An Information-Theoretic Model of Voting Systems

Ben Hosp*, Poorvi L. Vora*
Dept. of Computer Science
George Washington University
Washington DC 20052
{bhosp,poorvi}@gwu.edu

Abstract

This paper presents an information-theoretic model of a voting system, consisting of (a) definitions of the desirable qualities of integrity, privacy and verifiability, and (b) quantitative measures of how close a system is to being perfect with respect to each of the qualities. It describes the well-known trade-off between integrity and privacy in this model, and defines a concept of *weak privacy*, which is traded off with system verifiability. This paper is a simultaneous submission to VSRW06 and WOTE06. The simultaneous submission has been approved by representatives from both Program Committees, as neither meeting will have printed proceedings.

1 Introduction

Elections in the United States have relied more and more upon computerized or electronic voting technology. Additionally, other democracies are also using electronic voting – examples include the UK's early internet voting trial, and India's use of a single type of dedicated electronic polling machine. Yet, the literature does not provide a standard model to compare the electronic voting systems with the electromechanical and paper-based systems they have replaced, or to compare them among themselves.

This paper presents a voting model that is based on information flow through an election system. Some of the more important desirable properties of voting systems – integrity, privacy and verifiability – are carefully defined in the model, and information theoretic metrics for the measurement of deviation from perfect are presented. The advantage of this model is (a) it provides a single framework in which to define and measure integrity, privacy and verifiability, and (b) the tradeoffs among these criteria are explicit in this model. In fact, the tradeoffs among these criteria arise exactly because all the information required to verify and release vote counts can only be obtained from the votes. This makes an information-theoretic approach the natural – if not the only – approach to study these.

We make some important points about our approach here. The measures we propose – for integrity, privacy, verifiability and usability – are based on the concept of entropy. The type of entropy – computational or information-theoretic – can, in principle, depend on what is best for the specific setting; the paper addresses, for the purpose of simplicity of presentation, only information-theoretic entropy. One may argue about whether the exact measures proposed are the best ones - that is, whether one uses an average or maximum entropy measure, whether one examines the system separately for each individual, or aggregates across individuals. The main focus of this paper, however, is not the exact details of the measures (though these are, in our opinion, the best of several alternatives), but the model itself and the manner in which it exposes all the tradeoffs.

Our uniform approach while determining measures has been that, when a system treats a particular user or groups of users differently from others, we are actually dealing with multiple systems, and each has its own measure. Hence, for example, if a system leaks more information on voters from location X than it does on voters from location Y, there are at least two measures of the system privacy: one for those voting in location X, and another for those voting in location Y. It is clear how this idea extends to, for example usability: the usability measure for a particular system for the visually handicapped would, in general, be different from that for those not visually handicapped. Thus, while an entropy-based measure averages over randomly-obtained outputs, it need not average over populations that are treated distinctly by the system.

This paper provides an initial attempt at formalizing the framework, and presents the types of questions that can be examined using it.

2 Prior Work

[12] contains one of the earliest list of voting system requirements, and many papers in the recent WOTE 2001 [14] and WEST 2002 [13] workshops

*These authors supported in part by NSF SGER 0505510

also include overviews of voting system requirements [4, 5, 7]. None provide a means of measuring performance with respect to the requirements. Papers on evaluating voting technologies include [2, 6], and several other papers from the NIST Workshop on Threats to Voting Systems [11], in particular [8, 10], provide an evaluation with respect to threats to count integrity. [1] provides a mathematical definition of voting system privacy, and a related entropy-based privacy measure, which our work draws heavily from.

3 Election Goals

This section provides a brief list of desirable properties of election systems; the goals have been drawn from prior work such as [12, 4, 5, 7, 9].

1. **Usability:** Ballots should be "cast as intended," meaning that an otherwise valid voter who intends to cast a vote for Candidate Alice should not be thwarted by election procedures or technology.

2. **Integrity:** Ballots should be "counted as cast," meaning that the voting system should declare that Candidate Bob received m votes if and only if exactly m ballots marked for Candidate Bob were cast.

