Stochastic Voting Protocol

Hiroaki Kikuchi† Jin Akiyama‡ Howard Gobioff

Gisaku Nakamura‡

February, 1998
CMU-CS-98-112

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

† Tokai Univeristy/Carnegie Mellon Univeristy
‡ Tokai University

Abstract

This paper proposes a voting scheme that protects voters privacy even when the Central Tabulating Facility reveals individual responses. The basic idea is to add random noise to the true opinion by randomizing in a way such that the true vote is chosen with higher probability than the other alternatives. A major part of this paper is to outline the statistical propeties of our proposed protocol. A commitment protocol is also proposed to cope with dishonest voters who do not randomize their votes. The primary result is that the accuracy of the voting result improves as the number of voters increases.

Keywords: Electronic Voting, Randomized Algorithm, Security Protocol, Confidence Interval, Commitment Protocol

1 Introduction

> *Alice is a student who is evaluating her instructor, Bob, by submitting an evaluation form through his Web page. She is afraid that Bob could discover her vote by examining the access logs and retaliate by giving her a low grade if she voices her true opinion (she find the course was borning!) Fearing retribution, Alice answered that the course was excellent.*

Electronic voting is sometimes skewed by fear that the the Central Tabulating Facility (CTF) might violate voters privacy, though it has the potential of being cheaper and less time consuming than the conventional voting. In this example, we make the following observations:

1. The privacy of the voter (Alice) is paramount. Neither the CTF (Bob) nor anyone else can associate vote to the voter who cast it.

2. Bob requires voter authentication in order to permit only eligible voters to vote and to ensures that each voter can vote only once.

3. Alice wishes to verify that her vote is counted in the tally, no vote is eliminated from the tally, and no bogus vote is tallied.

4. Highly accurate result is not required. A rough estimate is good enough to rate instructors.

5. An efficient and light-weight implementation is feasible in combination of current technologies.

The last two properties makes this problem different from the traditional electronic voting problem, which involve a considerable sacrifice in communication and computation costs to achieve high accuracy. Our goal is to provide a practical and efficient protocol for conducting surveys or votes, even though the accuracy may be lost. Here are a few possible applications of this protocol.

- Voting for choosing news groups to be carried in local site.

- Conducting surveys on topics of sensitive (or illegal) matter.

- Computing average salaries of a closed group.

- Determine how much money to donate to an institution.

- Rating Web sites.

To deal with the issue of privacy, many protocols based on cryptographic techniques have been proposed. Broadly, there are two categories.

1. Blind signature with anonymous channel[2, 3, 4, 5]

   Each vote is blindly signed by an authority and then collected by way of anonymous channel such as Mixnet[2] so that no one can associate a vote with a voter, but only authorized votes are tallied. This mechanism is equivalent to the class of digital cash protocols that satisfie the privacy of payment.

2. Multiparty computation [6, 9, 8, 10]

   Votes are divided up among independent CTFs so that no single (or up to a constant) CTF can determine any individual vote. Secret sharing and relevant techniques are used to sum up all votes without revealing anyone's vote. Verifiable secret sharing schemes and zero-knowledge protocols are often used for guarantee that both voters and the CTFs follows the protocol correctly.

Ideally, both approaches should be able to hold secure votes. Although there have been some attempts for anonymous channel[11, 12, 17], these involve large communication costs and hence not deployed over the internet. The multi-party computation protocol solves the privacy problem and holds some theoretically interesting properties such as receipt-freeness[7]; however, the protocol is quite computationally complex and does not scale well.

In this paper, we propose a new light-weight anonymous voting mechanism. Our solution for protecting voter's privacy is to randomize votes by adding noise to votes so that the true vote is chosen with higher probability than the other alternatives. The result would be estimated by discarding the random noises according to statistical properties of expected noise.

For simplicity, consider a single-bit (yes/no) vote, which will be extended to multiple candidates protocol in Section 4.1.

1. Each voter flips a coin. In the case of heads, the voter give his true vote. Otherwise, the voter flips a second time and votes on the basis of the second flip.

2. Votes are tallied and summed. The CTF publishes the tally m. Let $m = 60$ given the total voters, n, is 100.

Voter submits his vote with his identity from web page. The CTF immediately adds the vote to the tally (m). The voter can make sure the CTF's computation by seeing the difference of the tally from the one before he submitted. To ensure that no bogus vote is added to the tally, the CTF should add the voter's identity to the list of voters that have submitted votes and finally publish the list. His vote may be noticed by other voters, but no one can identify his true vote as the CTF can not do it.

Here we have an intuitive solution to discard the random noise from the tally. Since the tallied votes m includes a half random component, in which a half of the component is yes, the total of true votes is $m - n/4 = 60 - 25 = 35$, which is the result for the half population, thus we have the estimated result by doubling it, that is, $k^* = 2 \cdot 35 = 70$. In this way, given $m = 60$, about 70 voters are likely to have yes votes.

