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

May 8, 2006

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.

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 use of the model is illustrated by comparing a few common voting models with respect to integrity and privacy.

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 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.

∗These authors supported in part by NSF SGER
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 there should be a direct correspondence between the published and certified election results and the actual cast ballots. That is, the results 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.

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 these 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 values written on it); \( V_i \in \mathcal{V} \), 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 \( V^\Sigma \) 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, ...” \( V^\Sigma \) is therefore also a discrete random variable, but it is a deterministic function of \( V^* \).

Let \( \hat{V}^\Sigma \) 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 \( \hat{V}^\Sigma \) is a part of \( E \).) Assume that the voting system declares two algorithms, i.e. two sets of well-defined steps, \( \text{VoteCount} \) and \( \text{ElectionOutput} \), that, when applied to the votes, \( V^* \),
produce $\hat{(V^*)}$ and $E$ respectively. When VoteCount and ElectionOutput are known, we represent $\hat{V^*}$ and $E$ by $\text{VoteCount}(V^*)$ and $\text{ElectionOutput}(V^*)$ respectively. Note that this does not imply that either of VoteCount and ElectionOutput is necessarily deterministic, simply that $\hat{V^*}$ 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

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

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 $H(X|Y)$ is the uncertainty in $X$ if $Y$ is known,

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

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

$$I(X;Y) = H(X) - H(X|Y)$$

and it can be shown that:

$$I(X;Y) = I(Y;X)$$

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” 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 the use of only Shannon entropy in all cases.

### 4.2 Usability

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 1:** 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 2:** The **usability measure** of the interface for voter \( i \) is

\[
\Omega = \frac{I(V'_i; V_i)}{H(V_i)}
\]

Note that \( \Omega = 1 \) 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 determine it. We do not address the issue of usability further, and assume perfect usability in the rest of this paper, that is, \( V_i = V'_i \).

### 4.3 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, \( V^\Sigma \). Unlike previous definitions of integrity, ours does not include any notion of certainty in the election output. This notion is covered by the property of verifiability (see section 5).

Election integrity may be defined more precisely as follows:

**Definition 3:** An election system provides **perfect integrity** if \( VoteCount(V^*) = V^\Sigma \).

Even if the system does not provide perfect integrity, the uncertainty in \( V^\Sigma \) is generally reduced on knowledge of \( VoteCount(V^*) \). The reduction in uncertainty, \( I(V^\Sigma; VoteCount(V^*)) \) 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 \( H(V^\Sigma) \) (indicating perfect integrity). One could use a normalized value of this reduction in uncertainty to measure election integrity. However, there is usually information about an individual’s vote that comes from sources other than the election system (from bumper stickers and exit polls, for example). In a system with imperfect integrity, we would prefer to measure the integrity as the **least** uncertainty reduction over all possible values of the external information.

Let \( S \) be a random variable denoting the information obtained from sources other than the election system, for all the voters. Because \( S \) denotes information about individual votes, \( V^* \), it also denotes information about the tally, \( V^\Sigma \). We are interested in the worst case fractional reduction in uncertainty of the tally, conditioned on the externally-obtained information, \( S \).

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

\[
J = \min_{p_{V^*, S}} \frac{I(V^\Sigma|S, VoteCount(V^*))}{H(V^\Sigma|S)}
\]

where \( p_{VoteCount(V^*)|V^*} \) is held fixed (it represents the system) and \( p_{V^*, S} \) varies.

**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 $V^\Sigma$ 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.

**Example 2:** This example illustrates the impact of $S$ on the amount of information revealed by $VoteCount(V^\ast)$. Consider a system which counts all votes except those from a certain district. The votes from this district are simply dumped into the river. Let $V^\Sigma_{D_1}$ denote the votes, from District 1, that are counted, and $V^\Sigma_{D_2}$ the other votes, from District 2. Further, let $V^\Sigma_{D_1}$ and $V^\Sigma_{D_2}$ denote the tallies over the two districts, and assume they are independent as random variables. Then:

$$VoteCount(V^\ast) = V^\Sigma_{D_1}$$

and the uncertainty in the vote count of District 2 is not changed by knowledge of the system’s vote tally:

$$\mathcal{H}(V^\Sigma_{D_2}|VoteCount(V^\ast)) = \mathcal{H}(V^\Sigma_{D_2})$$

Hence, any information on $V^\Sigma_{D_2}$ is obtained through $S$.

