A capacity-achieving simple decoder for bias-based traitor tracing schemes

Citation for published version (APA):
Oosterwijk, J., Skoric, B., & Doumen, J. M. (2015). A capacity-achieving simple decoder for bias-based traitor tracing schemes. *IEEE Transactions on Information Theory*, *61*(7), 3882-3900.
https://doi.org/10.1109/TIT.2015.2428250

Document license:
TAVERNE

DOI:
10.1109/TIT.2015.2428250

Document status and date:
Published: 01/01/2015

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

A Capacity-Achieving Simple Decoder for Bias-Based Traitor Tracing Schemes

Jan-Jaap Oosterwijk, Boris Škorić, and Jeroen Doumen

***Abstract*— We investigate alternative suspicion functions for bias-based traitor tracing schemes, and present a practical construction of a simple decoder that attains capacity in the limit of large coalition size c. We derive optimal suspicion functions in both the restricted-digit model and the combined-digit model. These functions depend on information that is usually not available to the tracer—the attack strategy or the tallies of the symbols received by the colluders. We discuss how such results can be used in realistic contexts. We study several combinations of coalition attack strategy versus suspicion function optimized against some attack (another attack or the same). In many of these combinations, the usual codelength scaling $\ell \propto c^2$ changes to a lower power of c, e.g., $c^{3/2}$. We find that the interleaving strategy is an especially powerful attack. The suspicion function tailored against interleaving is the key ingredient of the capacity-achieving construction.**

***Index Terms*— Collusion resistance, traitor tracing.**

I. Introduction

A. Collusion Attacks on Watermarking

FORENSIC watermarking is a means for tracing the origin and distribution of digital content. Before distribution, the content is modified by embedding an imperceptible watermark, which plays the role of a personalized serial number. Once an unauthorized copy of the content is found, the identities of those users who participated in its creation can be determined. A tracing algorithm outputs a list of suspicious users.

The most powerful attacks against watermarking are *collusion attacks* [1], in which multiple attackers (the 'coalition') combine their differently watermarked versions of the same content; the observed differences point to the locations of the hidden marks.

In the past two decades several types of collusion-resistant codes have been developed. The most popular type in the recent literature is the class of *bias-based* codes. These were introduced by G. Tardos in 2003. The original paper [2] was followed by a flurry of activity, e.g. improved analyses [3]–[8], code modifications [9]–[11], decoder modifications [12]–[14] and various generalizations [15]–[18]. The advantage of bias-based versus deterministic codes is that they can achieve the asymptotically optimal relationship $\ell \propto c^2$ between the sufficient code length ℓ and the coalition size c.

Two kinds of tracing algorithm can be distinguished: (i) *simple decoders* [14], [16], [19], which assign a level of suspicion to single users and (ii) *joint decoders* [12], [13], [20], which look at sets of users. Joint decoders employ a simple decoder as a bootstrapping step.

Tardos' scheme worked with a binary code and a simple decoder. Its 'suspicion function' for computing a level of suspicion for single users was improved [16] and the scheme was generalized to q-ary alphabets. However, it turns out [21] that the suspicion function yields sub-optimal fingerprinting rates, i.e. rather far below the fingerprinting capacity [22], [23] and far below the best achieved dynamic code rate [24].

Alternative suspicion functions for the binary case were introduced [13], where an Expectation Maximization (EM) algorithm was used. A candidate coalition is selected, which (if the guess is sufficiently good) makes it possible to estimate the employed attack strategy; a suspicion function is then used which is optimized against that strategy. This leads to a new ranking of users, giving a new candidate coalition, and the whole process is repeated until it converges.

B. Contributions

This paper is an extended version of earlier work on optimal suspicion functions [25]. The current work analyzes the worst attack against the interleaving defense. It turns out that there is no stronger attack than the interleaving attack. This implies that the interleaving defense actually achieves capacity asymptotically. The current version also includes all proofs.

- We generalize the work of Charpentier et al. [13] to q-ary alphabets. Using functional derivation methods we obtain suspicion functions that for large c maximize the expected score for the coalition. We present results for the Combined-Digit Model and the Restricted-Digit Model.
- We consider a set of often-considered attack strategies. We substitute these attacks into the generic formulas and obtain closed-form expressions for the asymptotically optimal suspicion functions associated with these attacks.
- We tabulate the performance for each combination of attack and suspicion function. For some cases we prove theorems analytically and for all binary cases we have

Manuscript received July 1, 2013; revised December 12, 2014; accepted March 13, 2015. Date of publication April 30, 2015; date of current version June 12, 2015. This work was supported in part by the Dutch Technology Foundation STW, a part of the Netherlands Organization for Scientific Research, and in part by the Ministry of Economic Affairs. This paper was presented at the First ACM workshop on Information Hiding and Multimedia Security.

J.-J. Oosterwijk is with the Eindhoven University of Technology, Eindhoven 5612 AZ, The Netherlands, and also with Irdeto B.V., Hoofddorp 2132 LS, The Netherlands (e-mail: jan-jaap.oosterwijk@irdeto.com).

B. Škorić is with the Eindhoven University of Technology, Eindhoven 5612 AZ, The Netherlands (e-mail: b.skoric@tue.nl).

J. Doumen is with Irdeto B.V., Hoofddorp 2132 LS, The Netherlands (e-mail: jdoumen@irdeto.com).

Communicated by Y. Liang, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2428250

numerical results. Naturally, in case of a match the sufficient code length ℓ is small; for all considered strategies but the interleaving attack we even find $\ell \propto c^{3/2}$. For the interleaving attack and its matching suspicion function we find an asymptotic fingerprinting rate $(q-1)/(2c^2 \ln q)$, which is exactly the q-ary asymptotic fingerprinting capacity.

In non-matching cases the results differ widely. In some cases, as expected, the mismatched defense fails completely, while in others the code length remains $\ell \propto c^2$ (often with a smaller coefficient than with the Tardos suspicion function), and in many cases we find $\ell \propto c^{3/2}$ even for a mismatch.

- The suspicion function tailored against the interleaving attack is very special. When this suspicion function is adopted as the basis of a simple decoder, the minimax game for the asymptotic code rate (attack strategy versus bias distribution function) has a saddle point when the interleaving attack is used and the distribution function is the Dirichlet distribution with concentration parameter 1/2. In the saddle point the asymptotic rate equals the asymptotic capacity. The saddle point is the same point that was found by Huang and Moulin [23] for the mutual information minimax game. Thus, we have identified a simple decoder that asymptotically achieves capacity.

In Sections III-A and XI we comment on possible ways to exploit our results for the construction of improved decoders by using several suspicion functions in parallel, and/or deploying a tally-dependent suspicion to strengthen the EM algorithm, and/or to validate candidate coalitions in general.

II. Preliminaries

A. General Notation

We denote random variables by capital letters and their realizations in lower case. We write vectors in boldface. We define $[\ell] = \{1, \dots, \ell\}$. The q-ary alphabet is $\mathcal{A}$, which is sometimes set to $\mathcal{A} = \{0, \dots, q-1\}$.

We use multi-index notation, e.g. $\boldsymbol{p}^\kappa = \prod_{\alpha\in\mathcal{A}} p_\alpha^\kappa$, $\boldsymbol{p}^{\boldsymbol{m}} = \prod_{\alpha\in\mathcal{A}} p_\alpha^{m_\alpha}$, and $\binom{c}{\boldsymbol{m}} = c!/\prod_{\alpha\in\mathcal{A}} m_\alpha!$.