3. With the published total number of yes votes, m, voters can estimate the result, k^*, as follows:

$$k^* = \frac{1}{1-p}(m - npq) \tag{1}$$

   where both p and q denote 1/2 probability of coin flipping.

We define the basic protocol by the above three steps. Obviously, the basic protocol has the following properties:

- It is an one-shot protocol. Even one round between voters and the CTF is not involved.
- Low computational costs. Neither computational nor information-theoretical assumption is required.
- No trusted third party is assumed. Even the CTF can not learn whether voters are true or random.
- The protocol is scalable to a number of voters. (Beside, the accuracy of vote increases as more voters participate in voting.)

However, the proposed protocol has the following difficulties:

1. **Statistical Analysis** Equation (1) is derived by intuition without theoretical analysis, and not always correct. For instance, consider what would happen when $m = 80$. According to Equation (1), the estimate of true yes votes is $k^* = 2(80 - 100/4) = 110$, which exceeds the total number n ?! What does this mean?

   We will answer this question by statistical analysis of the basic protocol in Section 3.2.

2. **Accuracy** The proposed protocol sacrifices the accuracy of the result to protect voter's privacy. The solution mentioned above gives an *approximation* of the votes; however, we have no guarantee of the accuracy of estimate. What can we say about the difference between the estimate, k^*, and the true number of yes votes, k?

   According to the statistical propeties of randomness, the fraction of error would be relatively small as more voters participate in the voting. We will answer to this question by showing the confidence interval of estimate.

3. **Dishonest voter** Since each voter randomizes his vote secretly, some malicious voters can vote the way they want without adding random element, which will skew the final result.

   We present a revised protocol so that no voter can cast a vote without adding random noise.

In this paper, after we give a basic formalization of proposed protocol in Section 2, we study probability distributions associated with some random variables m and k in Section 3. The probability distribution follows an expected value and the most likely value for k in Section 3.2. The correctness of the intuitive solution will be clarified mathematically. In Section 3.3, we discuss the variance of k with regards to some parameters in the protocol in order to figure out a confident interval given n and m under an assumption the random variable of k is distributed normally.

To prevent dishonest voters from disrupting voting, in Section 4.1, we propose a modification of the basic protocol using an one-way function as commitment of vote.

2 Basic Protocol

First, we define a basic protocol which deals with a simple binary election where individual votes are either 1 (means "yes") or 0 (means "no").

2.1 Notation

The following is a list of symbols.

n	number of voters
m	number of ”Yes” votes observed in the protocol
k	number of voters who have true ”Yes”
l	number of candidates ($l = 2$ in single vote)
p	probability that a voter picks random vote
q	probability of random vote being ”Yes” ($q = 1/l$)
$P(m\|k)$	probability that number of yes votes is m given number of true votes is k
$P(k\|m)$	probability that number of true votes is k given number of yes votes is m
$P(m)$	prior probability of m
$P(k)$	prior probability of k
M	random variable in $\{0, .., n\}$ of m
K	random variable in $\{0, .., n\}$ of k
$\sigma(m\|k)$	standard deviation of M given k
$\sigma(k\|m)$	standard deviation of K given m
$\mu(k\|m)$	($E[k\|m]$) an expected value of M given k

2.2 Protocol Definition

We have n voters and a single Central Tabulating Facility (CTF).

Protocol 1

Step 1 Voter i has true vote v_i in $\{0, 1\}$. Voter i randomly composes a ballot b_i such that

$$b_i = \begin{cases} v_i & \text{with probability } 1 - p \\ r \in \{0, 1\} & \text{with probability } p \end{cases}$$

where r is a random number such that a probabilities of r being 1 is q and being 0 is $1 - q$. The voter sends b_i to the CTF.

Step 2 The CTF tallies ballots and publishes the result m, that is, $m = b_1 + \cdots + b_n$.

Step 3 Given m, voters and the CTF learn the estimated number of true yes votes by

$$k^* = \begin{cases} 0 & \text{if } m < npq \\ n & \text{if } m > n(1 - p + pq) \\ \frac{1}{1-p}(m - pqn) & \text{otherwise} \end{cases}$$

We call v_i a *true vote* and r a *random vote.*

In the following section, we will show how the estimate k^* is statistically derived and examine the confidence of the estimate.

To prevent ineligible voters from voting and eligible voters from casting multiple votes, the CTF requires voter a password, PIN or digial signature with his certificate at Step 1. The CTF checks the voter on a list of eligible voters. Once a vote is submitted, it may be immediately added to the tally so that the voter can ensure that his vote is counted correctly. Since the vote is randomized, the tally may be made public without revealing his true opinion. This property makes it easier for voters to verify the CTF follows the protocol correctly.

