one of secrecy or verifiability is easy to design into an election protocol; the difficulty arises in trying to incorporate both criteria.

The "secret ballot voting methods" widely used today protect secrecy at the expense of verifiability. (Unfortunately, the increasingly popular absentee ballots effectively provide neither.) While there are many safeguards in place to ensure accurate results, the individual voter is asked to trust that these safeguards are working and that the administering body is trustworthy.

Several protocols have been proposed that utilize cryptography as a means of allowing secrecy and verifiability simultaneously. Some of the more recent proposals include those of Neff [5–7], the visual cryptography of Chaum [1], the Prêt-à-Voter scheme of Chaum and Ryan [2], and the Scratch-and-Vote enhancement of Adida and Rivest as described in [8]. For simplicity, we conflate these schemes into a "cryptologic method" we now describe.

Each voter gets a receipt, albeit with the votes (more or less) listed only in encrypted form. These receipts are publicly posted. Since they are encrypted, secrecy is not compromised. Each voter can confirm that her (encrypted) vote made it out of the polling station. The administering body then decrypts these receipts so that the results of the election can be determined. A division-of-powers approach along with clever cryptologic protocols enables this all to occur while preserving both secrecy and verifiability. Within certain constraints, the security of the protocol can be proven mathematically. The voter does not need to trust the administering body. However, she must still trust cryptologic constructions and mathematical proofs that will be incomprehensible to the average voter.

These cryptologic protocols are creative and robust systems; we do not claim or attempt to improve on them here. But while fraud in these systems is harder to perpetuate and errors can be caught, it is unclear that the average voter will trust the process any more than she trusts our current system. Here we pursue an alternate approach that offers a measure of secrecy and verifiability without relying on encrypted votes (though we do utilize standard digital signatures).

In this paper we suggest a simple approach that lets *some* voters directly verify that their votes have been accurately recorded and that they are appropriately contributing to the final tallies. This is done while minimizing coercion and bribery. Third-party auditors play a crucial role in verifying votes and guarding against vote stuffing. No sophisticated network algorithms, encryption techniques, or blind faith in an administering body is asked.

In our "randomized method," most voters leave the polling station with a receipt. The receipt might correspond to someone else's vote. These receipts can be compared to the contents of a plaintext list on a

publicly readable bulletin board. The voters do not need to trust that a proper linkage exists between a list of encrypted votes and a list of plaintext votes. What we give up is the universality in which all voters get to check their own votes.

We can crudely compare the three methods we have introduced by summarizing the degree to which they incorporate verifiability and protect/ensure secrecy.

Verifiability

- *Secret ballot method:* A voter is unable to directly verify that her vote is included in the final tally.
- *Cryptologic method:* A voter can indirectly verify that her vote is included in the final tally by accepting mathematical proofs and that certain algorithms have been carried out properly.
- *Randomized method: Some* voters can directly verify that their votes are included in the final tally.

Secrecy

- *Secret ballot method:* The voter cannot prove (or be forced to prove) how she voted.
- *Cryptologic method:* The voter cannot prove (or be forced to prove) how she voted.
- *Randomized method:* Some voters can present evidence to having voted a certain way, but the evidence cannot be construed as a proof of having done so.

2. Sketch of proposed voting process

The essence of our protocol is that each voter (with some exceptions) exits the polling place with a voting receipt, but that there are even odds as to whether or not the receipt was generated in response to her vote, or to that of an earlier voter. In light of these odds, any voter with a receipt can plausibly deny that it corresponds to her vote. This randomization allows plaintext receipts while minimizing the threat of coercion.

We follow the terminology of Karlof, Sastry and Wagner [4] whenever possible. As is common in describing voting protocols such as these, we do not address logistical matters such as authenticating voters or preventing recording in voting booths.

We now describe the voting protocol in more detail. However, there are several details we defer discussing in order to avoid obfuscating the main components of the protocol.

(1) **Election initialization**

Some number $n > 0$ of independent auditors is registered for each polling location. Each auditor is given a place in the polling station, along with a box for collecting receipts of the votes cast, and a can for mixing receipts. A mechanism for choosing a random number between 0 and n is placed between the voting booths and the auditors. The value generated by this mechanism should be public. For simplicity, we think of this mechanism as an $(n+1)$-sided die.

