2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Fraud Detection and Deterrence in Electronic Voting Machines: A Game-Theoretic Approach

Anuj S. Vora
Delft Center for Systems and Control, TU Delft
2628CD Delft, NL
a.vora@tudelft.nl

Ankur A. Kulkarni
Systems and Control Engineering, IIT Bombay
Mumbai-400076, India
kulkarni.ankur@iitb.ac.in

***Abstract*— We study a setting where a *detector* wishes to detect and deter adversarial manipulation in an electronic voting machine. An *adversary* tries to win the election by tampering the votes while obfuscating its manipulation. We pose this problem as a game between the detector and the adversary and characterize the equilibrium payoffs for the players and the asymptotic nature of these payoffs. We find that if the detector is too cautious, then in equilibrium the adversary wins with a probability *higher* than its prior probability of winning. We derive an expression for the deterrence threshold, i.e., the minimum level of false-alarm that the detector should endure so that the adversary is not any better off by the manipulation. With this, asymptotically, the detector can ensure that the probability of missed-detection becomes zero by appropriately adjusting the rate of decay of probability of false-alarm. But if this rate of decay is too 'fast', then the adversary can get an arbitrarily high probability of winning in spite of having a vanishing prior probability of winning. We then extend the results to a setting where the detector has incomplete information about the adversary.**

I. Introduction

An Electronic Voting Machine (EVM) is a digital voting machine that records the votes of individuals and is used in numerous countries during elections [1]. In an EVM, votes are recorded in a storage device by pressing a button allotted to a certain party or individual. Elections being one of the pillars of democracy, it is of interest to detect and deter any fraudulent manipulation. Suppose an investigating entity (*detector*) can only read the votes that are recorded in the EVM. We ask the following question – how well can this entity detect any fraudulent manipulation in the voting process? Is there a way to deter the manipulation by any adversary? And when does the adversary succeed?

EVM manipulation has a few peculiarities. The EVM can be hacked by an *adversary* who can replace the original votes recorded in it by any set of votes. This is in contrast with a paper ballot where the manipulation effort grows with the number of votes manipulated. In elections, usually an independent provision (say through social identity cards) tracks the *number* of votes cast. Thus, in any manipulation the adversary is constrained to keep the total votes constant.

Elections are also a sensitive and expensive affair. The detector faces a dilemma of balancing the real-world requirement of both catching fraud, but also of being very cautious and certain when calling for it. Frequent or indiscriminate raising of alarms does catch and deter frauds, but also leads to significant degradation in the public's faith in the electoral process, and damages the detector's credibility. The knowledge that the detector is faced with this dilemma makes it possible for an adversary to manipulate stealthily to improve its chances of winning. Motivated by this, we propose a game-theoretic model to address the above questions.

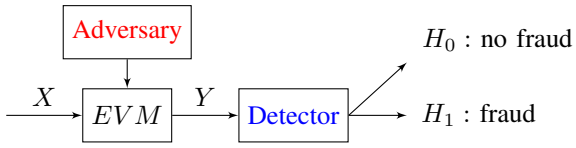

Fig. 1: A detector tries to detect fraud using EVM data

We consider a setting where an EVM can be manipulated by an adversary such that the recorded votes are modified. A detector can read the votes from the EVM and declare 'no manipulation' (H_0) or 'some manipulation' (H_1). This is depicted in Figure 1. When the detector declares H_1, the entire election is scrapped and the adversary's efforts, if any, are in vain. We assume that the votes are generated by a fixed distribution that is known by adversary and the detector.

We study this problem under a hypothesis testing-like framework where the detector encounters two types of errors based on its decisions – *false-alarm* and *missed-detection*. A false-alarm error occurs when the detector declares H_1 where there was no manipulation. A missed-detection error occurs when the detector declares H_0 when there was some manipulation. The adversary aims to modify the votes so that they lie in a certain winning set *and* that the detector decides H_0. We call this as the event of *winning*. We pose the problem as a game between the detector and the adversary where the detector minimizes the missed-detection probability subject to a bound on the false-alarm probability. The adversary maximizes the probability of winning.

We allow randomized strategies and study the problem under two settings. In the first setting, the detector perfectly knows the winning set. Although the game is a non-zero sum game, we show that under a non-degeneracy requirement, it reduces to a constant-sum game. We get the remarkable finding that if the false-alarm bound is below a certain threshold, then in any equilibrium, the adversary wins with a probability *higher* than its prior probability of winning. This leads us to define the notion of *deterrence threshold*,

This work was done when Anuj S. Vora was a postdoctoral fellow with Systems and Control group, IIT Bombay.

that is the minimum false-alarm error that the detector must bear to ensure that the posterior probability of winning is no more than the prior probability of winning. The deterrence threshold is given by $p_w(1-p_w)$, where p_w is the prior probability of winning. We then study the asymptotic nature of the equilibrium by assuming votes are generated i.i.d. according to a certain distribution, and determine conditions under which the probabilities of false-alarm and missed-detection approach zero and determine their rates of convergence. We show that despite the adversary bearing no cost for its manipulation, the detector can ensure that the missed-detection probability becomes zero by appropriately adjusting the rate of decay of false-alarm probability. We also show that if this rate is too 'fast', i.e., the detector is too cautious in making a false-alarm, then the adversary can get an asymptotically arbitrarily high probability of winning. This holds despite of a vanishing prior winning probability.

We finally study a setting where the detector does not exactly know the winning set of the adversary. We show that under a similar non-degeneracy requirement, the game reduces to a constant-sum game and the equilibria of the game are characterized by a linear program. We then present analogous conditions on the threshold for deterrence.

The adversarial hypothesis problem was introduced by Barni and Tondi in [2]. They study a setting where an adversary wishes to deceive a detector by manipulating the samples subject to a cost. Further discussion and variants can be found in [3]. The work closest to our setting is the work of Yasodharan and Loiseau in [4] where they study a game between an adversary and a detector where the adversary deceives by choosing the distribution for generating the samples. They show the existence of the Nash equilibrium and determine the asymptotic nature of the equilibrium. The adversarial setting is studied in sequential hypothesis testing problem by Zhang and Zou in [5], Jin and Lai in [6], Pan et al. in [7] and Cao et al. in [8]. Our setting where the adversary is interested in winning and faces no cost, has, to the best of our knowledge, not been studied before.

The problem is formulated in Section II. The reduction to a constant-sum game and its analysis is discussed in Section III. The asymptotics are discussed in Section IV. In Section V, we discuss a case when detector has imperfect information of the winning set. We conclude in Section VI.