3 Estimate

3.1 Probability Distribution Functions of Voters

In this section, we give a probability distribution function of the actual total number of "yes" votes k given m.

Lemma 3.1 *Let p be a probability of picking random vote in Protocol 1. The probability of observing m votes given k voters who have true yes vote is given by*

$$\begin{aligned} P(m|k) &= \sum_{i=max(0,m+k-n)}^{\min(m,k)} \binom{k}{i} \binom{n-k}{m-i} \\ &\quad \times (1-p+pq)^i (p-pq)^{k-i} \\ &\quad \times (pq)^{m-i} (1-pq)^{n-k-m+i}. \end{aligned} \tag{2}$$

(All proofs and derivation of the equations are in the appendix.)

Obviously, the behavior of this function is similar to a Binomial distribution, though the Eq. (2) is more complicated. Figure 1 shows the probability distributions of $P(m|k)$ when $k = 0, 6, 10, 14, 20$, $n = 20$, and $p = q = 1/2$. We also indicate the results of computer simulation ($k = 14$) on the figure.

We are interested in the posterior probability $P(k|m)$, which is the true number of yes votes, k, given the observed m votes. Bayes theorem is used to calculate the posterior probability $P(k|m)$ from the prior probability $P(k)$, $P(m)$ and $P(m|k)$.

Since we have no prior knowledge that particular value of k is more likely than another, it is reasonable to assign the same prior probability to every value of k. Furthermore, since the possible value of k are in $[0, n]$, these prior probabilities sum to 1. Hence, we should assume that

$$P(k) = 1/(n+1)$$

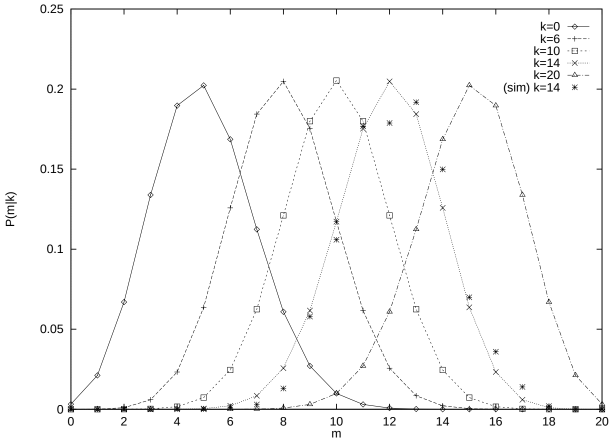

Figure 1: Probability Distribution $P(m|k)$

and we have

$$P(m) = \sum_{k=0}^{n} P(m|k)P(k) = 1/(n+1)\sum_{k=0}^{n} P(m|k)$$

Figure 2 illustrates the probability distribution of m when $n = 20$. The result of the computer simulation is also indicated. The probability of m highest when $k = nq$ and is independent of p.

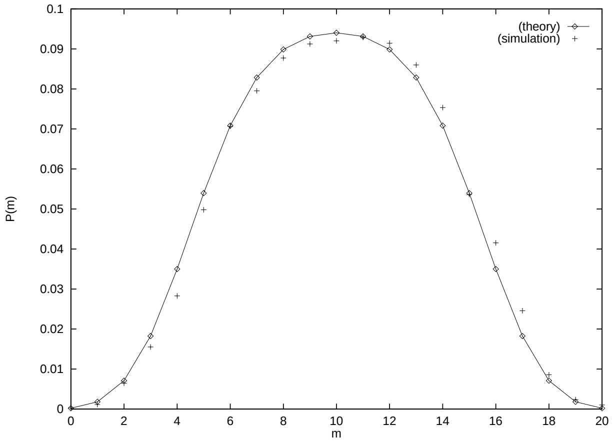

Figure 2: Probability Distribution $P(m)$

Finally, we obtain the desired posterior probability $P(k|m)$ which provides an estimated value of a total number of yes votes.

Theorem 3.1 *Suppose m is observed in Protocol 1 and we have no prior knowledge of k. The posterior probability of number of true yes votes is computed as follows:*

$$P(k|m) = \frac{P(m|k)}{\sum_{i=0}^{n} P(m|k=i)}, \tag{3}$$

where $P(m|k)$ is defined in Eq. (2).

Figure 3 shows an example of probabilities of the number of true votes k for $m = 2, 6, 10, 14, 18$ and $n = 20$. Notice that in comparison with Figure 1, the probability distributions varies depending on m.