In order to guard against forgery, certain important data, *though not the votes themselves*, will be digitally signed using public key cryptography. These signatures act like traditional longhand signatures, though they are mathematically difficult to forge. The essence of the approach is to sign data using your *private key* (which no one else knows). Other parties can check, by decrypting your signature with your *public key*, that you, and not some impostor, created the signature. Public and private key pairs are chosen for the administering body and each of the voting machines.

Receipt-bundles such as those illustrated in Figure 1 are printed prior to the election. Each receipt-bundle is perforated into n sections. We reserve the term "receipt" for any one of these sections. These receipts each have two areas. The first area contains two items 1) a serial number unique to the receipt (constant on the receipts of a given receipt-bundle) and 2) a digital signature of the serial number which is encrypted using the administrator's private key. The second area contains room for a printout of a voter's choices, for a second digital signature (described below), and an ID number for the machine used. These areas will be referred to as the SN- and V-areas, respectively. At the beginning of the election, the SN-areas should all be filled out; the V-areas remain blank until a vote is cast by an individual voter. Receipts such as those in Figure 1 are *initialized*, those such as in Figure 2 are *complete*.

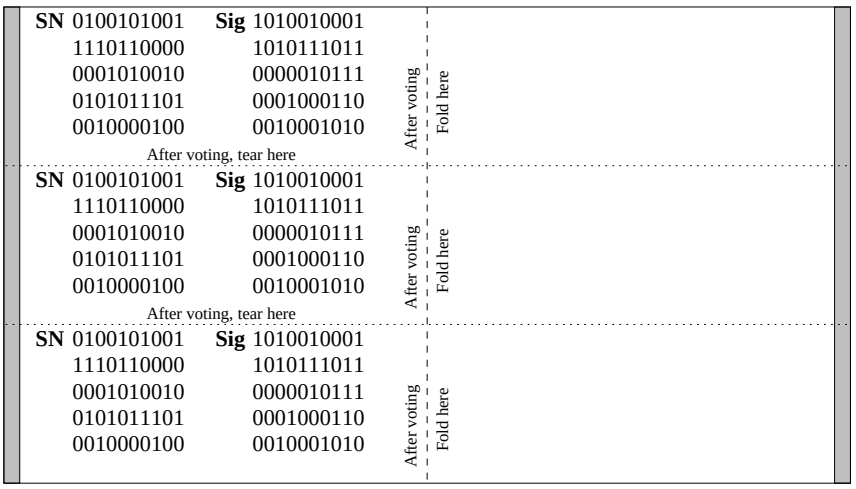

FIGURE 1. Sample receipt-bundle with $n = 3$ containing initialized, but uncompleted, receipts.

Each receipt, when separated from the rest of the receipt-bundle, can be folded in half such that neither the SN-area nor the V-area is visible. Sticky areas along the left and right edges of the receipt-bundle can be used to keep a receipt closed once it is folded.

A bin of initialized, randomly-ordered receipt-bundles is made available to those voters confirmed as eligible to vote.

(2) **Ballot preparation**

Each voter takes a receipt-bundle, enters a privacy booth and casts a ballot. We leave the physical implementation unspecified, though we do require a voter-verified paper audit trail to be generated. The voter is prompted to insert her receipt-bundle into a machine for printing of her vote and the machine's ID on the V-areas of each receipt. The machine also adds a digital signature utilizing the machine's private key. This signature encodes the serial number of the receipt and the voter's selections. All of the completed receipts on a given receipt-bundle should be identical. The voter is directed to separate the receipts of each receipt-bundle and to fold them shut so that her selections are not visible.

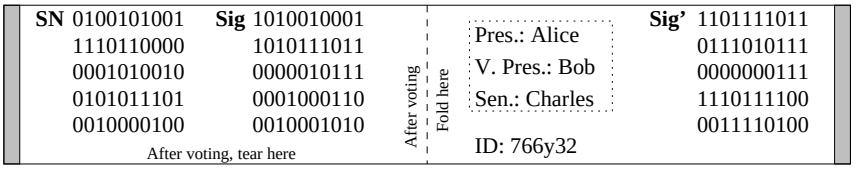

FIGURE 2. Sample completed receipt.