II. Problem Formulation

A. Notation

All random variables are discrete and denoted by upper-case letters, say, X, Y, Z and their instances by x, y, z. Script letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ denote their alphabet spaces. P_{X} denotes the probability of the source. The upper-case letter Q, with an appropriate subscript, denotes the strategies of the detector and adversary. $\mathcal{P}(\mathcal{X})$ denotes the space of probability distributions on $\mathcal{X}$. For a sequence $x \in \mathcal{X}^n$, P_x denotes the empirical distribution where $P_x(i) := |\{k : x_k = i\}|/n$, $\forall i \in \mathcal{X}$. $\mathbb{P}(A)$ denotes the probability of an event or a set A computed under the underlying measure. $\mathrm{supp}(P)$ denotes the support of a distribution P. We write $\sum_{x,y\in\mathcal{C}^n} P_X(x)Q_{Y|X}(y|x)Q_{Z|Y}(H_0|y) = (P_X Q_{Y|X} Q_{Z|Y})(H_0)$, $\sum_{x\in\mathcal{C}^n} P_X(x)Q_{Y|X}(y|x) = (P_X Q_{Y|X})(y)$ and $\sum_{i\in\Theta,x,y\in\mathcal{C}^n} P_\theta(i)P_X(x)Q_{Y|X}(y|x)\times Q_{Z|Y}(H_0|y) = (P_\theta P_X Q_{Y|X,\theta} Q_{Z|Y})(H_0)$.

B. The EVM manipulation game

An EVM records the votes of n voters as n-length strings X from $\mathcal{C}^n$ where $\mathcal{C} = \{0, \ldots, q-1\}$ is a finite set of candidates. The votes are assumed to be generated i.i.d. according to $P_X(x) = \prod_i P_{\mathsf{X}}(x_i)$, $x \in \mathcal{C}^n$ for a fixed $P_{\mathsf{X}} \in \mathcal{P}(\mathcal{C})$. We denote the votes that are stored in the EVM as $Y \in \mathcal{C}^n$. In the absence of any interference, $Y = X$. In the presence of an adversary, there may be a manipulation such that Y may not equal X. Note that the i.i.d. nature of P_X and string structure of X are invoked for the asymptotics; they are not required for the one-shot formulation.

The detector, on observing Y, has to take a decision of no adversarial manipulation ($H_0 \cong Y = X$) or some adversarial manipulation ($H_1 \cong Y \neq X$). The decision of the detector is denoted as $Z \in \mathcal{H} = \{H_0, H_1\}$ and is chosen according to the conditional distribution $Q_{Z|Y} \in \mathcal{P}(\mathcal{C}^n|\mathcal{C}^n)$. The adversary can also observe the votes stored in the EVM and can alter them according to the distribution $Q_{Y|X} \in \mathcal{P}(\mathcal{C}^n|\mathcal{C}^n)$. The false-alarm error is the event $\{Y = X, Z = H_1\}$ and a missed-detection error is the event $\{Y \neq X, Z = H_0\}$. The respective probabilities are

$$\begin{aligned} P_F(Q_{Y|X}, Q_{Z|Y}) &= \mathbb{P}(Y = X, Z = H_1) \\ &= \sum_{x\in\mathcal{C}^n} P_X(x)Q_{Y|X}(x|x)Q_{Z|Y}(H_1|x), \\ P_M(Q_{Y|X}, Q_{Z|Y}) &= \mathbb{P}(Y \neq X, Z = H_0) \\ &= \sum_{x,y\in\mathcal{C}^n, y\neq x} P_X(x)Q_{Y|X}(y|x)Q_{Z|Y}(H_0|y). \end{aligned}$$

The adversary is concerned with a *winning set* $\mathcal{W} \subseteq \mathcal{C}^n$. The adversary wins in the event $\{Y \in \mathcal{W}, Z = H_0\}$. The probability of winning is then given as

$$\begin{aligned} P_W(Q_{Y|X}, Q_{Z|Y}) &= \mathbb{P}(Y \in \mathcal{W}, Z = H_0) \\ &= \sum_{x\in\mathcal{C}^n, y\in\mathcal{W}} P_X(x)Q_{Y|X}(y|x)Q_{Z|Y}(H_0|y). \end{aligned} \tag{1}$$

We formulate this problem as a simultaneous-move game between the detector and the adversary. We define

$$\begin{aligned} u_D(Q_{Y|X}, Q_{Z|Y}) &= P_M(Q_{Y|X}, Q_{Z|Y}), \\ u_A(Q_{Y|X}, Q_{Z|Y}) &= P_W(Q_{Y|X}, Q_{Z|Y}). \end{aligned}$$

The detector minimizes $u_D(Q_{Y|X}, Q_{Z|Y})$ while ensuring that $P_F(Q_{Y|X}, Q_{Z|Y})$ stays below a threshold $\gamma \in [0, 1]$. The adversary maximizes $u_A(Q_{Y|X}, Q_{Z|Y})$. We refer to this as the *EVM manipulation game* and solve for the NE.

Definition 2.1: A pair $(Q^*_{Y|X}, Q^*_{Z|Y})$ is a Nash equilibrium of the game if $P_F(Q^*_{Y|X}, Q^*_{Z|Y}) \leq \gamma$ and

$$\begin{aligned} u_A(Q^*_{Y|X}, Q^*_{Z|Y}) &\geq u_A(Q_{Y|X}, Q^*_{Z|Y}) \quad \forall\, Q_{Y|X}, \\ u_D(Q^*_{Y|X}, Q^*_{Z|Y}) &\leq u_D(Q^*_{Y|X}, Q_{Z|Y}) \quad \forall\, Q_{Z|Y}. \end{aligned}$$

This game has coupled constraints [9], and standard results on existence of equilibria [10] may not directly apply. However, we show that it can be analyzed by such techniques.

III. Nash Equilibrium of the Game

In this section, we determine a class of equilibrium strategies of the game. We then show that under a restriction on the strategies of the adversary, the game reduces to a constant-sum game whose saddle point corresponds to an equilibrium from this class. Define $f(\gamma, \overline{\mathcal{W}}) = \max\left\{1 - \gamma/\mathbb{P}(\overline{\mathcal{W}}), 0\right\}$.

Theorem 3.1: For all $\pi \in [0,1]$ and $\overline{\mathcal{W}} \subseteq \mathcal{W}$, the pair $(Q^*_{Y|X}, Q^*_{Z|Y})$ are a class of Nash equilibria where

$$Q^*_{Y|X}(y|x) = \begin{cases} 1 & y = x, x \in \overline{\mathcal{W}} \\ P_X(y)/\mathbb{P}(\overline{\mathcal{W}}) & y \in \overline{\mathcal{W}}, x \notin \overline{\mathcal{W}} \end{cases}, \quad (2)$$

$$Q^*_{Z|Y}(H_0|y) = \begin{cases} f(\gamma, \overline{\mathcal{W}}) & y \in \overline{\mathcal{W}} \\ \pi & y \notin \overline{\mathcal{W}} \end{cases}.$$

Further, $P^*_F(Q^*_{Y|X}, Q^*_{Z|Y}) = \gamma$, $P^*_M(Q^*_{Y|X}, Q^*_{Z|Y}) = (1 - \mathbb{P}(\overline{\mathcal{W}}))f(\gamma, \overline{\mathcal{W}})$, $P^*_W(Q^*_{Y|X}, Q^*_{Z|Y}) = f(\gamma, \overline{\mathcal{W}})$. In particular, a Nash equilibrium exists.