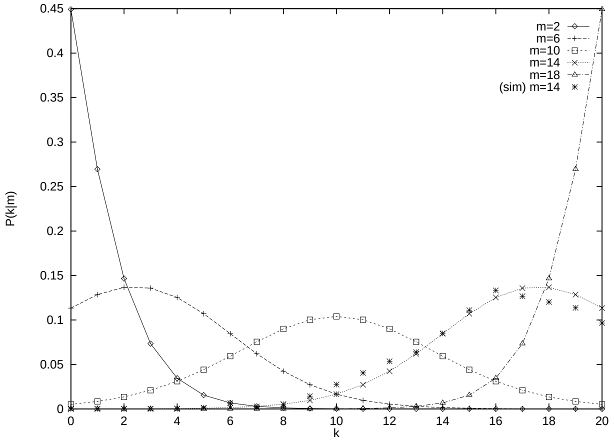

Figure 3: Probability Distribution $P(k|m)$

3.2 Expected Value and The Most Likely Value of Votes

Given the probability distribution function we can obtain the expected value, which should be approximately k^* as defined by Equation (1).

The expected value $E[X]$ of random variable X, defined by $E[X] = \sum_{x \in X} xP(x)$, provides the exact mean number of votes. The most likely value is a value of X which maximizes $P(X)$. Note that the expected value does not always maximize the probability.

Theorem 3.2 *The expected value of M is identical to the most likely M and obtained as follows:*

$$E[M|k] = L[M|k] = k(1-p) + npq \tag{4}$$

Notice that this result holds the Equation (1) by having correspondences $k^* = k$ and $m = E[M|k]$.

$$k = (E[M|k] - npq)/(1-p) = k^*$$

In other words, the derivation of k^* was solving the inverse problem of the expected value of M to be identical to the observing m.

Next, we consider the expected value of K given m. Straightforwardly, an expected value $E[K|m]$ is computed in the following formula,

$$E[K|m] = \sum_{k \in \{0,\ldots,n\}} P(k|m)k,$$

where $P(k|m)$ is computed in Eq. (3). However, we have no closed form for $E[K|m]$. Instead, we consider the most likely K given m, written by $L[K|m]$, which can be easily computed and nearly equal to $E[K|m]$.

Theorem 3.3 *The most likely value $L[K|m]$ is given by*

$$L[K|m] = \begin{cases} n & \text{if } m \geq n(1-p+pq) \\ \lfloor \frac{1}{1-p}(m - pqn) \rfloor & \text{if } n(1-p+pq) > m \geq np \\ \lceil \frac{1}{1-p}(m - pqn) \rceil & \text{if } np > m > pqn \\ 0 & \text{if } pqn \geq m \end{cases} \tag{5}$$

Figure 4 illustrates the expected value $E[K|m]$ and the most likely value $L[K|m]$. We see both values are nearly identical for most m. Let us recall the difficulty of our first estimate of k^* that $m = 80$ gives estimate of $k^* = 110$, which exceeds $n = 100$. In Figure 4, this case happens at m being greater than 30, where there is a big difference between $L[K|m]$ and $E[K|m]$.

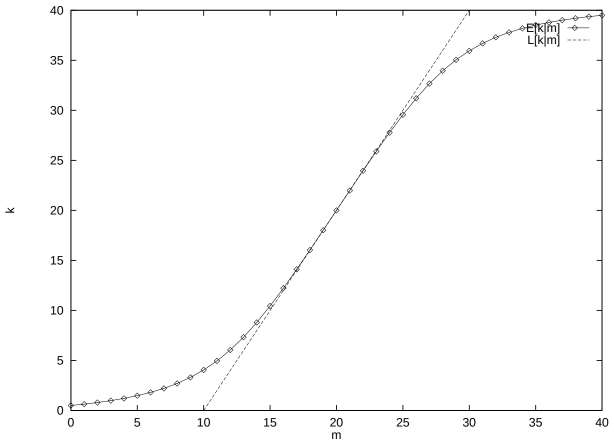

Figure 4: Expected value $E[K|m]$ and Most Likely value $L[K|m]$

3.3 Confidence Interval of Votes

Now, we consider the confidence intervals of the estimated values. One common way to deal with the uncertainty in our estimate is to give an interval within the true value is expected to fall, along with the probability.

In order to figure out the size of interval, we assume that the exact probability distributions can be approximated with the Normal distributions having the same expected value μ_k $(= E[K|m])$ and the standard derivation $\sigma_k(= \sqrt{Var[K|m]})$. Then, we have the well-known fact that the true k will fall into the interval

$$\mu_k \pm 1.96\sigma_k$$

with 95 % probability[13].

For example, consider a vote of $n = 10$ and $m = 7$. According to the previous results, we have the expected value and the most likely value of K as $E[K|m] = 7.57$ and $L[K|m] = (m - pqn)/(1-p) = (7-2.5)2 = 9$. Then, by letting $\sigma_k = 1.97$, we have the 95 % confident interval as follows,

$$7.57 \pm 1.96 \times \sigma_k = [3.71, 11.4].$$

The ratio of the confident interval to the range of a random variable will shrunk as n gets larger. This implies that the accuracy of the estimate will improve as more voters participate in voting. We examines the influence of the variance of k with regards to the parameters including p, m and n.