**Case 1:** $S$ consists only of information on $V^\Sigma_{D_1}$. Hence the reduction in tally uncertainty due to knowledge of $VoteCount(V^\ast)$ is:

$$\mathcal{I}(V^\Sigma_{D_1}|S, VoteCount(V^\ast)) = \mathcal{H}(V^\Sigma_{D_1}|S) - \mathcal{H}(V^\Sigma_{D_1}|S, V^\Sigma_{D_2})$$

**Case 2:** $S$ consists only of information on $V^\Sigma_{D_2}$. Hence knowledge of $VoteCount(V^\ast)$ reduces completely the uncertainty in $V^\Sigma_{D_1}$:

$$\mathcal{I}(V^\Sigma_{D_1}|S, VoteCount(V^\ast)) = \mathcal{H}(V^\Sigma_{D_1}|S) - \mathcal{H}(V^\Sigma_{D_1}|S, V^\Sigma_{D_2}) = \mathcal{H}(V^\Sigma_{D_1})$$

The system provides more information in Case 2 than in Case 1.

In an extreme example of Case 1, when $S$ provides enough information to determine $V^\Sigma_{D_1}$ completely, i.e. $\mathcal{H}(V^\Sigma_{D_1}|S) = 0$, the reduction in vote tally uncertainty due to knowledge of $VoteCount(V^\ast)$ is zero. Hence the integrity of this system is zero, if it is possible for $\mathcal{H}(V^\Sigma_{D_1}|S)$ to be zero.

We assume, wlog, that an integrity value of one implies perfect integrity – that is, that $VoteCount(V^\ast)$ 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 $VoteCount(V^\ast)$.) Note that this is true whether the measure used is “Shannon entropy” or computational entropy, because a computational integrity of one means that $V^\Sigma$ can be determined from $VoteCount(V^\ast)$ in the computational model).

**4.4 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 use the fact that $E = ElectionOutput(V^\ast)$, 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 $ElectionOutput(V^\ast)$, is covered in section 5.2.5. Note that $ElectionOutput$ is not restricted to $VoteCount$, it includes any other information the system may reveal.
**Definition 5 [1]:** An election system provides **perfect privacy** if \( V^* \) is conditionally independent of \( \text{ElectionOutput}(V^*) \) after conditioning on \( S \), i.e.,

\[
p_{V^*|S}(v^*; s) = p_{V^*|S,\text{ElectionOutput}(V^*)}(v^*; s, \text{ElectionOutput}(v^*))
\]

for all \( v^*, s, \text{ElectionOutput}(v^*) \).

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

**Definition 6:** The amount of privacy loss, \( \mathcal{L} \), of a voting system and process is

\[
\mathcal{L} = \max_{p_{V^*}, S} \frac{I(V^*; S, \text{ElectionOutput}(V^*))}{\mathcal{H}(V^*|S)}
\]

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

\( \mathcal{L} \) 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 \( \text{ElectionOutput}(V^*) \) is known.

**Example 3:** 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.

### 4.5 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]. Because our model, unlike past ones except for [1], includes information from sources external to the voting system, we qualify this statement slightly in our model. Perfect integrity and perfect privacy are not simultaneously achievable unless the election is completely redundant – that is, unless the information from external sources is sufficient to determine the vote count exactly.

**Theorem 1:** Perfect privacy and perfect integrity are achieved simultaneously only if \( \mathcal{H}(V^\Sigma|S) = 0 \).

**Proof:** See Appendix.

**Example 4:** 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, \( \text{VoteCount}(V^*) = V^\Sigma \). The system that provides the most privacy while achieving perfect integrity, provides no more information about \( V^*|S \) other than \( V^\Sigma|S \).
Definition 7: An election system is said to provide **maximal privacy** if \( V^* \) is conditionally independent of \( \text{ElectionOutput}(V^*) \) after conditioning on \( S \) and \( V^\Sigma \), i.e.,

\[
p_{V^*|S,V^\Sigma}(v^*; s, v^\Sigma) = p_{V^*|S,V^\Sigma,\text{ElectionOutput}(V^*)}(v^*; s, v^\Sigma, \text{ElectionOutput}(v^*))
\]

for all \( v^*, s, v^\Sigma, \text{ElectionOutput}(v^*) \).