For concreteness, we abbreviate the contents of a complete receipt by the five-tuple (SN, Sig, V, Sig', ID): the serial number SN, the digital signature Sig of SN, the vote V, the digital signature Sig' of the pair SN and V, and the machine identification number, ID.

(3) **Receipt distribution**

Upon exiting the booth, each voter rolls a die to receive a number k, $0 \leq k \leq n$. The voter first distributes one receipt of the receipt-bundle to each auditor *except* auditor k. If $k = 0$, then the voter has no receipts remaining and proceeds to exit the voting station. If $k > 0$, then auditor k picks one receipt at random from his box and puts it, along with the voter's remaining receipt, into the auditor's mixing can. The can is closed and then shaken to randomize the positions of the two receipts inside. The voter retrieves one receipt and the auditor retrieves the other. The auditor puts his chosen receipt in his collection box. The voter exits the polling station with her chosen receipt.

(4) **Tallying:**

After the end of voting, the administering body posts the results on a publicly readable bulletin board. In particular, each line of the posting consists of the data found on one of the completed receipts: (SN, Sig, V, Sig', ID). The data should be formatted such that the votes are easily tallied by anyone with a spreadsheet. Also posted should be the public key of the administrator and the public keys of the various voting machines.

(5) **Verification:**

Different parties have different opportunities to verify the results of the election to different extents and in different ways.

(a) Any person can view the bulletin board and perform three checks. First, using the administrator's public key, he can check that Sig is the correct signature for SN. Confirmation of this means that, unless the administrator's private key has been compromised, the receipt comes from an official receipt-bundle. Second, using the public key corresponding to machine ID, he can check that the vote was printed by the machine claimed. Third, he can tally the votes in the list and compare to the publicly announced totals.

(b) A person with a (valid) receipt can check that the held receipt actually appears in the posted list.

(c) A voter with her own receipt has the added knowledge that *her own* vote is actually being counted. The advantage of this check over that of a voter with someone else's receipt is purely psychological.

(d) The first check an auditor can make is to compare the number of people he recorded to have voted with the total number of posted votes. The second check is to confirm that the receipts he has collected are all distinct and that they all appear on the posted list. This second check guards against the administering body throwing out receipts or generating receipt-bundles with the same serial number. The two checks together guard against "virtual votes" (votes not actually cast by eligible voters) creeping into the publicly posted list of votes. If there were such a virtual vote in the list, since the number of total voters is known, at least one valid receipt would have to have been omitted from the posted list. The probability of this action going undetected depends on several parameters, but can be assumed to be fairly low if the auditors do their jobs; this issue is addressed more in Section 5.

3. A mock election

We illustrate the trajectories of the ballots in a mock election with $T = 6$ voters and $n = 2$ observers. The first two voters are forced to leave without any ballot (i.e., forced to roll $k = 0$). We suppose there are two positions being elected and we indicate a voter's choice as two names separated by a comma. Suppose the completed receipts are as depicted in Table 1. (For simplicity, we use short, and therefore easily memorizable, serial numbers and signatures that would be inappropriate for a real election.)

Voter	SN	Sig	V	Sig'	ID
1	14	7q	Alice,Charles	3e	111
2	11	2u	Bob,Charles	2f	112
3	12	8u	Alice,Charles	7x	112
4	15	4v	Bob,David	2w	111
5	16	9i	Alice,David	4g	112
6	13	1k	Alice,Charles	7y	111

Table 1. All of the receipts from the mock election.

We illustrate in Table 2 what each of the two auditors might have in their boxes after each voter passes through. Note that the first two voters are required to leave the polling station without a receipt in order to seed the collections of each auditor. The third and fourth voters both switch their own receipts with one held by an auditor, but the third voter gets a receipt with selections identical to those on her own receipt. She will not know that a switch has occurred unless she has memorized the serial number or one of the

signatures. The fifth voter rolls a 0 and walks away without any receipt. The sixth voter rolls a 2, but happens to choose her own receipt.