We define the norm of a vector as $|\boldsymbol{p}| = \sum_{\alpha\in\mathcal{A}} |p_\alpha|$. For probability mass/density functions we use abbreviated notation of the form $f_{y|\boldsymbol{p}} = f_{Y|\boldsymbol{P}}(y|\boldsymbol{p})$ when it does not cause ambiguity.

In conditional expectation values we sometimes use the abbreviation $\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[\cdots] = \mathbb{E}_{\boldsymbol{M}}[\cdots|\boldsymbol{P} = \boldsymbol{p}]$. An $\mathbb{E}$ without subscripts is an expectation over *all* probabilistic degrees of freedom. We use $\delta_{x,y}$ to denote the Kronecker delta function, which is 1 when $x = y$ and 0 when $x \neq y$.

The notation $\frac{\partial A}{\partial p_x}||_{|\boldsymbol{p}|=1}$ is defined as follows. First the derivative $\partial A/\partial p_x$ is taken *without* taking the constraint $\sum_\alpha p_\alpha = 1$ into account. After differentiation the constraint is enforced.

We will use the shorthand notation $a_k := (p_0 + \cdots + p_{k-1})$ and $a_{\mathcal{B}} = \sum_{\beta\in\mathcal{B}} p_\beta$.

B. Bias-Based Tracing; Simple Decoder

The content contains ℓ abstract 'locations' into which a q-ary symbol can be embedded. For each location $i \in [\ell]$ independently, the tracer draws a bias vector $\boldsymbol{P}_i = (P_{i,\alpha})_{\alpha\in\mathcal{A}}$ from a distribution $f_{\boldsymbol{P}}$. The biases satisfy $P_{i,\alpha} \geq 0$ and $|\boldsymbol{P}_i| = 1$. A symmetric Dirichlet distribution was taken [16], with concentration parameter $\kappa > 0$,

$$f_{\boldsymbol{p}} = \boldsymbol{p}^{\kappa-1}\Gamma(q\kappa)/[\Gamma(\kappa)]^q. \tag{1}$$

For $q = 2$ it is customary to set $\kappa = \frac{1}{2}$, turning (1) into the arcsine distribution for the component p_1. However, in that case the support has to be reduced to $p_1 \in [\delta, 1-\delta]$, with cutoff parameter $\delta > 0$, in order to avoid statistical problems due to extremely unlikely events. The probability density function then becomes

$$f_{p_1} = \frac{1}{2\arcsin(1-2\delta)} \frac{1}{\sqrt{p_1(1-p_1)}}. \tag{2}$$

As the cutoff parameter is typically chosen so small that it vanishes, we will neglect it in our analysis. The number of users is n. For each $i \in [\ell]$ and each $j \in [n]$, the tracer draws a random symbol $X_{i,j} \in \mathcal{A}$ according to the categorical distribution with parameter $\boldsymbol{P}_i$, i.e. $\mathbb{P}[X_{i,j} = \alpha|\boldsymbol{P}_i = \boldsymbol{p}_i] = p_{i,\alpha}$ independent of j. The symbol $X_{i,j}$ is embedded into the content of user j in location i.

The coalition of attackers is denoted as $\mathcal{C} \subset [n]$, with $|\mathcal{C}| = c$. In some attack models, e.g. the Combined-Digit Model (Section II-C), they are allowed to do signal processing attacks such as introducing noise and fusing symbols. In the Restricted-Digit Model (RDM) they are only allowed to select one colluder's symbol (denoted as y_i) in location i. In the *simple decoder* approach, the tracer determines a score S_j for each user j by adding independently computed sub-scores $S_{i,j}$ for each location i; these are based on $\boldsymbol{p}_i$, $X_{i,j}$ and the colluders' output in location i. If the score exceeds a threshold, user j is accused.

Tardos [2] introduced a (simple decoder) score system for the RDM at $q = 2$ that was later [16] symmetrized and generalized to $q > 2$. The sub-scores for each location are computed using a 'suspicion function' g as $S_{i,j} = g(x_{i,j}, y_i, \boldsymbol{p}_i)$ with

$$g(x, y, \boldsymbol{p}) = \begin{cases} \sqrt{(1-p_y)/p_y} & \text{if } x = y \\ -\sqrt{p_y/(1-p_y)} & \text{if } x \neq y. \end{cases} \tag{3}$$

It has the special property that the $S_{i,j}$ of innocent users has expectation 0 and variance 1.

Given the symmetries present in the code generation and accusation algorithm, it is usually assumed that the attackers apply a strategy that acts at every location independently. Furthermore, we assume that the colluders take equal risks. In such an attack model, the colluders' decision in location i depends only on the tallies $M_{i,\alpha} = |\{j \in \mathcal{C}|X_{i,j} = \alpha\}|$ (with $\alpha \in \mathcal{A}$). The tallies satisfy $|\boldsymbol{M}_i| = c$, and they are multinomial-distributed, with density $f_{\boldsymbol{m}|\boldsymbol{p}} = \binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}}$. The attack strategy may be probabilistic.

C. Combined-Digit Model (CDM)

The CDM [17] allows colluders to mix symbols and to introduce noise (see Figure 1). In each location, the symbols that are mixed are assumed to have equal power. The set of symbols that the colluders choose to mix is denoted

$$\boldsymbol{P} \xrightarrow[f_{\boldsymbol{M}|\boldsymbol{P}}]{\text{code generation}} \boldsymbol{M} \xrightarrow[f_{\boldsymbol{\Psi}|\boldsymbol{M}}]{\text{colluder mix}} \boldsymbol{\Psi} \xrightarrow[f_{\boldsymbol{\Phi}|\boldsymbol{\Psi}}]{\text{tracer detection}} \boldsymbol{\Phi}$$

Fig. 1. A schematic depiction of the CDM.

as $\boldsymbol{\Psi} \subseteq \mathcal{A}$ with $m_\alpha > 0$ for each $\alpha \in \boldsymbol{\Psi}$. The attack strategy is parametrized by a set of probabilities $f_{\boldsymbol{\psi}|\boldsymbol{m}}$. The tracer has a detector that outputs a set $\boldsymbol{\Phi} \subseteq \mathcal{A}$ of observed symbols. The joint effects of the noise and the mixing lead to probability distributions $f_{\boldsymbol{\Phi}|\boldsymbol{\Psi}}$, where it is possible that the noise introduces symbols in $\boldsymbol{\Phi}$ that are absent in $\boldsymbol{\Psi}$. Simple-decoder score systems were introduced in [17] and [18].

The CDM reduces to the RDM when the noise strength is sent to zero and the detector unerringly observes $\boldsymbol{\Phi} = \boldsymbol{\Psi}$, forcing the colluders to output a single symbol, $\boldsymbol{\Psi} = \{Y\}$. For the RDM, a strategy is parametrized by a set of probabilities $f_{y|\boldsymbol{m}}$.

D. Performance; Moments of the Scores

The performance of bias-based tracing schemes can for a large part be characterized by looking merely at the first and second moment of the innocent and guilty scores. (This holds especially at large c, where the large code length induces an almost-Gaussian shape of the score probability distributions.)