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 \( \text{VoteCount} \) of the voting system is known. We have not discussed, however, how we know that the voting system is actually following algorithm \( \text{VoteCount} \), 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 \( \widehat{V}^\Sigma = \text{VoteCount}(V^*) \), that is, information that is used to prove that the claimed algorithm, \( \text{VoteCount} \), was followed.

5.1 The Definition of Verifiability

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

Definition 8: An election system is **perfectly verifiable** when

\[
\mathcal{H}(T|S, P) = 0.
\]

Equivalently, \( I(T|S; P) = \mathcal{H}(T|S) \). 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 9: The **verifiability measure** of an election system is

\[
\mathcal{V} = \min_{p_{V^*}, S} \frac{I(T|S; P)}{\mathcal{H}(T|S)}
\]

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.

Example 5: Consider an election system that makes the following common “black box” DRE claim:

- During polling, \( V^* \) (and nothing else) goes in.
- After polling, \( V^\Sigma \) (and nothing else) comes out.
Here, $VoteCount(V^*) = V^\Sigma$, and the integrity is perfect. However, $E = \hat{V}^\Sigma$; that is, $P = \emptyset$. Hence, $\mathcal{H}(T|P) = \mathcal{H}(T)$ and $\mathcal{U} = 0$.

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 Trust in a Reasonable Result

We have also assumed that $S$ denotes information about an individual’s vote that comes from sources other than the election system. An observer hence has an estimate of the vote count, in the form of a probability distribution function, $p_{v^\Sigma|S}(v^\Sigma; s)$, that is independent of the vote tally provided by the system. The vote tally provided by the system, $\hat{V}^\Sigma$, is viewed in the context of $S$, and the more reasonable it seems, the more the voter will trust it without any other information from the system. This, however, is also not due to any proof provided by the system, and is hence not included in the measurement of system verifiability. As with Institutional Trust, if Trust in a Reasonable Result is used in any way to characterize verifiability, it would influence the denominator in the definition of the verifiability measure; that is, it would influence $\mathcal{H}(T|S)$.

5.2.3 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 $\mathcal{H}(T|S)$ 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. $t = (v^\Sigma = VoteCount(v^*))$. 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, $Y = (\hat{V}^\Sigma = VoteCount(V^*))$ 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 goes to zero only asymptotically.

Clearly, H(Y) → 0 does not imply H(T|S) → 0; in fact,

\[ \lim_{N \to 0} H(Y) = 0 \Rightarrow \lim_{N \to 0} H(T|S) = H(T,S,Y) \]

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

\[ \lim_{N \to 0} \mathcal{V} = \min_{p_{V^*},S} \frac{I(T|S;Y)}{H(T|S)} \]

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 I(T|S;Y).

5.2.4 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.

Example 7: Consider the case where P simply equals V*, that is, all votes and corresponding voter names are made available on a website. VoteCount(V*) = VΣ, and E = ElectionOutput(V*). This election provides no privacy, because H(V*|ElectionOutput(V*)) = 0. It provides perfect integrity, because VoteCount(V*) = VΣ. It has the potential to provide perfect verifiability, because any voter can complain if their ballot was not included correctly by looking it up; this is how legislatures vote. In a model where all voters do not check that their votes were correctly recorded, the verifiability measure is

\[ \mathcal{V} = 1 - \max_{p_{V^*},S} \frac{H(V^*_\text{unchecked} = V^*_\text{Σ} - V^*_\text{checked}|S,V^*_\text{checked})}{H(T|S)} \]

where V^*_\text{unchecked} is the tally of the unchecked votes, and V^*_\text{checked} that of the checked votes. V^*_\text{checked} is the set of checked votes.

\[ \mathcal{V} = 1 - \frac{H(V^*_\text{unchecked}|S)}{H(T|S)} \]

Note, however, that if verifiability is an issue – that is, if it is not certain that the public version of V* output by the system is indeed the correct value of V* till the voter checks it – then additional privacy is provided when voters do not check their votes. If one has some level of System Trust, however, the privacy provided by not checking one’s vote is less than perfect, because a reasonably trustworthy system is more likely than not to have provided the correct vote. Further, if many voters check their votes and claim they are correctly posted, it is even more likely that an unchecked vote is correctly posted; that is, this privacy depends on the verifiability and on the System Trust.
5.2.5 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 10** (this is identical to the definition in [1]): An election system provides **weak privacy** if $V^*$ is conditionally independent of $E$ after conditioning on $S$, i.e.,

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

for all $v^*, s, e$.