Most recent voter	Die roll	First auditor's box	Second auditor's box	Voter leaves with
1	0	14	14	Nothing
2	0	14,11	14,11	Nothing
3	2	14,11,12	11,12	14
4	1	14,12,15	11,12,15	11
5	2	14,12,15,16	11,12,15,16	Nothing
6	2	14,12,15,16,13	11,12,15,16	13

TABLE 2. Holdings of auditors and voters after randomization steps.

The contents of the public bulletin board should consist of the information in Table 3. Note that there is no link back to individual voters and that the information has been sorted in some predetermined manner.

SN	Sig	V	Sig'	ID
11	2u	Bob,Charles	2f	112
12	8u	Alice,Charles	7x	112
13	1k	Alice,Charles	7y	111
14	7q	Alice,Charles	3e	111
15	4v	Bob,David	2w	111
16	9i	Alice,David	4g	112

TABLE 3. Contents of bulletin board.

The third, fourth and sixth voters can each check that the information from the receipt she holds is accurately listed on the bulletin board. The fourth voter knows for sure that the receipt she holds is not her own. Each auditor can perform similar checks with the receipts he holds.

4. Coercion & bribery analysis

Our protocol is neither perfectly private nor perfectly secret in the sense of [3]. Nonetheless, as mentioned in that article, probabilistic information about a voter's selections can easily be gleaned from demographic data. As such, we do not view the failure of perfect privacy or secrecy as fatal flaws of our approach.

The usual problem with unencrypted receipts is that since each voter retains her own receipt, a voter can be punished or rewarded according to how she voted. This could, of course, influence how a person chooses to vote. In the protocol proposed in Section 2, any voter who retains a receipt can plausibly deny that it is her original receipt. This reduces the threat of coercion. Voters who don't retain receipts cannot be coerced (see part 8 of Section 6). On the other hand, a voter cannot guarantee that the receipt she holds reflects how she voted. This reduces the threat of bribery. While "coercion" (negative incentives) and "bribery" (positive incentives) are somewhat different in nature, we only analyze the susceptibility of our protocol to bribery.

We consider an election between two candidates, A and B, with equal popular support for each. We suppose there is someone buying votes for A. For simplicity, we assume the vote buyer is only able to contact a small percentage of all voters. Each voter falls into one of four possible classes according to which candidate they want to win and for which candidate they end up voting. Certainly those who want A to win will vote for A. (There is no incentive for someone who wants A to win to vote for B.) We obtain a table such as that in Figure 4. The percentage in the second-to-last column is that of the total voting population; the last column indicates the portion of these who actually do get paid. We assume in the following that anyone with a receipt for candidate A goes and gets paid (i.e., irregardless of whether she deserves the money). For simplicity of the analysis, we assume that $k = 0$ is never rolled. This assumption makes buying votes more efficient than it actually is, though the discrepancy grows smaller as the number of observers increases. We now present data for the two scenarios of unethical and ethical populations of people preferring B.

Prefers	Votes for	Keeps own receipt	% of total	Actually paid
A	A	Yes	25	All
A	A	No	25	Half
A	B	Yes	0	N/A
A	B	No	0	N/A

TABLE 4. Voters who want A to win.

Prefers	Votes for	Keeps own receipt	% of total	% paid	% of total	% paid
B	A	Yes	25	All	0	N/A
B	A	No	25	Half	0	N/A
B	B	Yes	0	N/A	25	None
B	B	No	0	N/A	25	Half

TABLE 5. Voters who want B to win. The fourth and fifth columns apply if this subpopulation is unethical and hence ready to change their votes; the sixth and seventh columns apply if this subpopulation is ethical and won't change their votes.

In the scenario with unethical voters, the vote buyer ends up paying 75% of the people, even though only one in every two voters changed her vote. In the scenario with ethical voters, the briber ends up paying 50% of the people, even though no one changed her vote.

5. Analysis of percentage of receipts held by auditors

In this section we consider the probability that none of the receipts from a given receipt-bundle end up in the hands of any of the auditors. Under the assumption that auditors will check the receipts they hold against the data on the bulletin board at a much higher rate than individual voters will, this calculation gives a measure of how hard it is for parties to remove votes from (or introduce invalid votes to) the official tally. The data in Table 6 shows that the removal or addition of votes will be caught with very high probability.