For an innocent user j, we define the mean and variance as

$$\tilde{\mu}_{\text{inn}} := \mathbb{E}[S_{i,j}] \tag{4}$$

$$\tilde{\sigma}^2_{\text{inn}} := \text{Var}[S_{i,j}] = \mathbb{E}[(S_{i,j} - \tilde{\mu}_{\text{inn}})^2] = \mathbb{E}[S^2_{i,j}] - \tilde{\mu}^2_{\text{inn}}, \tag{5}$$

where the index $i \in [\ell]$ is arbitrary. The expectation $\mathbb{E}$ is taken over the random variables $\boldsymbol{P}_i$, $X_{i,j}$, and Y_i (in the CDM $\boldsymbol{\Psi}_i$ and $\boldsymbol{\Phi}_i$ instead of Y_i). We call a suspicion function centered if it yields $\tilde{\mu}_{\text{inn}} = 0$ and normalized if $\tilde{\sigma}^2_{\text{inn}} = 1$. For the coalition we define $S_{i,\mathcal{C}} := \sum_{j\in\mathcal{C}} S_{i,j}$. The moments are

$$\tilde{\mu}_{\mathcal{C}} := \mathbb{E}[S_{i,\mathcal{C}}] \tag{6}$$

$$\tilde{\sigma}^2_{\mathcal{C}} := \text{Var}[S_{i,\mathcal{C}}] = \mathbb{E}[(S_{i,\mathcal{C}} - \tilde{\mu}_{\mathcal{C}})^2] = \mathbb{E}[S^2_{i,\mathcal{C}}] - \tilde{\mu}^2_{\mathcal{C}} \tag{7}$$

again with arbitrary index i.

The limit $c \to \infty$ (and correspondingly $n \to \infty$) is of particular interest, since it lends itself to analysis. The term ‘asymptotically’ will refer to this limit.

Due to the Central Limit Theorem, the probability distribution of the innocent users’ score is asymptotically Gaussian. For Gaussian distributions it was shown [8], [16] that the sufficient code length ℓ_{suff} is given by

$$\ell_{\text{suff}} = \frac{2\tilde{\sigma}^2_{\text{inn}}}{\tilde{\mu}^2_{\mathcal{C}}} c^2 \ln \frac{1}{\varepsilon_1} \tag{8}$$

where ε_1 is the maximum tolerated false accusation probability for any fixed innocent user. (Note that $1/\varepsilon_1$ is proportional to n).

The *fingerprinting rate* R of a code is defined as $R \overset{\text{def}}{=} \frac{\log_q n}{\ell}$. The asymptotic rate that follows from (8) is

$$R_{\text{asymp}} = \frac{\tilde{\mu}^2_{\mathcal{C}}}{2\tilde{\sigma}^2_{\text{inn}} \ln q} \cdot \frac{1}{c^2} \tag{9}$$

The *fingerprinting capacity* is the maximum achievable fingerprinting rate. For the RDM it was shown [22] that the capacity in the case of a joint decoder is given by $C^{\text{joint}}_{\text{RDM}} = \frac{q-1}{c^2 \cdot 2 \ln q}$. The mutual information game, in which the colluders choose an attack strategy $f_{y|\mathbf{m}}$ and the tracer chooses a bias distribution $f_{\mathbf{p}}$, has a saddlepoint [23] at $f_{y|\mathbf{m}} = m_y/c$ (the interleaving attack), $f_{\mathbf{p}} \propto \mathbf{p}^{-1/2}$ (the Dirichlet distribution with $\kappa q = 1/2$). It was also shown [23] that the simple-decoder capacity asymptotically becomes equal to the joint-decoder capacity.

III. Optimal Suspicion Functions

We consider suspicion functions h other than the function g given in (3). We derive suspicion functions that maximize the performance indicator $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$, in the CDM as well as the RDM. Without loss of generality, we will consider only suspicion functions that are centered ($\tilde{\mu}_{\text{inn}} = 0$) and normalized ($\tilde{\sigma}_{\text{inn}} = 1$). We use the standard approach of Lagrange functionals; we use constraint multipliers $\lambda_1, \lambda_2 \in \mathbb{R}$ to enforce the constraints $\tilde{\mu}_{\text{inn}} = 0$ and $\tilde{\sigma}_{\text{inn}} = 1$. We define the functional

$$L(h, \lambda_1, \lambda_2) = \tilde{\mu}_{\mathcal{C}} - \lambda_1 \tilde{\mu}_{\text{inn}} - \tfrac{1}{2}\lambda_2(\tilde{\sigma}^2_{\text{inn}} - 1), \tag{10}$$

where $\tilde{\mu}_{\text{inn}}$, $\tilde{\sigma}_{\text{inn}}$ and $\tilde{\mu}_{\mathcal{C}}$ depend on the function h as specified in (4-6). The optimal h is found by solving the set of equations $\delta L/\delta h = 0$, $\partial L/\partial \lambda_1 = 0$ and $\partial L/\partial \lambda_2 = 0$. The solution depends on the arguments of h: in the CDM the sub-score of user j in location i is typically a function of $X_{i,j}$, $\boldsymbol{\Phi}_i$ and $\boldsymbol{P}_i$; in the RDM a function of $X_{i,j}$, Y_i and $\boldsymbol{P}_i$.

A. Optimal Suspicion Functions in the Combined-Digit Model

We present a number of lemmas leading up to the main theorem of this section, which shows the solution obtained by the Lagrangian approach. The conditional probabilities that appear in the lemmas are related as follows: $f_{\boldsymbol{\psi}|\boldsymbol{p}} = \sum_{\boldsymbol{m}} f_{\boldsymbol{\psi}|\boldsymbol{m}} f_{\boldsymbol{m}|\boldsymbol{p}}$ and $f_{\boldsymbol{\phi}|\boldsymbol{p}} = \sum_{\boldsymbol{\psi}} f_{\boldsymbol{\phi}|\boldsymbol{\psi}} f_{\boldsymbol{\psi}|\boldsymbol{p}}$. The numbers $f_{\boldsymbol{\phi}|\boldsymbol{\psi}}$ are fixed parameters of the CDM independent of the strategy.

Lemma 1: An optimal suspicion function of the form $h(x, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{p})$ does not depend on $\boldsymbol{\phi}$. An optimal suspicion function of the form $h(x, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{m}, \boldsymbol{p})$ depends neither on $\boldsymbol{\phi}$ nor $\boldsymbol{\psi}$.

Proof Sketch: The set $\boldsymbol{\psi}$ contains more information about the attackers than the set $\boldsymbol{\phi}$. Likewise, the tallies $\boldsymbol{m}$ contain more information than $\boldsymbol{\psi}$. ■

We will give the full proof after Theorem 1.

To determine the optimal suspicion functions of the increasingly general forms $h(x, \boldsymbol{\phi}, \boldsymbol{p})$, $h(x, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{p})$, and $h(x, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{m}, \boldsymbol{p})$, it suffices to study the forms $h_{\boldsymbol{\Phi}}(x, \boldsymbol{\phi}, \boldsymbol{p})$, $h_{\boldsymbol{\Psi}}(x, \boldsymbol{\psi}, \boldsymbol{p})$, and $h_{\boldsymbol{M}}(x, \boldsymbol{m}, \boldsymbol{p})$, respectively.

Lemma 2: Let h be of the form $h_{\boldsymbol{\Phi}}(x, \boldsymbol{\phi}, \boldsymbol{p})$ and define

$$T_{\boldsymbol{\Phi}}(x, \boldsymbol{\phi}, \boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{c p_x f_{\boldsymbol{\phi}|\boldsymbol{p}}} = \frac{1}{c} \left. \frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x} \right|_{|\boldsymbol{p}|=1} + 1. \tag{11}$$

Then $\tilde{\mu}_{\mathcal{C}} = c \cdot \mathbb{E}[T_{\boldsymbol{\Phi}} h]$ and $\mathbb{E}[T_{\boldsymbol{\Phi}}] = 1$.