3. **Privacy:** The secret ballot principle should apply to the election; voter i should not have the contents of her ballot associated with her in any way by anyone – even with the collusion of many parties, including election officials and other voters. Notably, this privacy should be involuntary, in the sense that even a set of colluding parties that includes voter i herself should not be able to prove the contents of her ballot once she has left the polling place.

   One may note that a system that provides privacy also provides **fairness** [9]: partial election results should not be available to anyone during the election. (This requirement ensures that the election is fair to all candidates, as the revelation of partial counts might encourage supporters of a winning candidate to abstain from voting when they might have otherwise voted.) Fairness is implied by privacy because the revelation of partial vote counts reveals information about individual votes.

4. **Verifiability:** Both the general public – including non-voting observers – as well as the individual voter should be able to rest assured that the above goals have been met. Such assurance should not require real-time observation of election procedures or secret information.

   Dispute-freeness [9] is a special kind of verifiability, where disputes raised by various parties as to the validity of the election are decidable based on information that is publicly-available. In other words, the dispute resolution procedure is publicly-verifiable.

5. **Robustness:** Errors and failures can be detected and fixed without impact upon the other goals.

4 The Model

In this section we describe our model and translate some of the goals of the previous section into mathematical conditions in the model.

We will consider that there are n voters casting ballots in an election. Let V_i be a discrete random variable representing voter i's ballot (or rather the votes it contains); $V_i \in \mathcal{V}$, the set of all possible ballots in the election. We will use V^* as shorthand for $[V_1, V_2, \ldots V_{n-1}, V_n]$; $V^* \in \mathcal{V}^n$. Let V^Σ be the vote count, in the form: "In the race for Governor, 600 votes for Alice, 400 for Bob; on Proposition 242, 580 votes for Yes, 420 votes for No, ..." V^Σ is therefore also a discrete random variable, but it is a deterministic function of V^*.

Let $\widehat{V^\Sigma}$ represent the vote count output by the voting system used to conduct the election, and let E be its entire "output". From our point of view, E will be some vector or set of discrete random variables. (Note that $\widehat{V^\Sigma}$ is a part of E). Assume that the voting system declares two algorithms, i.e. two sets of well-defined steps, $VoteCount$ and $ElectionOutput$, that, when applied to the votes, V^*, produce $\widehat{V^\Sigma}$ and E respectively. In the literature on cryptographic voting schemes, E has been referred to as the *View*. We use the different terminology in order to emphasize that our model need not be restricted to the cryptographic or other algorithms, but can also be used to evaluate the implementation as well as the procedures. When $VoteCount$ and $ElectionOutput$ are known, we represent $\widehat{V^\Sigma}$ and E by $VoteCount(V^*)$ and $ElectionOutput(V^*)$ respectively. Note that this does not imply that either of $VoteCount$ and $ElectionOutput$ is necessarily deterministic, simply that $\widehat{V^\Sigma}$ and E are the outputs of the voting system after applying a set of well-defined, known, steps to V^*

4.1 Preliminaries

We use the notion of entropy to define the mathematical goals and to measure deviation from perfect. As defined by Shannon [3], entropy is a mathematical measure of the uncertainty in a random variable. We concern ourselves only with discrete random variables, and measure entropy in bits. The entropy of discrete random variable X that takes on value x with probability $p_X(x)$ is

$$\mathcal{H}(X) = -\sum_x p_X(x) log_2 p_X(x)$$

Roughly speaking, the entropy of a random variable is understood to be the average number of bits required to represent it.

When two random variables X and Y are not independent, knowing the value of one reduces the uncertainty of the other. If $\mathcal{H}(X|Y)$ is the uncertainty in X if Y is known,

$$\mathcal{H}(X|Y) = \sum_y p_Y(y)\mathcal{H}(X|y)$$

The reduction in entropy in one variable, due to the other being known, is termed mutual information. It is defined as follows:

$$\mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y)$$

and it can be shown that:

$$\mathcal{I}(X;Y) = \mathcal{I}(Y;X)$$

The computational entropy of a random variable, roughly speaking, is the average number of bits required to represent it under the constraint that the algorithm generating the bits from the random variable is feasible in the computational model [15]. In certain instances, when secrecy is provided by computational assumptions, it is more appropriate to use computational entropy over "Shannon" (or absolute) entropy. We will point out these instances when possible. While the use of computational entropy in the definitions is outside the scope of this paper, it appears to be a straightforward extension of this work. Further, the fact that we do not address computational entropy explicitly should not be taken to imply that we require or recommend the use of only Shannon entropy in all cases.