Theorem 3.4 *Given true number of yes votes k, the variance of observing yes votes M is*

$$Var[M|k] \quad = \quad pk(p-1)(2q-1) + npq(1-pq) \tag{6}$$

Note that when $q = 1/2$, the first term disappears and the variance is $Var[M|k] = npq(1-pq)$, which is a constant independent to k. In Figure 5, we show how m influence the standard derivation of k. Also, the difference with regards to probabilities $p = 1/4, 1/2, 3/4$ are indicated. In all cases, we see the standard deviations maximizes at $m = n/2$, which gives the upper bound of the confidence interval.

3.4 Estimate Error

Now, we consider estimate error, which is a ratio of the confidence interval to the number of voters, n. The ratio represents how good the expected value is estimated. It approaches 0 as the interval gets smaller relative to the number of voters. Since the confidence interval is proportional to the standard deviation, we simplify the metric to the ratio of standard deviation to n.

According to the Central Limit Theorem[13], regardless of the distribution of random variable, the probability distribution of a sample mean approaches

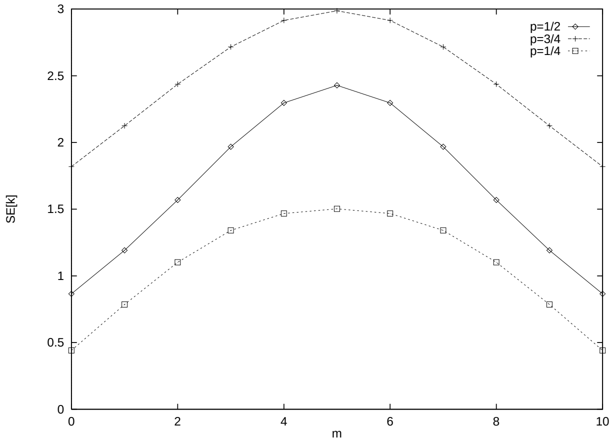

Figure 5: Standard Deviation σ_k wrt m, p

a normal distribution as $n \to \infty$. The standard deviation is proportional to $1/\sqrt{n}$ where n is sample size. Hence, the ratio of standard deviation of k to n will be also proportional to $1/\sqrt{n}$ because the votes are chosen by certain probability distributions, and summed in the protocol. Therefore, the accuracy of the proposed voting protocol improves as the number of votes increases.

For instance, we shows the behavior of estimate error ratio defined by $\sigma_k(n/2)/n$ with regards to the size of voting $n(= 10, 20, \ldots, 100)$ in Figure 6.

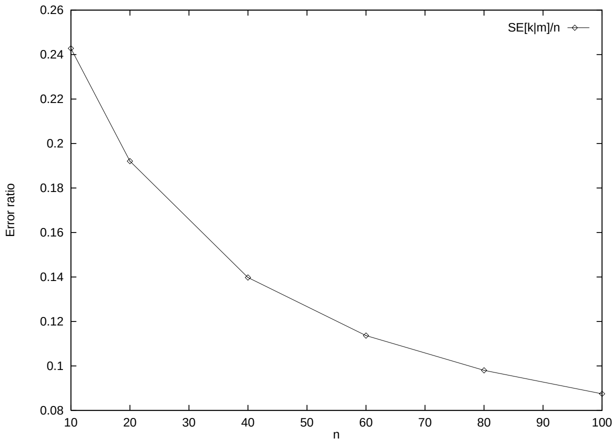

Figure 6: Estimate Error Ratio wrt n

4 Revised Protocol

4.1 Commitment Protocol

To prevent dishonest voters from casting invalid vote, n we propose a modification of the basic protocol using an one-way function as commitment of vote.

The weak point of Protocol 1 is to allow a voter to pick a random number by himself without coin flipping. There is the fair coin flipping protocols using the bit-commitment functions[16], in which voter shows to the CTF only the outcomes of one-way function for some votes that the fraction of yes votes represents the probability of coin being head (p), and then the CTF chooses one of them. The CTF can not figure out the vote from the outcome of one-way function, and the voter can not change his vote after one of the outcomes is chosen. However, it is not sufficient for our purpose; the malicious voter can prepare all yes (or no) votes, which disrupts the voting. Voter should prove to the CTF that her vote is randomized correctly without revealing his vote.

We present a simple and lightweight protocol for fair coin flipping without revealing the probability of coin flipping, which implies his true vote. Note that true yes voter casts 1 and 0 with probabilities of $1-p+pq = 3/4$ and $p-pq = 1/4$, respectively. While, true no voter follows $pq = 1/4$ and $1-pq = 3/4$ probablities.