Suppose there are T total voters, n auditors, and that we force a die roll of $k = 0$ for the first a voters. Consider some voter $i \geq a$ who rolls $k = n$. She deposits receipts $R_1, R_2, \ldots, R_{n-1}$ with each of the auditors $1, 2, \ldots, n-1$, respectively. She has a probability of $1/2$ of leaving R_n with auditor n. First we compute the probability that R_1 remains with the first auditor at the end of voting. Each successive voter has a probability of $1/(2n+2)$ of actually exchanging her receipt with one of the receipts held by the first auditor. Also, after j voters, each auditor is expected to hold approximately $a + (j-a)n/(n+1)$ receipts. So the probability that the j-th voter gets R_1 is

$$\frac{1}{2n+2} \cdot \frac{1}{a + (j-a)\frac{n}{n+1}} = \frac{1}{2(a+jn)}.$$

Hence, the probability that R_1 is not removed at this stage is $1 - 1/(2a+2jn)$. It follows that the probability that R_1 is removed by one of the later voters is

$$U = 1 - \prod_{j=i+1}^{T} \left(1 - \frac{1}{2(a+jn)}\right). \tag{1}$$

Raising U to the $(n-1)$-st power and multiplying by $(1+U)/2$ (to account for R_n) gives the probability that none of the receipts $R_1, R_2, \ldots, R_n$ remain with any of the auditors. (Note that receipts deposited by voters who roll $k = 0$ will be slightly more likely to remain in the hands of at least one auditor.) We give some sample probabilities in Table 6. Assuming the auditors are conscientious, we have just computed the odds that a *single* valid vote could be omitted without it being noticed. Successfully substituting for more than one vote is exponentially more difficult.

n	T = 100	T = 500	T = 1000
2	0.3	0.5	0.5
3	0.05	0.1	0.2
4	0.007	0.03	0.05
5	0.0007	0.006	0.01

TABLE 6. Approximate probability that none of receipts generated by the $(a+1)$-st voter remain with any of the n auditors after all T people have voted. The data shown is for $a = 10$.

6. Issues

(1) **Disposal of extra receipts:** If the number of receipts in a receipt-bundle is greater than the number of auditors, care must be taken for the disposal of the extra receipts. Otherwise a voter could secretly keep one of her own receipts.
(2) **Forging receipts:** A party could forge a receipt and claim that the forged receipt should be on the bulletin board. To do this, a party would have to have access to the private keys of both the administrator and one of the machines. A main purpose of the auditors is to protect against such attacks.
(3) **Bribery and coercion:** As described above, there is still a non-negligible risk of bribery and coercion. If it is deemed too high, during the step where the voter and auditor swap (or don't swap) receipts, the auditor could place more than one receipt from his collection in the can. This would correspondingly reduce the odds that any individual voter retains her own receipt. In fact, each voter who gets to leave the polling station with a receipt could be required to exchange it for someone else's. This would eliminates this avenue for bribery, but has the disadvantage of reducing the psychological effectiveness of the verification by individual voters.
(4) **Initial voters:** It is not possible for the first voter to exchange receipts. Hence we must force $k = 0$ for her. In fact, for the first few voters, allowing one to exchange could allow someone to deduce with high probability how one of these initial voters voted. For this reason, we should require $k = 0$ for the first several voters. The longer this restriction is in place, the less information is gained by following voters, but the lower the number of voters who get to leave the polling station with a receipt.
(5) **Rare selections:** A coercive party could require a rare combination of selections. He would be able to recognize these unusual arrays among the publicly posted votes. If this party did not see the requisite number of votes of this type, he could punish the person or group of persons who were instructed to vote with those combinations. One possible solution is to only post the choices for certain races on the ballot. For instance, if only the vote for president was listed, there wouldn't be enough races listed for this coercive technique to be workable. Another solution, though unwieldy, would be to have different receipts for each race being voted on.
(6) **Non-universality:** More than half of voters are denied the right to check their own ballots. However, a fraction $n/(n+1)$ of the n voters get to check somebody's ballot.
(7) **Order matching:** If the initialized receipt-bundles are not properly randomized before being selected by voters, any record of the order in which voters voted could be be matched to how they voted.
(8) **Memorization of serial numbers:** If the protocol is followed, the only receipts a voter sees by the time she exits the polling station are her original one and perhaps one that she exchanges for with an auditor. Suppose v is the vote combination a coercer/briber desires and that a voter leaves the polling station claiming that the receipt with serial number x has vote v. Once the votes are published on the bulletin board, the coercer/briber can confirm this. If confirmed, it says that the voter either voted with receipt-bundle x, or that she happened to exchange with an auditor for a receipt with $SN = x$ and $V = v$. Assuming that there are more than two possibilities for V, the latter case is unlikely. This means that the briber/coercer knows with some confidence how the voter voted.