Proof: We write (6) as

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{M}} \sum\nolimits_{x\in\mathcal{A}} M_x h(x,\boldsymbol{\Phi},\boldsymbol{P}) \tag{12}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{P}} \frac{f_{\boldsymbol{\Phi}|\boldsymbol{M}}}{f_{\boldsymbol{\Phi}|\boldsymbol{P}}}\mathbb{E}_{X|\boldsymbol{P}} \frac{M_X}{P_X} h(X,\boldsymbol{\Phi},\boldsymbol{P}) \tag{13}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}} \left[\frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}[M_X f_{\boldsymbol{\Phi}|\boldsymbol{M}}]}{P_X f_{\boldsymbol{\Phi}|\boldsymbol{P}}}\, h(X,\boldsymbol{\Phi},\boldsymbol{P})\right] \tag{14}$$

$$= c\;\mathbb{E}[T\cdot h]. \tag{15}$$

Furthermore, $\mathbb{E}_{X|\boldsymbol{p}}[m_X/p_X] = c$ and $f_{\boldsymbol{\phi}|\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{\boldsymbol{\phi}|\boldsymbol{M}}]$, so

$$\mathbb{E}_{X|\boldsymbol{p}}[T] = \mathbb{E}_{X|\boldsymbol{p}}\left[\frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_X f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{c p_X f_{\boldsymbol{\phi}|\boldsymbol{p}}}\right] = 1. \tag{16}$$

To be able to take the partial derivative $\frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x}$, the components $p_0,\ldots,p_{q-1}$ are assumed to be functionally independent. In particular, we do not assume $|\boldsymbol{p}|=1$ during differentiation. Since $f_{\boldsymbol{m}|\boldsymbol{p}} = \frac{1}{|\boldsymbol{p}|^c}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}}$, we find

$$f_{\boldsymbol{\phi}|\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{\boldsymbol{\phi}|\boldsymbol{M}}] = \frac{1}{|\boldsymbol{p}|^c}\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}} f_{\boldsymbol{\phi}|\boldsymbol{m}}. \tag{17}$$

$$\frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x} = \frac{\frac{1}{p_x}\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}} m_x f_{\boldsymbol{\phi}|\boldsymbol{m}}}{\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}} f_{\boldsymbol{\phi}|\boldsymbol{m}}} - \frac{c}{|\boldsymbol{p}|} \tag{18}$$

$$= \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{p_x f_{\boldsymbol{\phi}|\boldsymbol{p}}} - \frac{c}{|\boldsymbol{p}|}. \tag{19}$$

So $\frac{1}{c}\left.\frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1 = T_{\boldsymbol{\Phi}}(x,\boldsymbol{\phi},\boldsymbol{p})$. ■

When the colluders output is known the optimal suspicion function is derived as follows:

Lemma 3: Let h be of the form $h_{\boldsymbol{\Psi}}(x,\boldsymbol{\psi},\boldsymbol{p})$ and define

$$T_{\boldsymbol{\Psi}}(x,\boldsymbol{\psi},\boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\psi}|\boldsymbol{M}}]}{c p_x f_{\boldsymbol{\psi}|\boldsymbol{p}}} = \frac{1}{c}\left.\frac{\partial \ln f_{\boldsymbol{\psi}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1. \tag{20}$$

Then $\tilde{\mu}_{\mathcal{C}} = c\cdot\mathbb{E}[T_{\boldsymbol{\Psi}}h]$ and $\mathbb{E}[T_{\boldsymbol{\Psi}}] = 1$.

Proof: We write (6) as

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Psi}|\boldsymbol{M}} \sum\nolimits_{x\in\mathcal{A}} M_x h(x,\boldsymbol{\Psi},\boldsymbol{P}). \tag{21}$$

Note the similarity between (21) and (12). The proof proceeds analogously with $\boldsymbol{\Psi}$ instead of $\boldsymbol{\Phi}$. ■

When even the tallies of the coalition symbols are known the optimal suspicion function is derived as follows:

Lemma 4: Let h be of the form $h_{\boldsymbol{M}}(x,\boldsymbol{m},\boldsymbol{p})$ and define

$$T_{\boldsymbol{M}}(x,\boldsymbol{m},\boldsymbol{p}) := \frac{m_x}{c p_x} = \frac{1}{c}\left.\frac{\partial \ln f_{\boldsymbol{m}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1. \tag{22}$$

Then $\tilde{\mu}_{\mathcal{C}} = c\cdot\mathbb{E}[T_{\boldsymbol{M}}h]$, $\mathbb{E}[T_{\boldsymbol{M}}] = 1$, and $\mathrm{Var}[T_{\boldsymbol{M}}] = \frac{q-1}{c}$.

Proof: We write (6) as

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}} \sum\nolimits_{x\in\mathcal{A}} M_x h(x,\boldsymbol{M},\boldsymbol{P}) \tag{23}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{M_X}{P_X}\, h(X,\boldsymbol{M},\boldsymbol{P})\right] = c\;\mathbb{E}[T\cdot h]. \tag{24}$$

Furthermore, $\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x] = cp_x$, so

$$\mathbb{E}[T] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{M_X}{cP_X}\right] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}[1] = 1. \tag{25}$$

Also

$$\mathrm{Var}[T] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\left(\frac{M_X}{cP_X}-1\right)^2 \tag{26}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\mathrm{Var}_{\boldsymbol{M}|\boldsymbol{P}}\left[\frac{M_X}{cP_X}\right]$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{cP_X(1-P_X)}{c^2P_X^2}\right] \tag{27}$$

$$= \tfrac{1}{c}\mathbb{E}_{\boldsymbol{P}}\sum\nolimits_{x\in\mathcal{A}}(1-P_x) = (q-1)/c. \tag{28}$$

Also, $\frac{\partial \ln f_{\boldsymbol{m}|\boldsymbol{p}}}{\partial p_x} = \frac{m_x}{p_x} - \frac{c}{|\boldsymbol{p}|}$ and thus $\frac{1}{c}\left.\frac{\partial \ln f_{\boldsymbol{m}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1 = T_{\boldsymbol{M}}(x,\boldsymbol{m},\boldsymbol{p})$. ■

Theorem 1: In each of the cases above, the centered and normalized suspicion function that maximizes $\tilde{\mu}_{\mathcal{C}}$ is

$$h = (T-\mathbb{E}[T])\,/\sqrt{\mathrm{Var}[T]} \tag{29}$$

and the expected coalition score is $\tilde{\mu}_{\mathcal{C}} = c\cdot\sqrt{\mathrm{Var}[T]}$.

Proof: Define the Lagrangian

$$L(h,\lambda_1,\lambda_2) := c\;\mathbb{E}[Th] - \lambda_1\mathbb{E}[h] - \tfrac{1}{2}\lambda_2(\mathbb{E}[h^2]-1) \tag{30}$$

with the two Lagrange multipliers λ_1 and λ_2 enforcing that the function is centered and normalized respectively. Let h be such that $\frac{\delta L}{\delta h} = 0$. Then $D(cT-\lambda_1-\lambda_2 h) = 0$ (where D is the product of the probability densities of the random variables), i.e. $h = \frac{cT-\lambda_1}{\lambda_2}$. The first constraint, $\tilde{\mu}_{\text{inn}} = 0$, implies that $\lambda_1 = c\;\mathbb{E}[T]$ and the second constraint, $\tilde{\sigma}^2_{\text{inn}} = 1$, implies that $\lambda_2^2 = \mathbb{E}(cT-\lambda_1)^2 = c^2\mathrm{Var}[T]$.