Protocol 2

Step 1 Each voter picks four random numbers, r_1, r_2, r_3, r_3, where each has large size enough to feed an one-way function f. If a voter is willing to vote "yes", he composes four commitment numbers as follows;

$$c_1 = f(r_1|1),\ c_2 = f(r_2|1),\ c_3 = f(r_3|1),\ c_4 = f(r_4|0),$$

where $r_i|x$ means a bit concatenation. Otherwise, he composes

$$c_1 = f(r_1|0),\ c_2 = f(r_2|0),\ c_3 = f(r_3|0),\ c_4 = f(r_4|1).$$

The voter randomizes the order of the commitment numbers so that the CTF can not learn the assignment. Let us call the randomized commitments c'_1, c'_2, c'_3, and c'_4.

Step 2 The CTF randomly picks one of these commitment numbers, and ask the voter to open the committed value.

Step 3 The voter opens the corresponding input number $r_i|x$. She also opens one more input number $r_j|y$ so that x is not equal to y.

Step 4 The CTF verifies that x and y are different and the two input numbers $r_i|x$ and $r_j|y$ satisfies

$$c'_i = f(r_i|x) \text{ and } c'_j = f(r_j|y).$$

The CTF tallies x if only if both identities hold. Otherwise, the vote is rejected.

In step 2, a voter has five possibilities to composes for numbers. First, the commitment numbers $(1,1,1,0)$ and $(0,0,0,1)$ corresponding to "yes" and "no" are valid. Second, $(1,1,1,1)$ is invalid and should be rejected. Suppose the CTF chooses the first number. Then, at Step 3, the voter has to open not only the input $r_1|1$ but also the opposite input $r_j|0$ for some j, which does not exist in his commitment numbers. Therefore, the invalid vote always fails. The case of $(0,0,0,0)$ is prohibited in the same way. At last, $(1,1,0,0)$ gives an even probability of "yes" and "no", which can succeed the protocol without being detected, but has no effect on the result of voting. It may enlarge the probability p of random voting and the confidence interval. We assume that the influence is not so large because the canceling voters can not control the result as they like. Accordingly, any serious misbehavior of voters can be rejected in the protocol.

This protocol is a two round protocol and the messages are small. Consequently, the protocol is light-weight in communication and computation.

More generally, the commitment protocol can be extended to cope with arbitrary probability of p. By letting $\alpha_0, \alpha_1, \beta$ be intergers so that $\alpha_1/\beta = 1-p+pq$ and $\alpha_0/\beta = pq$ $(\alpha_1 > \alpha_0)$, a voter can commit his vote as

$$f(r_1|1), \ldots, f(r_{\alpha_1}|1), f(r_{\alpha_1+1}|0), \ldots, f(r_\beta|0)$$

or

$$f(r_1|1), \ldots, f(r_{\alpha_0}|1), f(r_{\alpha_0+1}|0), \ldots, f(r_\beta|0)$$

in the basis of his vote in Step 1. In response to the CTF at Step 3, the voter opens the selected commitment number $r_i|x$. Furthermore, letting $\Delta\alpha = \alpha_1 - \alpha_0 > 0$, the voter opens $\Delta\alpha$ opposite commitments number $r_{j_1}|y, r_{j_2}|y, \ldots, r_{j_{\Delta\alpha}}|y$ for $x \neq y$. The $\Delta\alpha$ is the maximum number that no more opposite commitment could reveal whether the vote is yes or no. Note that an invalid vote is rejected if the number of committed 1s exceeds α_1 or the number of committed 0s is less than α_0. As the same as the basic commitment protocol, it is is necessary condition but not sufficient for invalid votes; the cancelation is allowed by composing γ commitments of 1 such that $\alpha_0 < \gamma < \alpha_1$.

4.2 Multi Candidate Election

Protocol 1 can be extended to an election that has multiple candidates chosen from $\{0, 1, \ldots, l-1\}$. First, we assume that the candidates represent discrete quantities such as news groups or web pages.

Protocol 3

Step 1 Voter i has true vote v_i in $\{0, 1, \ldots, l-1\}$. Voter i randomly composes a ballot b_i such that

$$b_i = \begin{cases} v_i & \text{with probability } 1-p \\ r \in \{0, 1, \ldots, l-1\} & \text{with probability } p \end{cases}$$

where r is a random number of $\{0, 1, \ldots, l-1\}$ chosen with uniform probability of $q = 1/l$.

Step 2 The CTF collects ballots and compute l tallies for each $i \in \{0, \ldots, l-1\}$ as,

$$m_i = |\{b_j | b_j = i, j \in \{1, \ldots, n\}\}|$$

and publishes the results, $m_0, \ldots, m_{l-1}$.