4.2 Integrity

Election integrity requires that, if the voting system follows its declared algorithm $VoteCount$, there should not be any cast votes uncounted or any uncast votes counted. In other words, algorithm $VoteCount$, if followed, produces the *correct* vote count, V^Σ. Integrity does not address the issue of whether $VoteCount$ is indeed followed by the voting system; this is covered by the property of verifiability (see section 5).

Election integrity may be defined more precisely as follows:

Definition 1: An election system provides **perfect integrity** if $VoteCount(V^*) = V^\Sigma$.

Even if the system does not provide perfect integrity, the uncertainty in V^Σ is generally reduced on knowledge of $VoteCount(V^*)$. The reduction in uncertainty, $\mathcal{I}(V^\Sigma; VoteCount(V^*))$ could range anywhere from zero (indicating election results independent of the cast ballots, and hence an election with zero integrity) up to a maximum of $\mathcal{H}(V^\Sigma)$ (indicating perfect integrity). One could use a normalized value of this reduction in uncertainty to measure election integrity.

Definition 4: The **integrity measure** of an election system is

$$\mathfrak{I} = \frac{\mathcal{I}(V^\Sigma; VoteCount(V^*))}{\mathcal{H}(V^\Sigma)}$$

We assume that the system treats all voters similarly (if not, then there are really two election systems).

Example 1: Consider a voting system that produces a vote count through hand counting, where $VoteCount$ produces the average of N hand counts. The hand counts are not necessarily observed by the public, but, assuming that the algorithm is as declared, the uncertainty in V^Σ is the uncertainty due to hand counting. The integrity is not perfect, and the integrity measure increases with N. Whether the system actually does count the votes is addressed through the property of verifiability, see section 5.

We assume, wlog, that an integrity value of one implies perfect integrity – that is, that $VoteCount(V^*)$ does not produce a value $f(V^\Sigma)$ for f a deterministic invertible function that is not the identity. (If it did, the claimed vote count would be incorrect, but it would contain all the information necessary to obtain the correct vote count, which can be obtained by applying the inverse of f to $VoteCount(V^*)$. Note that this is true whether the measure used is "Shannon entropy" or computational entropy, because a computational integrity of one means that V^Σ can be determined from $VoteCount(V^*)$ in the computational model).

4.3 Privacy

Election privacy is the property that the election system should not reveal information about the values of individual or specific votes. Perfect privacy can be defined as in [1]. We state the definition almost verbatim here, except we ignore any vote information obtained from outside the system, and use the fact that $E = ElectionOutput(V^*)$; that is, the privacy definition assumes that the relationship between the output of the voting system and the individual votes is known. This makes for a stronger privacy requirement, and a weaker one, where E is assumed to not necessarily be $ElectionOutput(V^*)$, is covered in section 5.2.4. Note that *ElectionOutput* is not restricted to $VoteCount$, it includes any other information the system may reveal.

Definition 3 [1]: An election system provides **perfect privacy** if V^* is independent of $ElectionOutput(V^*)$, i.e.,

$$p_{V^*}(v^*) = p_{V^*|ElectionOutput(V^*)}(v^*; ElectionOutput(v^*))$$

for all $v^*, ElectionOutput(v^*)$.

Note, that, according to this definition, all election systems providing any information about V^Σ have imperfect privacy. This addressed through the concept of *maximal privacy* (Definition 5).

The measure of [1] is an appropriate measure of privacy loss, and we restate it here in normalized form.

Definition 4: The **amount of privacy loss**, $\mathfrak{L}$, of a voting system and process is

$$\mathfrak{L} = max_{p_{V^*}} \frac{\mathcal{I}(V^*; ElectionOutput(V^*))}{\mathcal{H}(V^*)}$$

where $p_{ElectionOutput(V^*)|V^*}$ is held fixed (it represents the system) and p_{V^*} varies.