Proof: Let $Q^*_{Y|X}$ be as given in (2). Then any strategy $Q_{Z|Y}$ in response to $Q^*_{Y|X}$ must satisfy

$$\begin{aligned} P_F(Q^*_{Y|X}, Q_{Z|Y}) &= \sum_{x \in \overline{\mathcal{W}}} P_X(x)(1 - Q_{Z|Y}(H_0|x)) \\ &= \mathbb{P}(\overline{\mathcal{W}}) - \sum_{x \in \overline{\mathcal{W}}} P_X(x) Q_{Z|Y}(H_0|x) \leq \gamma. \quad (3) \end{aligned}$$

Since $\mathrm{supp}(Q^*_{Y|X}(.|x)) = \overline{\mathcal{W}}$ for all x and $Q^*_{Y|X}(y|x) = 0$ if $x \in \overline{\mathcal{W}}$ and $y \neq x$, we have $u_D(Q^*_{Y|X}, Q_{Z|Y}) = \sum_{x \notin \overline{\mathcal{W}}, y \neq x, y \in \overline{\mathcal{W}}} P_X(x) Q^*_{Y|X}(y|x) Q_{Z|Y}(H_0|y) = (1 - \mathbb{P}(\overline{\mathcal{W}})) \times \sum_{y \in \overline{\mathcal{W}}} P_X(y) Q_{Z|Y}(H_0|y)/\mathbb{P}(\overline{\mathcal{W}})$

$$\geq (1 - \mathbb{P}(\overline{\mathcal{W}}))\Big(1 - \gamma/\mathbb{P}(\overline{\mathcal{W}})\Big) = (1 - \mathbb{P}(\overline{\mathcal{W}}))f(\gamma, \overline{\mathcal{W}}). \quad (4)$$

The inequality in (4) follows from (3). For $Q^*_{Z|Y}$, we get $\sum_{y \in \overline{\mathcal{W}}} P_X(y) Q^*_{Z|Y}(H_0|y) = \mathbb{P}(\overline{\mathcal{W}}) f(\gamma, \overline{\mathcal{W}})$. Thus, $Q^*_{Z|Y}$ satisfies (3) and achieves the lower bound in (4) and hence is a best response to $Q^*_{Y|X}$. Fixing $Q^*_{Z|Y}$, for any $Q_{Y|X}$, we have $u_A(Q_{Y|X}, Q^*_{Z|Y}) = f(\gamma, \overline{\mathcal{W}}) \sum_{y \in \overline{\mathcal{W}}} (P_X Q_{Y|X})(y)$. Since the payoff is the same for all $Q_{Y|X}$ supported on $\overline{\mathcal{W}}$, $Q^*_{Y|X}$ is a best response to $Q^*_{Z|Y}$. ∎

In equilibrium the adversary tries to ensure that all the votes lie in the set $\overline{\mathcal{W}}$ while obfuscating the output. The detector declares a fraud (H_1) with uniform probability $\forall y \in \overline{\mathcal{W}}$. For $y \notin \overline{\mathcal{W}}$, it can declare fraud ($H_1$) with any probability.

Observe that taking $\overline{\mathcal{W}}$ such that $\mathbb{P}(\overline{\mathcal{W}}) \leq \gamma$ leads to the equilibrium, where $P^*_M(Q^*_{Y|X}, Q^*_{Z|Y}) = 0$ and $P^*_W(Q^*_{Y|X}, Q^*_{Z|Y}) = 0$. Also, the adversary does not gain by manipulating x when it lies in $\mathcal{W}$. Yet our formulation could lead to equilibria where such manipulation is performed. We view these equilibria as *degenerate* equilibria arising primarily because there is no cost associated with manipulation. Moreover, the highest payoff of the adversary in the above equilibria is $\max\{1 - \gamma/\mathbb{P}(\mathcal{W}), 0\}$ which occurs when $\overline{\mathcal{W}} = \mathcal{W}$. With this motivation and to avoid degenerate equilibria, we restrict the strategies of the adversary to

$$Q_A = \left\{Q_{Y|X} : Q_{Y|X}(x|x) = 1 \ \ \forall\, x \in \mathcal{W}\right\}. \quad (5)$$

Taking $\overline{\mathcal{W}} = \mathcal{W}$, Theorem 3.1 gives a set of equilibria $(Q^*_{Y|X}, Q^*_{Z|Y})$ where $Q^*_{Y|X} \in Q_A$. The following result shows that the game reduces to a constant-sum game and hence, the payoffs are the same under all equilibria. Define

$$Q_D = \Big\{Q_{Z|Y} \in \mathcal{P}(\mathcal{C}^n|\mathcal{C}^n) : \sum_{x \in \mathcal{W}} P_X(x) Q_{Z|Y}(H_1|x) \leq \gamma\Big\}.$$

Theorem 3.2: Any equilibrium $(Q^*_{Y|X}, Q^*_{Z|Y})$ of the game, where $Q^*_{Y|X} \in Q_A$, also solves the minimax problem

$$\min_{Q_{Z|Y} \in Q_D} \max_{Q_{Y|X} \in Q_A} P_M(Q_{Y|X}, Q_{Z|Y}). \quad (6)$$

Further, for all such equilibria $P^*_M(Q^*_{Y|X}, Q^*_{Z|Y}) = (1 - \mathbb{P}(\mathcal{W}))f(\gamma, \mathcal{W})$ and $P^*_W(Q^*_{Y|X}, Q^*_{Z|Y}) = f(\gamma, \mathcal{W})$.

Proof: For all $Q_{Y|X} \in Q_A$ and $Q_{Z|Y}$, we write $(P_X Q_{Y|X} Q_{Z|Y})(H_0) =$

$$\begin{aligned} &\sum_{x,y \in \mathcal{C}^n, y \neq x} P_X(x) Q_{Y|X}(y|x) Q_{Z|Y}(H_0|y) \\ &+ \sum_{x,y \in \mathcal{C}^n, y = x} P_X(x) Q_{Y|X}(x|x) Q_{Z|Y}(H_0|x) \\ &= P_M(Q_{Y|X}, Q_{Z|Y}) + \sum_{x \in \mathcal{C}^n} P_X(x) Q_{Y|X}(x|x) \\ &- P_F(Q_{Y|X}, Q_{Z|Y}). \end{aligned}$$