Step 3 For each m_i, voters and the CTF computes the estimated number votes by

$$k_i^* = \begin{cases} 0 & \text{if } m_i < npq \\ n & \text{if } m_i > n(1-p+pq) \\ \frac{1}{1-p}(m_i - pqn) & \text{otherwise} \end{cases}$$

The statistical propeties of the basic protocol still hold in multi candidate protocol. Since the l estimated values $k_0^*, \ldots, k_{l_1}^*$ may not sum up to n, normalization of k_i^* is required. Clearly, this protocol includes Protocol 1 as the special case $l = 2$.

Next, we consider the candidates represent continuous quantities such as money or rating value.

Protocol 4

Step 1 Same as Protocol 3.

Step 2 The CTF collects ballots and compute tally $m = b_1 + \cdots + b_n$ and publishes the results m.

Step 3 Given m, voters and the CTF computes the estimated total of votes by

$$k_i^* = \begin{cases} 0 & \text{if } m < np\mu_R \\ n & \text{if } m > n(1-p+p\mu_R) \\ \frac{1}{1-p}(m - np\mu_R) & \text{otherwise} \end{cases}$$

where μ_R is a mean of random variable R, that is, $\mu_R = \frac{l-1}{2}$.

With the expected value of uniform random variable R, we derive the estimated total votes, k^*, which should be divided by n to be a meaningful quantity, e.g., the average money to donate, or the course grade determined by n parties.

4.3 Information Leak

The proposed protocols are not perfect in the sense that it may reveal partial information about votes. For example, consider a randomized vote b_i with $p = q = 1/2$. What can we say about the true vote from the observing vote?

We know the following probability of random variable of votes B given true vote V:

$$\begin{aligned} P(B=1|V=1) = &\quad 1-p+pq, \quad P(B=0|V=1) = p-pq, \\ P(B=1|V=0) = &\quad pq, \quad P(B=0|V=0) = pq, \end{aligned}$$

and the Bayes theorem provides the posterior probability of true vote V given observing vote B as follows:

$$P(V|b) = \frac{P(b|V)}{\sum_{v \in V} P(b|v)}$$

For example, when a voter casts randomized vote of $b_i = 1$, his true vote v_i would be 1 with probability of 3/4. The leaking information depends on the probability p for random voting. However, approaching p to 1 implies low accuracy of the result. We have a tradeoff between the leaking information and the accuracy of the result of voting.

5 Conclusion

We have proposed a light-weight electronic voting protocol that protects privacy of voters. Based on the statistical analysis of the protocol, we have clarified a meaningful estimate of the voting result. We have presented a commitment protocol that prevents dishonest voters from voting without randomizing their votes. The main result is that the accuracy of the voting result improves with the number of of voters increases. The proposed protocol is expected to be applied in a variety of electronic commerce protocols as a primitive technique.

Acknowledgement

The authors thank Prof. Doug Tygar of Carnegie Mellon University for his encouragement to pursue this work.

References

[1] Bruce Schneier, *Applied Cryptography*, 2nd Edition, John Wiley & Sons, 1996

[2] Chaum, D., Untraceable Electronic Mail, Returen Addresses, and Digital Pseudonyms, Communications of the ACM, Vol. 24, No.2, pp.84-88, 1981

[3] Chaum, D., Blind signatures for untraceable payments, Crypto 82, 1983, pp.199-203

[4] Chaum, D., Elections with Unconditionally-Secret Ballots and Disruption Equivalent to Breaking RSA, Proc. of Eurocrypto'88, pp. 177-182, 1988

[5] Fujioka, A., Okamoto, T., and Ohta, K., A practical secret voting scheme for large scale elections, Auscrypt'92, 1993, pp.244-251

[6] Benaloh, J. and Yung, M., Distributing the Power of a Government to Enhance the Privacy of Voter, Proc. of ACM Symposium on Principles of Distributed Computing, pp.52-62, 1986

[7] Benaloh, J. and Tuinstra, D., Receipt free secret-ballot elections, Proc. of ACM Symposium on the Theory of Computing, 1994, pp.544-533

[8] Oded Goldreich, Silvio Micali and Avi Wigderson, How To Play Any Mental Game or A Completeness Theorem for Protocols with Honest Majority, *ACM STOC*, p.218-229, 1987

[9] M. Ben-Or, S. Goldwasser and A. Wigderson, Completeness theorems for non-cryptographic fault-tolerant distributed computation, STOC88, pp.1-10

[10] Sako, K. and Kilian, J., Secure Voting Using Partially Compatible Homomorphisms, Proc. of Crypto'94, pp. 411-424, 1994

[11] Paul F. Syverson, David M. Goldschlag, and Michael G. Reed, Anonymous Connections and Onion Routing, *1997 IEEE Symposium on Security and Privacy*, pp.44-54, 1997

[12] Michael K. Reiter and Aviel D. Rubin, Crowds: Anonymity for web transactions, DIMACS Technical Report 97-15, April 1997.