To avoid a version of the above scenario, neither the serial number x, nor the two digital signatures Sig and Sig' should be easily memorizable. Even partial memorization could be enough for bribery or coercion if the portion memorized suffices to uniquely identify a receipt-bundle.

As an example of how to stymie memorization, one could write the serial number as a 5×10 binary grid where twenty of the positions are meaningful and the rest of the grid is filled in randomly. Upon posting on the bulletin board, only the meaningful entries need to be posted. If the positions of the meaningful entries are protected as the administrator's private key is, a voter would not know which entries to memorize.

(9) **Discrepencies:** If the private keys are uncompromised, the risk of ungrounded accusations of fraud against the administering body should be slight.

(10) **Subliminal channels:** The protocols of Chaum and of Neff are at least potentially susceptible to subliminal channels in the encrypted ballots (cf. [4]). Such attacks are not possible in our randomized scheme if the receipt-bundles are initialized and signed properly. In particular, if the serial numbers are printed onto the receipt-bundles before being selected by an individual voter, these numbers cannot encode any information about the voter's identity. Furthermore, the digital signatures on the receipts are computed in a deterministic manner according to data on the receipt-bundle and certain private keys. As such, they cannot encode information linking the ballot to the voter.

(11) **Unnoticed switches:** As described, no voter can be positive that she leaves the polling station with her original receipt. This arises from our assumption that the serial numbers and digital signatures are not memorizable along with the possibility that she switch with a receipt containing an vote identical to the one she cast. This ambiguity could be removed, for instance, by allowing a window in the mixing can through which the voter could peer as the two receipts are being mixed.

References

[1] D. Chaum, *Secrete-ballot receipts: True voter-verifiable elections*, IEEE Security & Privacy Magazine **2** (Feb 2004 Jan), no. 1, 38–47.

[2] D. Chaum, P. Y. A. Ryan, and S. Schneider, *A practical, voter-verifiable election scheme*, Technical Report CS-TR-880, University of Newcastle upon Tyne, 2004.

[3] Lillie Coney, Joseph Lorenzo Hall, Poorvi L. Vora, and David Wagner, *Towards a privacy measurement criterion for voting systems*, Proceedings of the 2005 national conference on Digital government research (2005), ACM International Conference Proceedings Series, 2005, pp. 287-288, available at `http://doi.acm.org/10.1145/1065226.1065324`.

[4] D. Graham-Rowe, *Scratch-and-Vote system could help eliminate election fraud*, Technology Review (August 2006).

[5] Jonathan K. Hodge and Richard E. Klima, *The mathematics of voting and elections: a hands-on approach*, Mathematical World, vol. 22, American Mathematical Society, Providence, RI, 2005. MR **2139211 (2005k:**91097)

[6] C. Karlof, N. Sastry, and D. Wagner, *Cryptographic voting protocols: A systems perspective*, Proceedings of the Fourteenth USENIX Security Symposium (August, 2005).

[7] A. Neff, *A verifiable secret shuffle and its applications to e-voting*, Conference on Computer and Communications Security (2001).

[8] ———, *Practical high certainty intent verification for encrypted votes*, 2006. Ver. 1.0.

[9] ———, *Verifiable mixing (shuffling) of ElGamal pairs*, 2006. Ver. 1.2.

[10] Donald G. Saari, *Geometry of voting*, Studies in Economic Theory, vol. 3, Springer-Verlag, Berlin, 1994. MR **1297124 (96d:**90022)