For all $x \in \mathcal{C}^n$, the payoff of the adversary in (1) is only determined by $Q_{Y|X}(y|x)$ where $y \in \mathcal{W}$. Hence, without loss of optimality we assume that $\mathrm{supp}(Q_{Y|X}(\cdot|x)) \subseteq \mathcal{W}$ $\forall x \in \mathcal{C}^n$. Using this, we have $(P_X Q_{Y|X} Q_{Z|Y})(H_0) = P_W(Q_{Y|X}, Q_{Z|Y})$. Moreover, $Q_{Y|X} \in Q_A$, gives $\sum_x P_X(x) Q_{Y|X}(x|x) = \mathbb{P}(\mathcal{W})$. Thus, $P_W(Q_{Y|X}, Q_{Z|Y}) = P_M(Q_{Y|X}, Q_{Z|Y}) + \mathbb{P}(\mathcal{W}) - \sum_{x \in \mathcal{W}} P_X(x) Q_{Z|Y}(H_1|x)$. This equality implies that the adversary maximizes $P_M(Q_{Y|X}, Q_{Z|Y})$ by choosing $Q_{Y|X} \in Q_A$. While the detector minimizes $P_M(Q_{Y|X}, Q_{Z|Y})$ subject to $\sum_{x \in \mathcal{W}} P_X(x) Q_{Z|Y}(H_1|x) \leq \gamma$. If $\gamma \geq \mathbb{P}(\mathcal{W})$ then trivially $Q_{Z|Y}(H_1|x) = 1$ $\forall x \in \mathcal{W}$. If $\gamma < \mathbb{P}(\mathcal{W})$ then $\sum_{x \in \mathcal{W}} P_X(x) Q_{Z|Y}(H_1|x) = \gamma$. In any case, $P_W(Q_{Y|X}, Q_{Z|Y}) - P_M(Q_{Y|X}, Q_{Z|Y})$ is a constant in any equilibrium. Thus any equilibrium solves (6). The minimax theorem holds since Q_A and Q_D are closed, convex sets and the objective is linear in the strategies. Finally, $(Q^*_{Y|X}, Q^*_{Z|Y})$ from Theorem 3.1 with $\overline{\mathcal{W}} = \mathcal{W}$ solves (6). ∎

Henceforth, $(Q^*_{Y|X}, Q^*_{Z|Y})$ stand for the pair of strategies from Theorem 3.1 where $\overline{\mathcal{W}} = \mathcal{W}$.

A. Deterrence by controlling the threshold γ

The adversary would benefit from manipulation only if the equilibrium winning probability is more than the prior probability of winning. Since the equilibrium winning probability decreases with increasing threshold γ, there is a minimum false-alarm probability that the detector must bear so that this

probability is not more than the prior probability of winning. Formally, we define this as a *deterrence* threshold.

Definition 3.1: A deterrence threshold, denoted as γ_D, is the smallest value such that for all $\gamma \geq \gamma_D$, we have $P_W^*(Q_{Y|X}^*, Q_{Z|Y}^*) \leq \mathbb{P}(\mathcal{W})$.

Lemma 3.3: The deterrence threshold γ_D is given by $\gamma_D = \mathbb{P}(\mathcal{W})(1 - \mathbb{P}(\mathcal{W}))$.

Proof: Suppose $\gamma \leq \mathbb{P}(\mathcal{W})$. Then, $P_W^*(Q_{Y|X}^*, Q_{Z|Y}^*) \leq \mathbb{P}(\mathcal{W})$ gives $1 - \gamma/\mathbb{P}(\mathcal{W}) \leq \mathbb{P}(\mathcal{W})$ and the value of γ_D. ■

Observe that γ_D is maximum when $\mathbb{P}(\mathcal{W}) = 0.5$. Intuitively, this implies that when the adversarial candidate has a 50% chance of winning *without* any interference, then the detector must endure the highest possible false-alarm in order to minimize the chance of missing the detection of interference.

IV. Asymptotics

Any election has finitely many and large number of voters. Thus in addition to the limiting value of probabilities, it is also of interest to know the rate at which these probabilities approach their limits. Suppose the adversary tries to manipulate votes to ensure that the candidate 0 wins. Let n denote the number of votes. The winning set and the set of winning distributions, denoted as $\mathcal{W}^n$ and $\mathcal{P}_{\mathcal{W}}$ respectively, are

$$\mathcal{W}^n = \{y \in \mathcal{C}^n : P_y(0) \geq P_y(i) \quad \forall\, i \in \mathcal{C}\}, \tag{7}$$

$$\mathcal{P}_{\mathcal{W}} = \{P \in \mathcal{P}(\mathcal{C}) : P(0) \geq P(i) \ \ \forall\, i \in \mathcal{C}\}. \tag{8}$$

We take the threshold on the false-alarm probability to be γ_n. We define $P_F^*(n) = \gamma_n$, $P_M^*(n) = (1 - \mathbb{P}(\mathcal{W}^n))f(\gamma_n, \mathcal{W}^n)$ and $P_W^*(n) = f(\gamma_n, \mathcal{W}^n)$. Here $f(\gamma_n, \mathcal{W}^n)$ is as defined in Section III. Further, we define $P_i^* = \lim_n P_i^*(n)$ for $i \in \{F, M, W\}$, assuming these limits exist. The rate of convergence of the probabilities to their limits is defined as

$$R_i^* = \limsup_{n\to\infty} -\frac{1}{n}\log(P_i^* - P_i^*(n)) \quad i \in \{F, M, W\}.$$

If there exists $N \in \mathbb{N}$ such that $P_i^*(n) = 0\ \forall n \geq N$, then we define $R_i^* := \infty$. We now compute the asymptotics of $\mathbb{P}(\mathcal{W}^n)$. For $p \in \mathcal{P}_n(\mathcal{C})$, define $U_p^n = \{x \in \mathcal{C}^n : P_x = p\}$, where P_x is the empirical distribution of x. For P_{X}, the P_{X}-typical set for any $\epsilon > 0$ is defined as $T_{P_{\mathsf{X}},\epsilon}^n = \{x \in \mathcal{C}^n : |P_x(i) - P_{\mathsf{X}}(i)| < \epsilon\}$. Let $\mathcal{P}_0 \subseteq \mathcal{P}(\mathcal{C})$ be any set. Then, from [11] Ch.2, problem 2.12, we have that

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}\left(\bigcup_{P_n \in \mathcal{P}_0} U_{P_n}^n\right) = \inf_{\bar{Q}\in\mathcal{P}_0} D(\bar{Q}||P_{\mathsf{X}}). \tag{9}$$

Theorem 4.1: Let $\mathcal{W}^n$ be as in (7) and $\mathcal{P}_{\mathcal{W}}$ be as in (8).

1) If $P_{\mathsf{X}}(0) < P_{\mathsf{X}}(i)$ for some $i \in \mathcal{C}, i \neq 0$, then $\lim_n \mathbb{P}(\mathcal{W}^n) = 0$ and $\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\mathcal{W}^n) = \inf_{\bar{Q}\in\mathcal{P}_{\mathcal{W}}} D(\bar{Q}||P_{\mathsf{X}})$.
2) If $P_{\mathsf{X}}(0) > P_{\mathsf{X}}(i)\ \ \forall i \in \mathcal{C}, i \neq 0$, then $\lim_n \mathbb{P}(\mathcal{W}^n) = 1$ and $\lim_{n\to\infty} -\frac{1}{n}\log(1 - \mathbb{P}(\mathcal{W}^n)) = \inf_{\bar{Q}\in\mathcal{P}(\mathcal{C})\setminus\mathcal{P}_{\mathcal{W}}} D(\bar{Q}||P_{\mathsf{X}})$.