From the previous lemmas, we conclude that

$$\tilde{\mu}_{\mathcal{C}} = c\;\mathbb{E}[Th] = c\;\mathbb{E}[T-\mathbb{E}[T]]/\sqrt{\mathrm{Var[T]}}$$

$$= c\;\mathrm{Var[T]}/\sqrt{\mathrm{Var[T]}} = c\;\sqrt{\mathrm{Var[T]}}. \tag{31}$$

■

Now that we have seen the proof technique for Theorem 1, we can state the full proof of Lemma 1:

Proof of Lemma 1: To determine the optimal suspicion function of the form $h(x,\boldsymbol{\psi},\boldsymbol{p})$ in the proof of Theorem 1 we defined the Lagrangian

$$L(h,\lambda_1,\lambda_2) := c\;\mathbb{E}[Th] - \lambda_1\mathbb{E}[h] - \tfrac{1}{2}\lambda_2(\mathbb{E}[h^2]-1). \tag{32}$$

where $\mathbb{E}[\ldots] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Psi}|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}[\ldots]$. The Euler-Lagrange equation was $D(cT-\lambda_1-\lambda_2 h) = 0$ with $D = f_{\boldsymbol{p}} f_{\boldsymbol{\psi}|\boldsymbol{p}} f_{x|\boldsymbol{p}}$.

Instead, to determine the optimal suspicion function of the form $h(x,\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{p})$, we would define the same Lagrangian, but now with $\mathbb{E}[\ldots] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Psi}|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{\Psi}}\mathbb{E}_{X|\boldsymbol{P}}[\ldots]$. We obtain the same Euler-Lagrange equation but now with $D = f_{\boldsymbol{p}} f_{\boldsymbol{\psi}|\boldsymbol{p}} f_{\boldsymbol{\phi}|\boldsymbol{\psi}} f_{x|\boldsymbol{p}}$.

In both cases, we draw the same conclusion: that $cT-\lambda_1-\lambda_2 h = 0$. We therefore find that the optimal suspicion function of the form $h(x,\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{p})$ is the one we found in Lemma 3 of the form $h(x,\boldsymbol{\psi},\boldsymbol{p})$.

Likewise, the optimal suspicion function of the form $h(x,\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{m},\boldsymbol{p})$ is the one we found in Lemma 4 of the form $h(x,\boldsymbol{m},\boldsymbol{p})$. ■

Our suspicion functions have a close relation with Neyman-Pearson scores, as shown in the following proposition.

Proposition 1: In all three cases $T_{\boldsymbol{\Phi}}(x, \boldsymbol{\phi}, \boldsymbol{p})$, $T_{\boldsymbol{\Psi}}(x, \boldsymbol{\psi}, \boldsymbol{p})$ and $T_{\boldsymbol{M}}(x, \boldsymbol{m}, \boldsymbol{p})$ for the function T it holds that

$$T(x, \square, \boldsymbol{p}) = \frac{\mathbb{P}[x, \square, \boldsymbol{p} | j \in \mathcal{C}]}{\mathbb{P}[x, \square, \boldsymbol{p} | j \notin \mathcal{C}]}, \tag{33}$$

and thus T is a Neyman-Pearson score.

Proof: The Neyman-Pearson score for testing a hypothesis H given evidence e is given by the likelihood ratio $\mathbb{P}[e|H = \text{True}]/\mathbb{P}[e|H = \text{False}]$. Our hypothesis is $H = (j \in \mathcal{C})$ for a user $j \in [n]$, and we consider the evidence $e = (x, \boldsymbol{\phi}, \boldsymbol{p})$ available in one location. (The proof for all the other cases is analogous.) Then the Neyman-Pearson score is

$$\frac{\mathbb{P}[x, \boldsymbol{\phi}, \boldsymbol{p} | j \in \mathcal{C}]}{\mathbb{P}[x, \boldsymbol{\phi}, \boldsymbol{p} | j \notin \mathcal{C}]} = \frac{\mathbb{P}[x, \boldsymbol{\phi}, \boldsymbol{p}, j \in \mathcal{C}]}{\mathbb{P}[x, \boldsymbol{\phi}, \boldsymbol{p}, j \notin \mathcal{C}]} \frac{\mathbb{P}[j \notin \mathcal{C}]}{\mathbb{P}[j \in \mathcal{C}]} \tag{34}$$

$$= \frac{f_{\boldsymbol{p}} f_{x|\boldsymbol{p}} f_{\boldsymbol{\phi}|x,\boldsymbol{p},j\in\mathcal{C}}}{f_{\boldsymbol{p}} f_{x|\boldsymbol{p}} f_{\boldsymbol{\phi}|\boldsymbol{p}}} \tag{35}$$

$$= \frac{f_{\boldsymbol{\phi}|x,\boldsymbol{p},j\in\mathcal{C}}}{f_{\boldsymbol{\phi}|\boldsymbol{p}}} \tag{36}$$

$$= \frac{1}{f_{\boldsymbol{\phi}|\boldsymbol{p}}} \sum_{\boldsymbol{m}: m_x \geq 1} \binom{c-1}{\boldsymbol{m} - \boldsymbol{e}_x} \boldsymbol{p}^{\boldsymbol{m} - \boldsymbol{e}_x} f_{\boldsymbol{\phi}|\boldsymbol{m}} \tag{37}$$

$$= \frac{1}{f_{\boldsymbol{\phi}|\boldsymbol{p}}} \sum_{\boldsymbol{m}} \frac{m_x}{c p_x} \binom{c}{\boldsymbol{m}} \boldsymbol{p}^{\boldsymbol{m}} f_{\boldsymbol{\phi}|\boldsymbol{m}} \tag{38}$$

$$= \frac{1}{f_{\boldsymbol{\phi}|\boldsymbol{p}}} \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}} \left[\frac{M_x}{c p_x} f_{\boldsymbol{\phi}|\boldsymbol{M}} \right]. \tag{39}$$

Here $\boldsymbol{e}_x$ is a length q vector containing a 1 in position x and zero elsewhere. ■

Several things are worth noting about these results.

(i) In the proof of Theorem 1 it is not necessary to specify the bias distribution. Though $\tilde{\mu}_{\mathcal{C}}$ is a functional of both h and $f_{\boldsymbol{P}}$, the optimization of h does not depend on $f_{\boldsymbol{P}}$.

(ii) In all three cases the result for h depends on information that the tracer usually does not have. (The strategy $f_{\boldsymbol{\psi}|\boldsymbol{m}}$ in Lemmas 2 and 3; the tallies $\boldsymbol{m}$ in Lemma 4). When a function $h_{\boldsymbol{\Phi}}$, for some guessed strategy, is used to compute scores, there is no guarantee that the attackers are actually adhering to that guessed strategy. Such 'mismatched' situations will be discussed (for the RDM) in the remainder of this paper.