$\mathfrak{L}$ ranges between zero (indicating that the election system reveals nothing at all about any ballots) and one (indicating that the election system reveals all ballots exactly), when the algorithm relating E to $ElectionOutput(V^*)$ is known. We do not distinguish between voluntary and involuntary privacy, but this distinction exists in the literature, in particular in [1] from which our definition is drawn.

Example 2: Consider a precinct with a single polling machine that provides VVPAT records on a paper reel which maintains the order of the vote. Suppose further that election officials maintain a record of who voted, in the order of arrival. Trivially, E consists of the ordered list of votes, and the ordered list of voters. Hence $\mathcal{H}(V^*|E) = 0$, and the privacy loss of this system is one.

4.4 Integrity/Privacy Relationship

The goals of integrity and privacy are not independent, and, furthermore, the literature states that perfect privacy and perfect integrity are not simultaneously achievable; see, for example, [9].

Example 3: Consider the fraudulent election: Candidate Alice is declared the winner independent of the votes. The integrity of the election is zero, and the privacy perfect, because the election output reveals no information at all about the vote.

Because perfect integrity is incompatible with perfect privacy except for the most trivial election, any election system must make tradeoffs between the two. As the purpose of the election is to obtain the vote count, the approach in the literature has been to require perfect integrity and the maximum privacy given that integrity is perfect.

Let us consider the case where perfect integrity is achieved; in other words, $VoteCount(V^*) = V^\Sigma$. The system that provides the most privacy while achieving perfect integrity, provides no more information about V^* other than V^Σ.

Definition 5: An election system is said to provide **maximal privacy** if V^* is conditionally independent of $ElectionOutput(V^*)$ after conditioning on V^Σ, i.e.,

$$p_{V^*|V^\Sigma}(v^*; v^\Sigma) = p_{V^*|V^\Sigma, ElectionOutput(V^*)}(v^*; v^\Sigma, ElectionOutput(v^*))$$

for all $v^*, v^\Sigma, ElectionOutput(v^*)$.

V^Σ denotes the legally required public vote count(s). For example, in the US, precinct-level vote counts are often required. In Arlington, VA, vote counts at the machine level are part of public record. Thus, while the election may be one for President of the United States (for which vote counts at the state level contribute to the determination of the election outcome) V^Σ will denote the vote count over a precinct or a machine, as required by law. The notion of maximal privacy allows us to determine if there are sources of privacy leakage other than those mandated by election procedures; in particular, it allows us to determine if the voting machine or voting system is leaking information.

5 Verifiable Elections

So far, we have assumed a model of perfect trust in the voting system, that is, we have assumed that the algorithm $VoteCount$ of the voting system is known. We have not discussed, however, how we know that the voting system is actually following algorithm $VoteCount$, that is, we have not studied

the verification requirements of the system. This section addresses verification.

Consider E as being divided into two elements:

1. $\widehat{V^\Sigma}$, a purported vote count.
2. P, information that can be used to prove $\widehat{V^\Sigma} = VoteCount(V^*)$, that is, information that is used to prove that the claimed algorithm, $VoteCount$, was followed.

5.1 The Definition of Verifiability

Perfect verifiability occurs when there is no uncertainty whether the election system used $VoteCount$ to produce $\widehat{V^\Sigma}$, given the correctness proof provided by the system. We denote by T the random variable representing the truth of $VoteCount(V^*) = \widehat{V^\Sigma}$.

Definition 6: An election system is **perfectly verifiable** when $\mathcal{H}(T|P) = 0$.

Equivalently, $\mathcal{I}(T;P) = \mathcal{H}(T)$. Note that an election system may be perfectly verifiable even if the proof shows that the vote count was not obtained through its declared algorithm; we simply require that a system provide enough information to check its result.