Proof: We write $\mathcal{W}^n = \bigcup_{P_n\in\mathcal{P}_{\mathcal{W}}} U_{P_n}^n$.

1) Choose $\epsilon > 0$ such that $P_{\mathsf{X}}(0) + \epsilon < P_{\mathsf{X}}(i) - \epsilon$ holds for some $i \in \mathcal{C}, i \neq 0$ and consider the P_{X}-typical set $T_{P_{\mathsf{X}},\epsilon}^n$. With this ϵ, we have $T_{P_{\mathsf{X}},\epsilon}^n \cap \mathcal{W}^n = \emptyset\ \forall n$. Since $\lim_n \mathbb{P}(T_{P_{\mathsf{X}},\epsilon}^n) = 1$, we have $\mathbb{P}(\mathcal{W}^n) = 0$. Moreover, using (9), we have $\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\mathcal{W}^n) = \inf_{\bar{Q}\in\mathcal{P}_{\mathcal{W}}} D(\bar{Q}||P_{\mathsf{X}})$.

2) Choose $\epsilon > 0$ such that $P_{\mathsf{X}}(0) - \epsilon > P_{\mathsf{X}}(i) + \epsilon\ \forall i \in \mathcal{C}, i \neq 0$ and consider the P_{X}-typical set $T_{P_{\mathsf{X}},\epsilon}^n$. We have $T_{P_{\mathsf{X}},\epsilon}^n \subseteq \mathcal{W}^n\ \forall n$ and hence $\lim_n \mathbb{P}(\mathcal{W}^n) = 1$. Further, $1 - \mathbb{P}(\mathcal{W}^n) = \mathbb{P}(\bigcup_{P_n\in\mathcal{P}\setminus\mathcal{P}_{\mathcal{W}}} U_{P_n}^n)$. Using (9), $\lim_{n\to\infty} -\frac{1}{n}\log(1 - \mathbb{P}(\mathcal{W}^n)) = \inf_{\bar{Q}\in\mathcal{P}\setminus\mathcal{P}_{\mathcal{W}}} D(\bar{Q}||P_{\mathsf{X}})$. ■

Let $Q^* := \arg\min_{\bar{Q}\in\mathcal{P}_{\mathcal{W}}} D(\bar{Q}||P_{\mathsf{X}})$. Let $\gamma_n = \exp(-n\alpha)$ with $\alpha \geq 0$. Let $\gamma_0 := \lim_n \gamma_n/\mathbb{P}(\mathcal{W}^n)$. The next theorem characterizes the asymptotics of the equilibrium probabilities.

Theorem 4.2: Suppose $\lim_n \mathbb{P}(\mathcal{W}^n) = 0$. Then,

1) For all $\alpha > 0$, $P_F^* = 0$ and $R_F^* = \alpha$
2) $P_M^* = P_W^* = \max\{1 - \gamma_0, 0\}$
   - If $\alpha < D(Q^*||P_{\mathsf{X}})$, then $P_M^* = P_W^* = 0$ and $R_M^* = \infty$ and $R_W^* = \infty$
   - If $\alpha > D(Q^*||P_{\mathsf{X}})$, then $P_M^* = P_W^* = 1$ and $R_M^* = R_W^* = \alpha - D(Q^*||P_{\mathsf{X}})$

Suppose $\lim_n \mathbb{P}(\mathcal{W}^n) = 1$. Then,

A) For all $\alpha > 0$, $P_F^* = 0$ and $R_F^* = \alpha$
B) For all $\alpha \geq 0$, $P_M^* = 0$ and $R_M^* = \inf_{Q\in\mathcal{P}(\mathcal{C})\setminus\mathcal{P}_{\mathcal{W}}} D(Q||P_{\mathsf{X}})$
C) For all $\alpha > 0$, $P_W^* = 1$ and $R_W^* = \alpha$

Proof: Let $\lim_n \mathbb{P}(\mathcal{W}^n) = 0$. Since $P_F^*(n) = \gamma_n$, 1) follows. For 2), we have $P_W^* = \lim_n f(\gamma_n, \mathcal{W}^n) = \max\{1 - \gamma_0, 0\}$. Since $P_M^*(n) = (1 - \mathbb{P}(\mathcal{W}^n))P_W^*(n)$ the claim follows for P_M^*. If $\alpha < D(Q * ||P_{\mathsf{X}})$, then γ_n vanishes slowly than $\mathbb{P}(\mathcal{W}^n)$ and hence $\gamma_0 > 1$. Thus there $\exists\ N \in \mathbb{N}$ such that $\gamma_n/\mathbb{P}(\mathcal{W}^n) > 1\ \forall n \geq \mathbb{N}$ and hence $P_M^*(n) = P_W^*(n) = 0\ \forall n \geq N$. If $\alpha > D(Q^*||P_{\mathsf{X}})$, then $\gamma_0 = 0$ and $\max\{1 - \gamma_0, 0\} = 1$. Further, R_M^* and R_W^* are determined by the rate at which $\gamma_n/\mathbb{P}(\mathcal{W}^n)$ vanishes which is $\alpha - D(Q^*||P_{\mathsf{X}})$.

Now let $\lim_n \mathbb{P}(\mathcal{W}^n) = 1$. Part A) follows from 1). Further, $P_M^* = \lim_n(1 - \mathbb{P}(\mathcal{W}^n))P_W^*(n) = 0\ \forall\alpha$. From Theorem 4.1, we get $R_M^* = \inf_{Q\in\mathcal{P}(\mathcal{C})\setminus\mathcal{P}_{\mathcal{W}}} D(Q||P_{\mathsf{X}})$. If $\alpha > 0$, then $\lim_n \gamma_n/\mathbb{P}(\mathcal{W}^n) = 0$ and hence $P_W^* = 1$. The rate R_W^* is same as rate of convergence of γ_n to zero. ■

Notice that in the case where $\lim_n \mathbb{P}(\mathcal{W}^n) = 0$, the adversary loses in the absence of any manipulation. However, if the adversary manipulates the votes and if the detector cannot tolerate a rate of decay of $P_F^*(n)$ slower than $D(Q^*||P_{\mathsf{X}})$, then asymptotically, the adversary can win with arbitrarily high probability with increasing n. Thus, if the detector is too cautious about making a false-alarm error, it pays with a higher missed-detection error which the adversary can use to win *surely*. When the rate is slower than $D(Q^*||P_{\mathsf{X}})$, the adversary can ensure that $P_M^*(n)$ and $P_W^*(n)$ become zero at finite n. For the case where $\lim_n \mathbb{P}(\mathcal{W}^n) = 1$, the adversary wins in any case. However, the detector can control the 'rate' at which the adversary wins by choosing an appropriate α. Observe that the asymptotic results depend only on $f(\gamma_n, \mathcal{W}^n)$ and not on the i.i.d. nature of the sequence of votes x. This assumption is required only to compute the rate of convergence of the probabilities.