(iii) We can think of two ways in which the $\boldsymbol{m}$-dependent result of Lemma 4, $h(x, \boldsymbol{m}, \boldsymbol{p}) = (\frac{m_x}{c p_x} - 1)\sqrt{\frac{c}{q-1}}$, can be used in practice. First, it could be employed in the EM algorithm [13]. The EM procedure estimates a strategy based on the symbols received by the candidate coalition, and then uses this estimate to adapt the suspicion function. Our h function could be used to directly assign scores to all users, *skipping the strategy estimation step*. This would speed up each iteration of the EM algorithm and avoid the statistical inaccuracies in the estimation. (Of course, inaccuracies due to a wrongly guessed coalition remain, and may even increase.)
Secondly, this h function can be used as a consistency check in the following way. Suppose that, by some means, a candidate coalition $\hat{\mathcal{C}}$ has been tentatively identified. Then one computes a score $(\frac{m_x}{c p_x} - 1)\sqrt{\frac{c}{q-1}}$ for all users, where the tally m_x is based on $\hat{\mathcal{C}}$ and the user's symbol x. If $\hat{\mathcal{C}}$ equals the actual coalition, one should see a huge score difference between innocent users and the colluders. Exploration of these ideas is left for future work.

(iv) The expression $\partial \ln f / \partial p_x$ in all three cases has the form of a Fisher score, being the derivative of the logarithm of a conditional probability with respect to the conditioning variable. We suspect that this form is no coincidence. However, the intuitive meaning of the associated 'game' (guessing $\boldsymbol{p}$ from y) is not immediately obvious. Asymptotically $\boldsymbol{m}$ tends to $c\boldsymbol{p}$. We hypothesize that the game 'guess $\boldsymbol{p}$ from y' is asymptotically equivalent to 'guess $\boldsymbol{m}$ from y'. The latter is a known formulation of the tracing problem.

(v) Our result in Proposition 1 is different from the Neyman-Pearson score in [14] and [20], where the whole sequence $(Y_i)_{i \in [\ell]}$ was considered.

B. Optimal Suspicion Functions in the Restricted-Digit Model

The Restricted-Digit Model is a special case of the Combined-Digit Model.

Corollary 1: Let h be of the form $h_Y(x, y, \boldsymbol{p})$ and define

$$T_Y(x, y, \boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{y|\boldsymbol{M}}]}{c p_x f_{y|\boldsymbol{p}}} = \frac{1}{c} \left. \frac{\partial \ln f_{y|\boldsymbol{p}}}{\partial p_x} \right|_{|\boldsymbol{p}|=1} + 1. \tag{40}$$

Then $\tilde{\mu}_{\mathcal{C}} = c \cdot \mathbb{E}[T_Y h]$ and $\mathbb{E}[T_Y] = 1$.

Proof: The optimal h function in the RDM case follows straightforwardly from Lemma 2 and Theorem 1 by taking the limit of zero noise and perfect detection of all mixed symbols, leading to $\boldsymbol{\Phi} = \boldsymbol{\Psi} = \{Y\}$, with $Y \in \mathcal{A}$. ■

In the RDM, Lemma 4 and Theorem 1 hold without change. Note that the Marking Assumption is not invoked to obtain Corollary 1. Hence Corollary 1 is valid in a more general setting, as long as the colluders produce a single symbol which is unerringly detected by the tracer.

Note also that (40) with $q = 2$ matches the expression given by Charpentier et al. [13] (which only considered the binary case).

C. Strongly Centered and Normalized Suspicion Functions

In (10) we required our optimal score functions to be centered and normalized. The normalization was done without loss of generality, since scores can be rescaled arbitrarily. The symmetric Tardos suspicion function was chosen to satisfy stronger properties: it is both centered and normalized, no matter what the pirate symbol y or the bias vector $\boldsymbol{p}$ are (and no matter what the attack strategy or the bias distribution is for that matter). These properties are captured in the following definition.

We call a suspicion function $h(x, y, \boldsymbol{p})$ *strongly centered* if $\mathbb{E}_{X|\boldsymbol{p}}[h(X, y, \boldsymbol{p})] = 0$ and *strongly normalized* if $\mathbb{E}_{X|\boldsymbol{p}}[h^2(X, y, \boldsymbol{p})] = 1$.

We show that even when the score function does not match the pirate strategy, the optimal score functions derived in the previous section remain centered but not necessarily normalized.

Lemma 5: Each optimal suspicion function (see Theorem 1) is strongly centered. So is the symmetric Tardos function.

Proof: This follows directly from (16). ■

If we wanted to find optimal suspicion functions that are both *strongly* centered and *strongly* normalized, like the symmetric Tardos suspicion function, in (10) we should require $\mathbb{E}_{X|\boldsymbol{p}}[h(X, y, \boldsymbol{p})] = 0$ and $\mathrm{Var}_{X|\boldsymbol{p}}[h(X, y, \boldsymbol{p})] = 1$. Since our optimal suspicion functions already turned out to be strongly centered, in Theorem 1 only the normalizing constant changes:

Corollary 2: The strongly centered and strongly normalized suspicion function that maximizes $\tilde{\mu}_{\mathcal{C}}$ is

$$h = \left(T - \mathbb{E}_{X|\boldsymbol{p}}[T]\right) / \sqrt{\mathrm{Var}_{X|\boldsymbol{p}}[T]} \tag{41}$$

and the expected coalition score is $\tilde{\mu}_{\mathcal{C}} = c \cdot \sqrt{\mathrm{Var}_{X|\boldsymbol{p}}[T]}$.

Proof: Define the Lagrangian

$$\begin{aligned} L(h, \lambda_1, \lambda_2) := c\, \mathbb{E}_{X|\boldsymbol{p}} &\left[T(X, y, \boldsymbol{p}) h(X, y, \boldsymbol{p})\right] + \\ &- \lambda_1 \mathbb{E}_{X|\boldsymbol{p}}[h(X, y, \boldsymbol{p})] + \\ &- \tfrac{1}{2}\lambda_2 (\mathbb{E}_{X|\boldsymbol{p}}[h^2(X, y, \boldsymbol{p})] - 1) \end{aligned} \tag{42}$$

with the two Lagrange multipliers λ_1 and λ_2 enforcing that the function is strongly centered and strongly normalized respectively. Let h be such that $\frac{\delta L(h,\lambda_1,\lambda_2)}{\delta h(x,y,\boldsymbol{p})} = 0$. Then

$$f_{x|\boldsymbol{p}}(cT(x, y, \boldsymbol{p}) - \lambda_1 - \lambda_2 h(x, y, \boldsymbol{p}) = 0, \tag{43}$$

i.e. $h(x, y, \boldsymbol{p}) = \frac{cT(x,y,\boldsymbol{p}) - \lambda_1}{\lambda_2}$.

The first constraint that $h(x, y, \boldsymbol{p})$ is strongly centered implies that $\lambda_1 = c \ \mathbb{E}_{X|\boldsymbol{p}}[T(X, y, \boldsymbol{p})]$ and the second constraint that $h(x, y, \boldsymbol{p})$ is strongly normalized implies that

$$\lambda_2^2 = \mathbb{E}_{X|\boldsymbol{p}}(cT(X, y, \boldsymbol{p}) - \lambda_1)^2 = c^2 \mathrm{Var}_{X|\boldsymbol{p}}[T(X, y, \boldsymbol{p})]. \tag{44}$$

■

D. Building a Traitor Tracing Scheme

Now that we have described our new optimal suspicion function, there is one caveat left to address. As noted before, when a suspicion function h, for some guessed strategy, is used to compute scores, there is no guarantee that the attackers are actually adhering to that guessed strategy. In particular, this means that we no longer have the property $\tilde{\sigma}_{\mathrm{inn}}^2 = 1$ which the Tardos suspicion function enjoys.