[13] Alberto Leon-Garcia, Probability and Random Processes for Electrical Engineering, second edition, Addison-Wesley, 1994

[14] K. Trivedi, Probability and statistics with reliability, queuing, and computer science applications, *Prentice-hall*, 1982

[15] L. Gonick, W. Smith, The cartoon guide to statistics, *HarperPerennial*, 1993

[16] M. Blum, "Coin Flipping by Telephone: A Protocol for Solving Impossible Problem, Proc. of the 24th IEEE Computer Conference (CompCon), pp.133-137, 1982

[17] Anonymizer, `http://www.anonymizer.com`

A Proof

Proof of Lemma 3.1 Assume that a voter has yes vote. The probabilities that $b_i = 1$ and 0 are $1 - p + pq$ and $1 - (1 - p + pq) = p - pq$, respectively. Otherwise, the probabilities that $b_i = 1$ and 0 are pq and $1 - pq$, respectively. Then, we have the probability generating function of $P(m|k)$ as follows,

$$f(z) = \sum_{m=0}^{n} P(m|k)z^m = ((1-p+pq)z + (p-pq))^k((pq)z + (1-pq))^{n-k}. \quad (7)$$

By expanding $f(z)$, we have

$$f(z) = \sum_{i=0}^{k} \binom{k}{i} (1-p+pq)^i(p-pq)^{k-i} \sum_{j=0}^{n-k} \binom{n-k}{j} (pq)^j(1-pq)^{n-k-j}z^{i+j},$$

where a coefficient of z^{i+j} corresponds to $P(m|k)$ with letting $m = i + j$. Thus, picking up all coefficients for i and j that holds $i + j = m$, we have

$$P(m|k) = \sum_{i=max(0,m+k-n)}^{\min(m,k)} \binom{k}{i} \binom{n-k}{m-i} (1-p+pq)^i(p-pq)^{k-i}(pq)^{m-i}(1-pq)^{n-k-m+i}.$$

(Q.E.D)

Proof of Theorem 3.1 According tothe Bayes theorem, we have

$$P(k|m) = \frac{P(k,m)}{P(m)} = \frac{P(m|k)P(k)}{\sum_{i=0}^{n} P(m|k=i)P(k=i)}$$

where $P(k) = 1/(n+1)$ is a constant, so by eliminating it, we have the theorem. (Q.E.D)

Proof of Theorem 3.2 According to the basic properties of probability generating function, we have the lemma in the following way.

$$\begin{aligned}
f'(z) &= \frac{d}{dz}f(z) = \sum_{m=0}^{n} P(m|k)mz^{m-1} \\
E[m|k] &= \sum_{m=0}^{n} P(m|k)m = f'(1) \\
&= k((1-p+pq)z + p - pq)^{k-1}(1-p+pq)(pqz + 1 - pq)^{n-k} \\
&\quad +((1-p+pq)z + p - pq)^k(n-k)(pqz + 1 - pq)^{n-k-1}pq \\
&= k(1-p) + npq
\end{aligned}$$

The identify of the most likely value is obvious from the behavior of the probability distribution function. (Q.E.D)

Proof of Theorem 3.3 Let $P_n(m|k)$ be a conditional probability of observing m votes given a true number of yes votes k. Then, we have the following recursive relationship on $P(m|k)$ with regards ot n:

$$P_{n+1}(m|k) = P_n(m-1|k-1)(1-p+pq) + P_n(m|k-1).$$

where $P_n(m|k)$ is probability of observing m given k in n voters. Then, the theorem is proved in mathematical induction with n. (Q.E.D)

Proof of Theorem 3.4 According to the propeties of probability generation function, we have the followings:

$$\begin{aligned}
f''(z) &= \frac{d}{dz}f(z) = \sum_{m=0}^{n} P(m|k)m(m-1)z^{m-2} \\
f''(1) &= \mu(m^2|k) - \mu(m|k) \\
&= k(k-1)(1-p+pq)^{k-2}(1-p+pq)^2(pqz+1-pq)^{n-k} \\
&\quad +2k(1-p+pq)^{k-1}(1-p+pq)(n-k)(pqz+1-pq)^{n-k-1}pq \\
&\quad +(1-p+pq)^k(n-k)(n-k-1)(pqz+1-pq)^{n-k-2}(pq)^2 \\
Var[m|k] &= \sum_{m=0}^{n} P(m|k)(\mu(m|k)-m)^2 \\
&= \mu(m|k)^2 - 2\mu(m|k)^2 + \mu(m^2|k) = f''(1) + f'(1) + f'(1)^2 \\
&= pk(p-1)(2q-1) + npq(1-pq)
\end{aligned}$$

(Q.E.D)