Figure 2 shows the rates of the equilibrium probabilities for the case when $\lim_n \mathbb{P}(\mathcal{W}^n) = 0$. For $\alpha < D(Q^*||P_{\mathsf{X}})$,

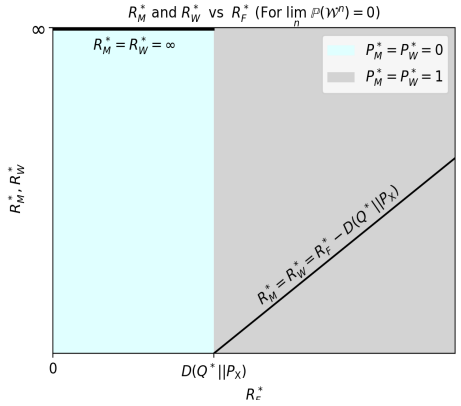

Fig. 2: R_M^* and R_W^* as a function of R_F^* when $\lim_n \mathbb{P}(\mathcal{W}^n) = 0$. The colour denotes the values of $P_M^* = P_W^*$. The black line is $R_M^* = R_W^*$ as a function of R_F^*.

$P_M^*(n)$ and $P_W^*(n)$ are zero after some finite n. For $\alpha > D(Q^*||P_X)$, they tend to one at the rate $\alpha - D(Q^*||P_X)$.

V. Adversary with Multiple Types

We now discuss a setting where the detector has imperfect information about the winning set of the adversary. We model this as an adversary having a type θ in a finite set of K types $\Theta = \{1, \ldots, K\}$. The corresponding winning set of an adversary with type $\theta = i$ is given as $\mathcal{W}_i$. We assume that $\mathcal{W}_i \cap \mathcal{W}_j = \emptyset \ \forall i, j \in \Theta, i \neq j$. The strategy of the adversary is now given as $Q_{Y|X,\theta}$. The detector has a belief over the set of types given by $P_\theta \in \mathcal{P}(\Theta)$. The strategy set of the detector is same as in Section II. The probabilities are

$$P_F(Q_{Y|X,\theta}, Q_{Z|Y}) = \mathbb{P}(Y = X, Z = H_1)$$
$$= \sum_{i \in \Theta, x \in \mathcal{C}^n} P_\theta(i) P_X(x) Q_{Y|X,\theta}(x|x,i) Q_{Z|Y}(H_1|x),$$
$$P_M(Q_{Y|X,\theta}, Q_{Z|Y}) = \mathbb{P}(Y \neq X, Z = H_0)$$
$$= \sum_{i \in \Theta, x \in \mathcal{C}^n, y \neq x} P_\theta(i) P_X(x) Q_{Y|X,\theta}(y|x,i) Q_{Z|Y}(H_0|y),$$
$$P_W(Q_{Y|X,\theta}, Q_{Z|Y}) = \mathbb{P}(Y \in \mathcal{W}, Z = H_0)$$
$$= \sum_{i \in \Theta, x \in \mathcal{C}^n, y \in \mathcal{W}_i} P_\theta(i) P_X(x) Q_{Y|X,\theta}(y|x,i) Q_{Z|Y}(H_0|y).$$

The detector minimizes the probability $P_M(Q_{Y|X,\theta}, Q_{Z|Y})$ subject to a constraint on $P_F(Q_{Y|X,\theta}, Q_{Z|Y})$ while the adversary maximizes $P_W(Q_{Y|X,\theta}, Q_{Z|Y})$. As in the Section III, we restrict the strategies of the adversary to

$$Q_A := \Big\{ Q_{Y|X,\theta} : \forall\, i \in \Theta,\ Q_{Y|X,\theta}(x|x,i) = 1 \ \ \forall\, x \in \mathcal{W}_i \Big\}.$$

With this, $P_F(Q_{Y|X,\theta}, Q_{Z|Y}) \leq \gamma$ restricts $Q_{Z|Y}$ to

$$Q_D = \Big\{ Q_{Z|Y} : \sum_{i \in \Theta, x \in \mathcal{W}_i} P_X(x) P_\theta(i) Q_{Z|Y}(H_1|x) \leq \gamma \Big\}.$$

Let $u_D(Q_{Y|X,\theta}, Q_{Z|Y}) = P_M(Q_{Y|X,\theta}, Q_{Z|Y})$ and $u_A(Q_{Y|X,\theta}, Q_{Z|Y}) = P_W(Q_{Y|X,\theta}, Q_{Z|Y})$. The Nash equilibrium of the game is defined as follows.

Definition 5.1: A pair $(Q^*_{Y|X,\theta}, Q^*_{Z|Y})$ is a NE if

$$u_A(Q^*_{Y|X,\theta}, Q^*_{Z|Y}) \geq u_A(Q_{Y|X,\theta}, Q^*_{Z|Y}) \quad \forall\, Q_{Y|X,\theta} \in Q_A,$$
$$u_D(Q^*_{Y|X,\theta}, Q^*_{Z|Y}) \leq u_D(Q^*_{Y|X,\theta}, Q_{Z|Y}) \quad \forall\, Q_{Z|Y} \in Q_D.$$

To determine a class of Nash equilibria for the game, we define the following linear program.

Definition 5.2: Consider an LP defined as

$$LP: \quad \min_{c \in [0,1]^{|\Theta|}} \sum_{i \in \Theta} P_\theta(i)(1 - \mathbb{P}(\mathcal{W}_i)) c_i$$
$$\text{s.t} \quad \textstyle\sum_{i \in \Theta} P_\theta(i) \mathbb{P}(\mathcal{W}_i)(1 - c_i) \leq \gamma.$$

We now present a class of Nash equilibrium for the game in terms of the optimal solution of the above LP.

Theorem 5.1: Let strategies of adversary be restricted to Q_A. Then, the pair $(Q^*_{Y|X,\theta}, Q^*_{Z|Y})$ is a class of Nash equilibrium of the game where, for all $i \in \Theta$,

$$Q^*_{Y|X,\theta}(y|x,\theta) = \begin{cases} 1 & x \in \mathcal{W}_i, y = x, \theta = i \\ \frac{P_X(y)}{\mathbb{P}(\mathcal{W}_i)} & x \notin \mathcal{W}_i, y \in \mathcal{W}_i, \theta = i \end{cases},$$
$$Q^*_{Z|Y}(H_0|y) = \begin{cases} c_i^* & y \in \mathcal{W}_i \\ 1 & \text{else} \end{cases},$$

where $c^* \in [0,1]^{|\Theta|}$ is an optimal solution of the LP in Definition 5.2. Furthermore, $P_F^*(Q^*_{Y|X,\theta}, Q^*_{Z|Y}) = \gamma$, $P_M^*(Q^*_{Y|X,\theta}, Q^*_{Z|Y}) = \sum_{i \in \Theta} P_\theta(i)(1 - \mathbb{P}(\mathcal{W}_i)) c_i^*$ and $P_W^*(Q^*_{Y|X,\theta}, Q^*_{Z|Y}) = \sum_{i \in \Theta} P_\theta(i) c_i^*$.