As a result, we can not simply plug our suspicion function into the Tardos traitor tracing scheme, since such a scheme typically accuses a user when he exceeds a fixed threshold. Traditionally, the use of a fixed threshold is possible since the scaling of the scores is taken care of by the property $\tilde{\sigma}_{\mathrm{inn}}^2 = 1$. Thus, ideally, we would like to have a normalized suspicion function. This can be achieved by scaling all scores (i.e. scaling the function h) by a factor $\tilde{\sigma}_{\mathrm{inn}}$. Since at this point both the bias vector $\boldsymbol{p}$ and the collusion symbols y have been determined, we can calculate the sample variance. Thus if we scale all scores by a factor $\tilde{\sigma}$ by replacing them with $S_j/\tilde{\sigma}$ for every user j, the traitor tracing scheme will perform well against any collusion strategy.

IV. Defending Against Common Collusion Strategies

From this point onward, we consider only the RDM. For a number of often-studied strategies we compute the optimal suspicion function. We investigate the situation where the actual attack is indeed the one for which the h-function was designed (a "match"), as well as mismatches. We will call the "optimal suspicion function against strategy A" the A-defense. The following sections will focus on defenses against five often-considered strategies. In short, these strategies can be described as follows:

1) *Interleaving Attack (Section V):* The interleaving attack randomly selects an attacker and outputs his symbol.
2) *All-High (All-1) Attack (Section VI):* The all-high attack is special as it breaks the symbol symmetry. It assumes that the alphabet can be ordered in some meaningful way, and outputs the largest received symbol. In the binary case $q = 2$ this attack is known as the all-1 attack, as it will output a 1 if the coalition has received one.
3) *Random-Symbol (Coin-Flip) Attack (Section VII):* The random-symbol attack randomly selects a received symbol, irrespective of the tally vector $\boldsymbol{m}$, and outputs it. In the binary case $q = 2$ this attack is known as the coin-flip attack.
4) *Majority Voting Attack (Section VIII):* The majority voting attack outputs the symbol that was received most often by the coalition. In case multiple symbols are received equally often, a random symbol is chosen among them.
5) *Minority Voting Attack (Section IX):* The minority voting attack outputs the symbol that was received least often (but at least once) by the coalition. When multiple symbols are received equally often, a random symbol is chosen among them.

A detailed description of each attack will be given at the start of its section. We also dedicate a section (Section X) to analyzing the performance of the traditional symmetrized Tardos suspicion function against these attacks.

V. Interleaving Defense

A. Optimal Defense

The interleaving attack $f_{y|\boldsymbol{m}} = m_y/c$ randomly selects an attacker and outputs his symbol.

Proposition 2: Against the interleaving attack, the quantity T is given by $T(x, y, \boldsymbol{p}) = 1 + \frac{1}{c}(\delta_{x,y}/p_y - 1)$, and the optimal suspicion function is

$$h(x, y, \boldsymbol{p}) = \frac{1}{\sqrt{q-1}} \left(\frac{\delta_{x,y}}{p_y} - 1\right). \tag{45}$$

Proof: We find

$$f_{y|\boldsymbol{p}} = \frac{1}{c|\boldsymbol{p}|^c} \sum_{\boldsymbol{m}} \binom{c}{\boldsymbol{m}} \boldsymbol{p}^{\boldsymbol{m}} m_y = \frac{p_y}{c|\boldsymbol{p}|^c} \frac{\partial |\boldsymbol{p}|^c}{\partial p_y} = \frac{p_y}{|\boldsymbol{p}|}. \tag{46}$$

Thus

$$\left. \frac{\partial \ln f_{y|\boldsymbol{p}}}{\partial p_x} \right|_{|\boldsymbol{p}|=1} = \frac{\delta_{x,y}}{p_y} - 1, \tag{47}$$

so $T(x, y, \boldsymbol{p}) = 1 + \frac{1}{c}(\delta_{x,y}/p_y - 1)$. Also,

$$\begin{aligned}
\mathrm{Var}[T] &= \mathbb{E}(T-1)^2 && (48)\\
&= \tfrac{1}{c^2}\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\left(\delta_{X,Y}/p_Y - 1\right)^2\right] && (49)\\
&= \frac{1}{c^2}\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_Y\left(\frac{1-P_Y}{P_Y}\right)^2 + \sum_{x\neq Y} P_x\right] && (50)\\
&= \frac{1}{c^2}\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1-P_Y}{P_Y}\right] && (51)\\
&= \frac{1}{c^2}\,\mathbb{E}_{\boldsymbol{P}}[q-1] = \frac{q-1}{c^2}. && (52)
\end{aligned}$$

■

The performance of the interleaving defense against the interleaving attack is given in the following lemma.

Proposition 3: When the interleaving attack is used against the interleaving defense, then $\tilde{\mu}_{\mathcal{C}} = \sqrt{q-1}$, achieving capacity for any $f_{\boldsymbol{P}}$.

Proof: From Theorem 1 we know that $\tilde{\mu}_{\mathcal{C}} = c\cdot\sqrt{\mathrm{Var}[T]}$. Combining this with (52) yields $\tilde{\mu}_{\mathcal{C}} = \sqrt{q-1}$.

We compute the asymptotic code rate according to (9), using $\tilde{\sigma}_{\mathrm{inn}} = 1$. This gives $R_{\mathrm{asymp}} = \frac{q-1}{c^2\cdot 2\ln q}$, which is exactly the asymptotic fingerprinting capacity (see Section II-D). In the derivation of this result, the bias distribution $f_{y|\mathbf{p}}$ was not used at any point. ■

Finding a suspicion function that (even asymptotically) achieves capactity is a remarkable result. In the rest of this section we will first analyze its performance against other known attack strategies. In the next Section (V-B), we will focus on its performace against generic attack strategies, and show that the interleaving defense achieves capacity against any attack strategy.

When $x = y$, the h is positive and increasing in p_y (rare events raise more suspicion). When $x \neq y$, it is negative and constant, in contrast to (3). The h is independent of c.

Lemma 6: If the tracer uses the interleaving defense, then, no matter what attack is used,

$$\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}}\left(-1 + \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[T(Y, Y, \boldsymbol{P})\right]\right) \qquad (53)$$

and

$$\tilde{\sigma}^2_{\mathrm{inn}} = \frac{1}{q-1}\left(-1 + \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1}{P_Y}\right]\right). \qquad (54)$$

where T belongs to the attack.

Proof: Using the interleaving defense from (45), we find

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}[T\cdot h] = \frac{c}{\sqrt{q-1}}\left(\mathbb{E}\left[T(X, Y, \boldsymbol{P})\frac{\delta_{X,Y}}{P_Y}\right] - 1\right). \qquad (55)$$

Also

$$h^2(x, y, \boldsymbol{p}) = \frac{1}{q-1}\left(\frac{\delta_{x,y}}{p_y}\left(\frac{1}{p_y} - 2\right) + 1\right), \qquad (56)$$

$$\begin{aligned}
\text{so}\quad \tilde{\sigma}^2_{\mathrm{inn}} &= \mathbb{E}[h^2] && (57)\\
&= \frac{1}{q-1}\left(1 + \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1}{P_Y} - 2\right]\right). && (58)
\end{aligned}$$

■

We can explicitly calculate the performance against the all-high attack (which is formalized in Proposition 10). Recall that $a_k := (p_0 + \cdots + p_{k-1})$.