Definition 7: The **verifiability measure** of an election system is

$$\mathfrak{V} = min_{p_{V^*}} \frac{\mathcal{I}(T;P)}{\mathcal{H}(T)}$$

Notice that we do not define verifiability as a measure of the uncertainty in $V^\Sigma = \widehat{V^\Sigma}$; that is, we do not define it in terms of how close the purported vote count is to the true one. Such a definition would be a combination of our definition of verifiability (which connects the purported vote count to that achieved by the declared algorithm) and our definition of integrity (which connects the output of the declared algorithm to the correct vote count). For verifiability, we simply require the system to demonstrate that it is indeed using the declared algorithm, which was quantified by its integrity measure. In other words, we are not requiring that the election channel introduce no noise at all, but that it introduces no more noise than does the declared algorithm $VoteCount$; if Z represents the noise introduced in $\widehat{V^\Sigma}$ by $VoteCount$, then T represents the truth of $\widehat{V^\Sigma} - V^\Sigma = Z$.

Example 4: Consider an election system that makes the following common "black box" DRE claim:

- During polling, V^* (and nothing else) goes in.
- After polling, V^Σ (and nothing else) comes out.

Here, $VoteCount(V^*) = V^\Sigma$, and the integrity is perfect. However, $E = \widehat{V^\Sigma}$; that is, $P = \emptyset$. Hence, $\mathcal{H}(T|P) = \mathcal{H}(T)$ and $\mathfrak{V} = 0$.

5.2 Means of Verification

There are several means used to obtain verification that the system uses the declared algorithm, $VoteCount$.

5.2.1 Institutional Trust

In practice, the voter probably has some non-trivial amount of trust in the Election Authority, unless she actively believes in a conspiracy to falsify elections. We categorize this as Institutional Trust, the "baseline" willingness of an observer to believe T is *True*. We do not, however, use this reduction in uncertainty, obtained without a provision of proof, as a contribution to the system's verifiability. This reduction in uncertainty is not caused by the specific election system or procedures, but by the voters' willingness to trust the system, obtained, perhaps, from other interactions with related authorities and systems– for example, the government, and the authentication system of the Department of Motor Vehicles. If Institutional Trust is used in any way to characterize verifiability, it would influence the denominator in the definition of the verifiability measure, by determining the prior probability distribution on T; for example, perhaps $p_T(t) = 0.8$ when $t = true$.

5.2.2 System Trust

P may consist of information about the election system itself, such as physical security procedures, source code, circuit diagrams, and / or parallel audit procedures and results. This sort of information is information about the **type** of election being run. For example, if the election is being run on a Brand X Voting Machine, P would include information about the general reliability of Brand X Voting Machines; if parallel audits are being run, P could include information about the set of Brand X Voting Machines delivered to the Election Authority. Information about the election system itself is not, however, information about the **specific** voting machine or election in question

Information about the election system serves to reduce $\mathcal{H}(T)$ for a specific set of voting machine(s) used in the election in the following way. Let t be the value of T for the specific set of machines for the election, i.e. $t = (\widehat{v^\Sigma} = VoteCount(v^*))$. Any

testing of the voting machine(s) before election day, and any testing of similar machines on election day or at another time, involves determining the values of Y, a similar, random variable, related to T. That is, $Y = (\widehat{V^{\Sigma}} = VoteCount(V^*))$ for the same machine(s) on another day, or for similar machine(s) on the same/another day. Repeatedly sampling the value of Y, say N times, enables the statistical characterization of the distribution of Y. That is, it enables an estimation of the probability with which the machines tested are using $VoteCount$ at the time they are tested. The larger the value of N, the lower the uncertainty in the estimation of the probability distribution of Y. If the machines are always using $VoteCount$, the value of Y will always be *true*, however, the uncertainty in Y is zero only asymptotically.

Clearly, $\mathcal{H}(Y) \rightarrow 0$ does not imply $\mathcal{H}(T) \rightarrow 0$; in fact,

$$\lim_{N \rightarrow 0} \mathcal{H}(Y) = 0 \Rightarrow \lim_{N \rightarrow 0} \mathcal{H}(T) = \mathcal{H}(T|Y)$$

When the only proofs provided are those of the statistical behavior of similar systems, the asymptotic value of the verifiability is

$$\lim_{N \rightarrow 0} \mathfrak{V} = min_{p_{V^*}} \frac{\mathcal{I}(T;Y)}{\mathcal{H}(T)}$$

It can be increased by improving the testing and more tightly coupling the tested systems to those used on election day, that is, by increasing $\mathcal{I}(T;Y)$.