Theorem 5.2: Let the strategies of the adversary be restricted to the set Q_A. Then, any Nash equilibrium of the game is also a solution of the minimax problem

$$\min_{Q_{Z|Y} \in Q_D} \max_{Q_{Y|X,\theta} \in Q_A} P_M(Q_{Y|X,\theta}, Q_{Z|Y}). \tag{10}$$

Proof: For all $Q_{Y|X,\theta} \in Q_A$ and $Q_{Z|Y}$, we write $(P_\theta P_X Q_{Y|X,\theta} Q_{Z|Y})(H_0) = P_M(Q_{Y|X,\theta}, Q_{Z|Y}) - P_F(Q_{Y|X,\theta}, Q_{Z|Y}) + \sum_{x \in \mathcal{C}^n} P_X(x)(P_\theta Q_{Y|X,\theta})(x|x)$. As in the proof of Theorem 3.2, without loss of optimality we can assume that the adversary with type i plays $Q_{Y|X,\theta}$ where $\sum_{y \in \mathcal{W}_i} Q_{Y|X,\theta}(y|x,i) = 1 \ \forall x \in \mathcal{C}^n$. Thus $(P_\theta P_X Q_{Y|X,\theta} Q_{Z|Y})(H_0) = P_W(Q_{Y|X,\theta}, Q_{Z|Y})$. Moreover, $Q_{Y|X,\theta} \in Q_A$ gives $\sum_{x \in \mathcal{C}^n} P_X(x)(P_\theta Q_{Y|X,\theta})(x|x) = \sum_{i \in \Theta} P_\theta(i) \mathbb{P}(\mathcal{W}_i)$ and $P_F(Q_{Y|X,\theta}, Q_{Z|Y}) = \sum_{i \in \Theta, x \in \mathcal{W}_i} P_\theta(i) P_X(x) Q_{Z|Y}(H_1|x)$. Thus, $P_W(Q_{Y|X,\theta}, Q_{Z|Y}) = P_M(Q_{Y|X,\theta}, Q_{Z|Y}) + \sum_{i \in \Theta} P_\theta(i) \mathbb{P}(\mathcal{W}_i) - \sum_{i \in \Theta, x \in \mathcal{W}_i} P_\theta(i) P_X(x) Q_{Z|Y}(H_1|x)$. Final arguments are similar to the proof of Theorem 3.2 ■

We now prove Theorem 5.1.

Proof: From Theorem 5.2 it suffices to solve (10) to determine a Nash equilibrium. Assume that $\gamma < \sum_{i \in \Theta} P_\theta(i) \mathbb{P}(\mathcal{W}_i)$. For any strategy $Q_{Z|Y}$, the best response of the adversary is $Q_{Y|X,\theta} \in Q_A$ where $\forall i \in \Theta$ and $x \notin \mathcal{W}_i$, $\sum_{y \in \mathcal{W}_i^*} Q_{Y|X,\theta}(y|x,i) = 1$ with $\mathcal{W}_i^* := \arg\max_{y \in \mathcal{W}_i} Q_{Z|Y}(H_0|y)$. Thus, $u_D(Q_{Y|X,\theta}, Q_{Z|Y}) = \sum_{x \notin \mathcal{W}_i, y \in \mathcal{W}_i^*} P_X(x)(P_\theta Q_{Y|X,\theta})(y|x) Q_{Z|Y}(H_0|y) = \sum_{i \in \Theta} P_\theta(i)(1 - \mathbb{P}(\mathcal{W}_i)) \max_{y \in \mathcal{W}_i} Q_{Z|Y}(H_0|y)$. Thus, the minimax problem is given as

$$\min_{Q_{Z|Y} \in Q_D} \sum_{i \in \Theta} P_\theta(i)(1 - \mathbb{P}(\mathcal{W}_i)) \max_{y \in \mathcal{W}_i} Q_{Z|Y}(H_0|y). \tag{11}$$

We show that the optimal strategy for the detector is a strategy where $Q_{Z|Y}(H_0|y) = c_i \; \forall y \in \mathcal{W}_i, \; \forall i \in \Theta$, with c_i being a constant. Let $\widehat{Q}_{Z|Y} \in Q_D$ and take $\widehat{Q}_{Z|Y}(H_0|y) = c_1 \;\; \forall y \in \mathcal{W}_1$. Now consider $Q'_{Z|Y} \in Q_D$ such that $Q'_{Z|Y}(H_0|y) = \widehat{Q}_{Z|Y}(H_0|y)$ for $y \notin \mathcal{W}_1$ and $Q'_{Z|Y}(H_0|y)$ is *not* a constant on $\mathcal{W}_1$. Since $\gamma < \sum_{i\in\Theta} P_\theta(i)\mathbb{P}(\mathcal{W}_i)$, the constraint in Q_D holds with equality and hence there exists a $y' \in \mathcal{W}_1$ such that $Q'_{Z|Y}(H_0|y') > c_1$. Thus, $\max_{y\in\mathcal{W}_1} \widehat{Q}_{Z|Y}(H_0|y) = c_1 < \max_{y\in\mathcal{W}_1} Q'_{Z|Y}(H_0|y)$. We can argue similarly for all types $i \in \Theta$. Thus, the minimum in (11) is attained by a strategy $Q^*_{Z|Y}$ where

$$Q^*_{Z|Y}(H_0|y) = \begin{cases} c_i & y \in \mathcal{W}_i \\ 1 & \text{else} \end{cases}.$$

with $c \in [0,1]^{|\Theta|}$ being constants. Substituting in (11), we get that the optimal values of $\{c_i\}_{i\in\Theta}$ is given by the LP. ■

The following defines the deterrence thresholds for $i \in \Theta$.

Definition 5.3: A deterrence threshold for the type $i \in \Theta$, denoted as γ^i_D, is the smallest value such that for all $\gamma \geq \gamma^i_D$ and all equilibria $(Q^*_{Y|X,\theta}, Q^*_{Z|Y})$, we have $\mathbb{P}(Z = H_0, Y \in \mathcal{W}_i|\theta = i) \leq \mathbb{P}(\mathcal{W}_i)$.