The corresponding normalized measure is:

**Definition 11:** The **amount of weak privacy loss**, $L_w$, of a voting system and process is

$$L_w = \max_{p_{V^*}, S} \frac{I(V^*; S, E)}{H(V^*|S)}$$

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

Like $L$, $L_w$ 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 $L$ 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 $ElectionOutput$.

The above example demonstrates that a finite amount of information about the specific election, $H(V^*)$, 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 $V^\Sigma$; that is, a single bit of information about $V^*$ reduces $H(V^\Sigma)$ by the same amount. On the other hand, every parallel test used to characterize the statistical behavior of the voting system reduces $H(X)$ by a smaller amount than did the previous one. However, parallel audits do not reveal information about individual votes.

5.2.6 Completeness

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

5.3 Verifiability and Privacy

In this section, we discuss the impact of verifiability on privacy and weak privacy. System Trust 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 $V_i$. If the tests establishing system trust are conducted with a set of votes
that is representative of $V^*$, then system trust information may impact privacy by contributing to
information in $S$.

Audit Trail information is directly opposed to privacy by definition, since it is information about
$V^*$. We divide audit trail information into two parts: $P_v$ – proof information that requires voter
participation (such as checking that the displayed vote is, indeed, the correct one), and $P_s$ – proof
information that does not require voter participation, but can be accomplished by the system and
the election authority itself, such as the checking of cryptographic hashes. Notice that

$$L_w = \frac{I(V^*|S;E)}{H(V^*|S)} = \frac{H(V^*|S) - H(V^*|S,ElectionOutput(V^*)) - H(ElectionOutput(V^*)|S,E)}{H(V^*|S)}$$

$$= L - \frac{H(ElectionOutput(V^*)|S,E)}{H(V^*|S)}$$

and $L_w \leq L$ as long as all the proof is not provided, that is, $H(ElectionOutput(V^*)|S,E) \neq 0$.

6 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. This is work in progress. We
hope to use it to rank the four models to be examined at the workshop in time for the workshop.

References

[1] Lillie Coney, Joseph L. Hall, Poorvi L. Vora, and David Wagner. Towards a privacy measurement
2005.
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,
[10] National Institute of Standards and Technology (NIST). Developing an analysis of threats to voting
systems: Workshop summary.
Theorem 1: Perfect privacy and perfect integrity are achieved simultaneously only if $\mathcal{H}(V^\Sigma |S) = 0$.

Proof:

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

$$\Rightarrow \mathcal{H}(V^* | S) = \mathcal{H}(V^* | S, ElectionOutput(V^*)) \leq \mathcal{H}(V^* | S, VoteCount(V^*))$$

because $VoteCount(V^*)$ is part of $ElectionOutput(V^*)$, that is, is completely determined by $ElectionOutput(V^*)$, and cannot contain more information than $ElectionOutput(V^*)$. Further,

$$\mathcal{H}(V^* | S, VoteCount(V^*)) = \mathcal{H}(V^* | S, V^\Sigma)$$

because perfect integrity requires that $V^\Sigma = VoteCount(V^*)$. But uncertainty does not increase with the addition of information, hence, from the above equations,

$$\mathcal{H}(V^* | S) \geq \mathcal{H}(V^* | S, V^\Sigma) \Rightarrow \mathcal{H}(V^* | S) = \mathcal{H}(V^* | S, V^\Sigma)$$

Further,

$$\mathcal{H}(V^* | S) + \mathcal{H}(V^\Sigma | V^*, S) = \mathcal{H}(V^* | S, V^\Sigma) + \mathcal{H}(V^\Sigma | S)$$

and $\mathcal{H}(V^\Sigma | V^*, S) = 0$.

$$\Rightarrow \mathcal{H}(V^\Sigma | S) = 0$$