5.2.3 Audit Trail

Even if every single audit or test ever conducted using the election system in question has indicated that the system is trustworthy, there will still remain some uncertainty about whether the specific election under consideration has been compromised or was erroneous, because how can one ever know that the *next* election will be the one to be compromised? Information about the actual, specific election under examination can be included in P. We will name the portion of P that contains information about V^* (even indirect information such as bits of random number seeds, cryptographic keys, etc., that have no significant impact on computational security but do impact the information-theoretic privacy of V^*) the Audit Trail.

5.2.4 Weak Privacy

We define the weak privacy of a system as the privacy provided when E is a random variable that is not necessarily $ElectionOutput(V^*)$; the amount of privacy then depends on the uncertainty in E.

Definition 8 (this is identical to the definition in [1]: An election system provides **weak privacy** if V^* is independent of E, i.e.,

$$p_{V^*}(v^*) = p_{V^*|E}(v^*; e)$$

for all v^*, e.

The corresponding normalized measure is:

Definition 9: The **amount of weak privacy loss**, $\mathfrak{L}_w$, of a voting system and process is

$$\mathfrak{L}_w = max_{p_{V^*}} \frac{\mathcal{I}(V^*;E)}{\mathcal{H}(V^*)}$$

where $p_{E|V^*}$ is held fixed (it represents the system).

Like $\mathfrak{L}$, $\mathfrak{L}_w$ ranges between zero (indicating that the election system reveals nothing at all about any ballots) and one (indicating that the election system reveals all ballots exactly). However, it is typically smaller than $\mathfrak{L}$ because the output of the election system, E, is not necessarily proven to be $ElectionOutput$, hence its values contain more uncertainty than when it was assumed to be exactly $ElectionOutput$.

5.2.5 Completeness

No information other than System Trust information and Audit Trail information can reduce the uncertainty in T. This is because System Trust is information about the distribution of random variable T, and the Audit Trail is information about the specific value of this variable, $t = (\widehat{v^{\Sigma}} = VoteCount(v^*))$, which depends only on the known value $\widehat{v^{\Sigma}}$, the known algorithm $VoteCount$, and the value v^*.

5.2.6 Parallel Testing/System Trust and Audit Trails

Example 5 demonstrates that a finite amount of information about the specific election, $\mathcal{H}(V^*)$, can lead to perfect verifiability. On the other hand, an infinite number of parallel audit style tests are required to achieve that result. Further, assuming that every ballot counts equally, every bit (for example) of information about V^* contains the same amount of information about V^{Σ}; that is, a single bit of information about V^* reduces $\mathcal{H}(V^{\Sigma})$ by the same amount. On the other hand, every parallel test used to characterize the statistical behavior of the voting system reduces $\mathcal{H}(X)$ by a smaller amount than did the previous one. Thus parallel testing provides a more inefficient way of obtaining the same amount of verifiability.

On the other hand, parallel testing does not obtain its information from V^*, hence it does not reduce

privacy. System Trust/parallel audit information is information about the distribution of (supposedly) identical systems under similar conditions (including, importantly, different input). It does not include any information about V^* or any V_i. Audit Trail information is directly opposed to privacy by definition, since it is information about V^*.

5.3 Usability

In this section, we illustrate how our model may be used to examine usability. Not being usability experts, we are not able to characterize completely and in detail, the usability measures themselves. We present this material here to illustrate that the information-theoretic model does not need to be restricted to the understanding of integrity, privacy and verifiability.

Consider the perfect vote for voter i – the vote she intends to cast, V_i. Consider the vote recorded by the user interface, random variable V_i'. For various reasons – a bad ballot design, a user interface inaccessible to a person of her abilities, etc. – V_i' may not be identical to V_i. Unless the usability of the system is the worst possible, however, V_i' will not be independent of V_i. The dependence on V_i will be a function of the categories the voter falls in: perhaps the user interface is more difficult to use for a person with visual handicaps than it is for one without; perhaps it is more difficult to use for a person whose native tongue is not English. Thus the user interface may be characterized as a communication channel carrying the input V_i from the voter to provide the output V_i' as seen by the polling machine. $p_{V_i'|V_i}(v_i'; v_i)$ characterizes the communication channel, and perfect usability occurs when $V_i' = V_i$ with probability one for all values of i.