Theorem 5.3: Suppose $\mathbb{P}(\mathcal{W}_i) \neq \mathbb{P}(\mathcal{W}_j)$ for all $i, j \in \Theta, i \neq j$. Let $c^* \in [0,1]^{|\Theta|}$ be an optimal solution of the LP and let $k \in \Theta$ be such that $c^*_k > 0$. Then,

$$c^*_j = \begin{cases} 0 & \mathbb{P}(\mathcal{W}_j) < \mathbb{P}(\mathcal{W}_k) \\ 1 & \mathbb{P}(\mathcal{W}_j) > \mathbb{P}(\mathcal{W}_k) \end{cases}.$$

Further, $\gamma^k_D = \sum_{j\in\Theta:\mathbb{P}(\mathcal{W}_j)<\mathbb{P}(\mathcal{W}_k)} P_\theta(j)\mathbb{P}(\mathcal{W}_j) + P_\theta(k)(1 - \mathbb{P}(\mathcal{W}_k))\mathbb{P}(\mathcal{W}_k)$.

Proof: We write the Lagrangian of the LP as

$$L(c,\mu,\nu,\lambda) = \sum_{i\in\Theta} P_\theta(i)(1-\mathbb{P}(\mathcal{W}_i))c_i - \sum_{i\in\Theta} \mu_i c_i + \sum_{i\in\Theta} \nu_i(c_i - 1) + \lambda\Big(\sum_{i\in\Theta} P_\theta(i)\mathbb{P}(\mathcal{W}_i)(1-c_i) - \gamma\Big),$$

where $\mu, \nu \in \mathbb{R}^{|\Theta|}, \mu, \nu \geq 0$ are the Lagrange multipliers corresponding to the constraints $c_i \geq 0$ and $c_i \leq 1$, and $\lambda \geq 0$ is the multiplier corresponding to the false-alarm constraint. The optimal Lagrange multipliers, denoted by $(\mu^*, \nu^*, \lambda^*)$, must satisfy the KKT conditions, where $\forall i$,

$$P_\theta(i)(1-\mathbb{P}(\mathcal{W}_i)) - \lambda^* P_\theta(i)\mathbb{P}(\mathcal{W}_i) - \mu^*_i + \nu^*_i = 0, \quad (12)$$

$\lambda^*\big(\sum_{i\in\Theta} P_\theta(i)\mathbb{P}(\mathcal{W}_i)(1-c^*_i) - \gamma\big) = 0$, $\mu^*_i c^*_i = 0$ and $\nu^*_i c^*_i = 0$. Since $c^*_k > 0$, we have $\mu^*_k = \nu^*_k = 0$ which gives $\lambda^* = (1-\mathbb{P}(\mathcal{W}_k))/\mathbb{P}(\mathcal{W}_k)$. Now take $j \in \Theta, j \neq k$. Substituting the value of λ^* in (12), we get

$$P_\theta(j)(1-\mathbb{P}(\mathcal{W}_j)) - \frac{1-\mathbb{P}(\mathcal{W}_k)}{\mathbb{P}(\mathcal{W}_k)} P_\theta(j)\mathbb{P}(\mathcal{W}_j) - \mu^*_j + \nu^*_j$$
$$= (P_\theta(j)/\mathbb{P}(\mathcal{W}_k))(\mathbb{P}(\mathcal{W}_k) - \mathbb{P}(\mathcal{W}_j)) - \mu^*_j + \nu^*_j = 0.$$

If $\mathbb{P}(\mathcal{W}_j) < \mathbb{P}(\mathcal{W}_k)$, then it must be that $\mu^*_j > 0$ which gives $c^*_j = 0$. If $\mathbb{P}(\mathcal{W}_j) > \mathbb{P}(\mathcal{W}_k)$, then $\nu^*_j > 0$ and $c^*_j = 1$.

Now for $\gamma = \gamma^k_D$, we have $c^*_k = \mathbb{P}(\mathcal{W}_k)$. From the constraint, we get that $\sum_{j\in\Theta} P_\theta(j)\mathbb{P}(\mathcal{W}_j)(1-c^*_j) - \gamma^k_D = 0$. This gives that $\gamma^k_D = \sum_{j\in\Theta:\mathbb{P}(\mathcal{W}_j)<\mathbb{P}(\mathcal{W}_k)} P_\theta(i)\mathbb{P}(\mathcal{W}_j) + P_\theta(k)\mathbb{P}(\mathcal{W}_k)(1-\mathbb{P}(\mathcal{W}_k))$. This completes the proof. ■

The above theorem shows that if the adversary with type k is deterred by the adversary, then all the types $j \in \Theta$ are deterred where $\mathbb{P}(\mathcal{W}_j) < \mathbb{P}(\mathcal{W}_k)$. The deterrence threshold also depends on all $\mathbb{P}(\mathcal{W}_j)$ with $\mathbb{P}(\mathcal{W}_j) < \mathbb{P}(\mathcal{W}_k)$, but is independent of $\mathbb{P}(\mathcal{W}_j)$ with $\mathbb{P}(\mathcal{W}_j) > \mathbb{P}(\mathcal{W}_k)$.

VI. Conclusion

We studied a game-theoretic setting where a detector wishes to detect and deter adversarial manipulation in an EVM and performed a static and asymptotic analysis. We found that if the rate of decay of false-alarm probability is too fast, then the detector misses detection and adversary wins with arbitrarily high probability, while if it is low enough, detection is possible. We defined a notion of deterrence threshold on the false-alarm probability that ensures that the posterior probability of winning of the adversary is always lower than the prior winning probability. We then extended the results to the case where the detector has imperfect information about the winning set of the adversary.

References

[1] S. Kumar and E. Walia, "Analysis of electronic voting system in various countries," *International Journal on Computer Science and Engineering*, vol. 3, no. 5, pp. 1825–1830, 2011.

[2] M. Barni and B. Tondi, "The source identification game: An information-theoretic perspective," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 450–463, 2013.

[3] M. Barni, B. Tondi *et al.*, "Theoretical foundations of adversarial binary detection," *Foundations and Trends® in Communications and Information Theory*, vol. 18, no. 1, pp. 1–172, 2020.

[4] S. Yasodharan and P. Loiseau, "Nonzero-sum adversarial hypothesis testing games," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[5] R. Zhang and S. Zou, "A game-theoretic approach to sequential detection in adversarial environments," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1153–1158.

[6] Y. Jin and L. Lai, "On the adversarial robustness of hypothesis testing," *IEEE Transactions on Signal Processing*, vol. 69, pp. 515–530, 2020.

[7] J. Pan, Y. Li, and V. Y. Tan, "Asymptotic nash equilibrium for the m-ary sequential adversarial hypothesis testing game," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 831–845, 2022.

[8] S. Cao, R. Zhang, and S. Zou, "Adversarially robust sequential hypothesis testing," *Sequential Analysis*, vol. 41, no. 1, pp. 81–103, 2022.

[9] A. A. Kulkarni and U. V. Shanbhag, "A shared-constraint approach to multi-leader multi-follower games," *Set-Valued and Variational Analysis*, vol. 22, no. 4, pp. 691–720, 2014.

[10] J. Nash, "Non-cooperative games," *The Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, Sep. 1951.

[11] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.