Definition 10: A user interface provides **perfect usability** when $Pr[V_i' = V_i] = 1 \; \forall \; i$.

When this is not so, the *usability* of the interface is the ratio of the information in V_i' about V_i, to the information in V_i.

Definition 11: The **usability measure** of the interface for voter i is

$$\mathfrak{U} = \frac{\mathcal{I}(V_i'; V_i)}{\mathcal{H}(V_i)}$$

Note that $\mathfrak{U} = 1$ does not imply perfect usability, it simply implies that V_i' contains all the information necessary to determine V_i. One can imagine a voting system that obtains all the information necessary to determine V_i but does not attempt to to determine it. In other words, usability does not necessarily imply that ballots, once properly cast as intended will be properly recorded as cast. Whether these are recorded as cast or not is examined while measuring verifiability. We do not address the issue of usability further, and assume perfect usability in the rest of this paper, that is, $V_i = V_i'$.

6 Open Questions

In this section, we present some questions that this model will allow us to address in future research, along with some speculation about the results.

6.1 Verifiability and Privacy

We have described the tradeoffs that exist between integrity and privacy. In addition, we have noted that tradeoffs must exist between Audit Trail verifiability and privacy (because that form of verifiability necessarily includes information about V^* beyond the information in $\widehat{V^\Sigma}$. On the other hand, System Trust verifiability does not impact privacy (except indirectly, by decreasing the uncertainty that $\widehat{V^\Sigma} = V^\Sigma$) because it reveals no information about the actual election in question.

We note a fundamental difference between System Trust and Audit Trail verifiability. System Trust verifiability draws elections to audit from an infinite-size population of "potential elections" that could be performed with the election system in question. As more and more such audits are performed, the probability that an unfair or otherwise incorrect election exists in that population asymptotically approaches zero. However, when using System Trust information to verify a specific election, one must consider the chance that the election system "knows" (via some secret signal or switch) the difference between an audit and a "live" election. In other words, System Trust verifiability does not address the possibility that the population of auditable elections is not the same as the population of "live" elections. Thus, the uncertainty in the correctness of a specific "live" election can only be reduced asymptotically to a non-zero value τ. On the other hand, Audit Trail verifiability can reduce that uncertainty to zero, however, this will occur, typically, at a privacy cost.

It seems to us that by combining System Trust verifiability with Audit Trail verifiability, we can achieve better verifiability and privacy for a specific live election than by using either alone. We speculate that one could use System Trust verifiability (i.e.: parallel audits) to reduce the uncertainty in the correctness of the election to some distance from τ, and also use Audit Trail verifiability (in effect) to reduce τ by some amount. In other words, one could use Audit Trail verifiability to "spend" a cer-

tain amount of privacy to get a certain amount of verifiability, and then (possibly) use System Trust verifiability to augment that. However, it is not yet clear to what extent the two forms of verifiability actually can complement each other.

6.2 Cast As Intended, Recorded as Cast, and Counted as Recorded

We have described the difference between voluntary privacy and involuntary privacy, but we have only scratched the surface of the differences between these two concepts in our model. Similarly, we have not drawn a clear distinction between voter verifiability and public verifiability. To examine these differences in more detail we must focus on other stages of the election process than the "counted as recorded"[1] stage.

It appears to us that there are actually three channels involved in an election system: the "cast as intended" channel that converts voter intentions to cast ballots, the "recorded as intended" channel that converts those cast ballots into recorded ballots, and the "counted as cast" channel that converts those recorded ballots into a vote count. This paper has focused on the "counted as cast" channel. We speculate that the other two channels are easily added to our model, and will have similar properties of throughput and verifiability versus privacy.

In addition, we note that the above division into three stages is natural in that it is according to the type of verifiability that is possible in each one. Only the voter can verify that her intentions were accurately written on the ballot to be cast; the "cast as intended" stage is inherently not publicly verifiable. Similarly, in order to verify that the cast ballots were correctly recorded, (or to raise an objection that some set of cast ballots were not correctly recorded), some voters must reveal some form of receipt or other information about their cast ballots, impacting their privacy. Thus, we speculate that public verifiability of the "recorded as cast" stage requires some "expenditure" of voluntary (information-theoretic) privacy, and is not possible if all voters refuse to cooperate. However, voter verifiability of this stage (as well as of the "cast as intended" stage) apparently does not impact privacy. Since we have focused on the "counted as recorded" stage in this paper, we have been unable to differentiate between voter verifiability and public verifiability (because these are essentially identical for this stage) and between voluntary and involuntary privacy. Future examinations of the first two channels should provide more insight into these differences.

[1] We first heard these terms used by Alan Sherman in a lecture.

6.3 Contextual Information

In this paper, we have supposed that elections occur in a vacuum, in the sense that the only source of information about voters and their votes is the election results. Of course, this is not true. A voter's demographic information and other political behaviors and activity (such as political party affiliation or PAC donations) that would be available publicly (or otherwise) might well reduce an observer's uncertainty about that voter's (intended) vote. Accordingly, the act of actually voting, and making information about that vote available in the election results, would have a less severe effect on that voter's privacy than it would in the "vacuum" election. Thus, a system that leaks more information might be acceptable when such contextual information is taken into account than otherwise.

Future research into this model will provide more insight into this issue. We speculate that almost all the terms in the existing equations will simply become dependent or conditional on the contextual information, and that this might easily be represented through a different initial probability distribution (prior distribution) on the votes, V^*. The interesting question here is the extent to which various measures are affected by the contextual information; the conditioning on such information is likely to affect the normalizing terms of the integrity, privacy, and verifiability metrics (for example) quite differently.

7 Conclusions

We have presented the beginnings of an information-theoretic approach to rating voting systems for integrity, privacy and verifiability. We have used this framework to show that tradeoffs exist between integrity and privacy and between verifiability and privacy, and the propose a few directions of future research.

References

[1] Lillie Coney, Joseph L. Hall, Poorvi L. Vora, and David Wagner. Towards a privacy measurement criterion for voting systems. In *National Conference on Digital Government Research*, May 2005.

[2] F. G. Conrad. Usability and voting technology. Technical report, Bureau of Labor Statistics, xxxx.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[4] E. Gerck. Voting system requirements. In *WOTE*, 2001.

[5] David Jefferson. Requirements for electronic and internet voting systems in public elections. In *WOTE*, 2001.

[6] D. W. Jones. Evaluating voting technology. http://www.cs.uiowa.edu/ jones/voting/uscrc.html, 2001.

[7] D. W. Jones. End-to-end standards for accuracy in paper-based systems. http://www.cs.uiowa.edu/ jones/voting/west02/, 2002.

[8] Douglas W. Jones. Threats to voting systems. *NIST Workshop on Threats to Voting Systems*, October 2005.

[9] Aggelos Kiayias and Moti Yung. Self-tallying elections and perfect ballot secrecy. In David Naccache and Pascal Paillier, editors, *Public Key Cryptography — 5th International Workshop on Practice and Theory in Public Key Cryptosystems*, volume 2274 of *Lecture Notes in Computer Science*, pages 141–158, Paris, France, 2002. Springer.

[10] National Institute of Standards and Technology (NIST). Developing an analysis of threats to voting systems: Workshop summary.

[11] National Institute of Standards and Technology (NIST). Workshop on threats to voting systems. http://vote.nist.gov/threats/, October 2005.

[12] Michael Ian Shamos. Electronic voting - evaluating the threat. *Computers, Freedom and Privacy*, 1993.

[13] WEST 2002 workshop on election standards and technology. http://www.vote.caltech.edu/west02/presentations.html, 2002.

[14] WOTE01 workshop on trustworthy elections. http://www.vote.caltech.edu/wote01/, 2002.

[15] A. C. Yao. Theory and applications of trapdoor functions. *Proc. of the 23rd Annual IEEE Symposium on Foundations of Computer Science*, 1982.