Proposition 4: If the tracer uses the interleaving defense, but the coalition uses the all-high attack, then

$$\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}}\sum_{y=0}^{q-2}\mathbb{E}_{\boldsymbol{P}}\left[A^{c-1}_{y+1}\right],\ \textit{and} \qquad (59)$$

$$\tilde{\sigma}^2_{\mathrm{inn}} = \frac{1}{q-1}\left(-1 + \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol{P}}\left[\frac{A^c_{y+1} - A^c_y}{P_y}\right]\right). \qquad (60)$$

Proof: Using Lemma 6 with (116), we find

$$\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}[T(Y, Y, \boldsymbol{P})] = \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol{P}}\left[A^{c-1}_{y+1}\right] = \sum_{y=0}^{q-2}\mathbb{E}_{\boldsymbol{P}}\left[A^{c-1}_{y+1}\right] + 1. \qquad (61)$$

and, with (121),

$$\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1}{P_Y}\right] = \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol{P}}\left[\frac{A^c_{y+1} - A^c_y}{P_y}\right]. \qquad (62)$$

■

If the Dirichlet distribution is used $\tilde{\mu}_{\mathcal{C}}$ will scale as $c^{1-\kappa}$ for large coalitions:

Proposition 5: Let $f_{\boldsymbol{P}}$ be the symmetric Dirichlet distribution with cutoff $\delta = 0$. If the tracer uses the interleaving defense, but the colluders use the all-high attack, then

$$\tilde{\mu}_{\mathcal{C}} = \frac{\Gamma(q\kappa)}{\Gamma([q-1]\kappa)}\frac{c^{1-\kappa}}{\sqrt{q-1}}[1 + \mathcal{O}(1/c)]. \qquad (63)$$

Proof: Lemma 6 gives $\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}}(\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}[T(Y, Y, \boldsymbol{P})] - 1)$. Next, using Proposition 10 we get $\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}}[-1 + \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol{P}}A^{c-1}_y]$ which can be simplified to $\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}}\sum_{y=0}^{q-2}\mathbb{E}_{\boldsymbol{P}}A^{c-1}_y$. The easiest way to evaluate the expectation is by using the marginal distribution of A_y, which is given by $M(a_y) = a_y^{y\kappa-1}(1 - a_y)^{[q-y]\kappa-1}/B(y\kappa, [q-y]\kappa)$. (See derivation at the end of this proof.) This yields

$$\begin{aligned}
\tilde{\mu}_{\mathcal{C}} &= \frac{c}{\sqrt{q-1}}\sum_{y=0}^{q-2}\frac{B([q-1-y]\kappa, [y+1]\kappa + c - 1)}{B([q-1-y]\kappa, [y+1]\kappa)} && (64)\\
&= \frac{c}{\sqrt{q-1}}\sum_{b=1}^{q-1}\frac{\Gamma(q\kappa)\Gamma(c-1+b\kappa)}{\Gamma(b\kappa)\Gamma(c-1+q\kappa)}. && (65)
\end{aligned}$$

Next we use the property $\Gamma(x+\alpha)/\Gamma(x+\beta) = x^{\alpha-\beta}[1 + \mathcal{O}(1/x)]$ which holds if $x \gg 1$, $a, b \ll x$, and a, b independent of x. (See [7, Lemma 7].) This gives

$$\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}}\sum_{b=1}^{q-1}\frac{\Gamma(q\kappa)}{\Gamma(b\kappa)}c^{(b-q)\kappa}[1 + \mathcal{O}\left(\frac{1}{c}\right)]. \qquad (66)$$

The dominant term is $b = q-1$, yielding (63). The smaller b values in the sum are terms of relative order $1/c$ or smaller.

Finally we derive the marginal distribution $M(a_y)$. We compute $M(a_y) = \mathbb{E}_{\boldsymbol{P}}\delta(a_y - \sum_{\alpha=0}^{y-1} P_\alpha)$,

$$M(a_y) = \int_0^1 \mathrm{d}^q p\ \delta(1-|\boldsymbol{p}|)\frac{\boldsymbol{p}^{\kappa-1}}{B(\kappa\mathbf{1}_q)}\delta(a_y - \sum_{\alpha=0}^{y-1} p_\alpha), \tag{67}$$

where $\mathbf{1}_q$ is a vector consisting of q ones and B is the generalized Beta function. We do the following change of integration variables: for $\alpha < y$ we write $p_\alpha = a_y t_\alpha$ and for $\alpha \geq y$ we write $p_\alpha = (1-a_y)s_\alpha$. This gives $\delta(a_y - \sum_{\alpha=0}^{y-1} p_\alpha) = a_y^{-1}\delta(1-|\boldsymbol{t}|)$ and $\delta(1-|\boldsymbol{p}|) = (1-a_y)^{-1}\delta(1-|\boldsymbol{s}|)$. Furthermore, $\mathrm{d}^q p\ \boldsymbol{p}^{\kappa-1} = \mathrm{d}^y t \mathrm{d}^{q-y} s\ a_y^{y\kappa}(1-a_y)^{[q-y]\kappa}\boldsymbol{t}^{\kappa-1}\boldsymbol{s}^{\kappa-1}$. Substitution into (67) gives

$$M(a_y) = \frac{a_y^{y\kappa-1}(1-a_y)^{[q-y]\kappa-1}}{B(\kappa\mathbf{1}_q)}[\int_0^1 \mathrm{d}^y t\delta(1-|\boldsymbol{t}|)\boldsymbol{t}^{\kappa-1}]$$
$$\cdot[\int_0^1 \mathrm{d}^{q-y}s\delta(1-|\boldsymbol{s}|)\boldsymbol{s}^{\kappa-1}] \tag{68}$$
$$= \frac{a_y^{y\kappa-1}(1-a_y)^{[q-y]\kappa-1}}{B(\kappa\mathbf{1}_q)}B(\kappa\mathbf{1}_y)B(\kappa\mathbf{1}_{q-y}). \tag{69}$$

Simplification of the Beta functions gives the density $M(a_y)$ as listed earlier in this proof. ■

We now investigate the binary case $q = 2$. We can then rephrase Proposition 4 as

Corollary 3: Let $q = 2$. *If the tracer uses the interleaving defense, but the coalition uses the all-1 attack, then* $\tilde{\mu}_{\mathcal{C}} = c\,\mathbb{E}_{\boldsymbol{P}}\big[P_0^{c-1}\big]$ *and* $\tilde{\sigma}^2_{\text{inn}} = -1 + \mathbb{E}_{\boldsymbol{P}}\Big[P_0^{c-1} + \frac{1-P_0^c}{P_1}\Big]$.

Proof: $A_1 = P_0$ and $A_2 = P_0 + P_1 = 1$. ■

In the binary case, we obtain explicit results for the coin-flip attack (as formalized in Proposition 15) against the interleaving defense:

Proposition 6: Let $q = 2$. *If the tracer uses the interleaving defense, but the coalition uses the coin-flip attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \tfrac{1}{2}c\ \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + P_1^{c-1}\right] \textit{and} \tag{70}$$

$$\tilde{\sigma}^2_{\text{inn}} = -1 + \mathbb{E}_{\boldsymbol{P}}\left[\frac{1+P_0^c - P_1^c}{2P_0} + \frac{1+P_1^c - P_0^c}{2P_1}\right]. \tag{71}$$

Proof: Using Lemma 6 with (165), we find

$$\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}[T(Y,Y,\boldsymbol{P})] = \tfrac{1}{2}\textstyle\sum_{y\in\mathcal{A}}\mathbb{E}_{\boldsymbol{P}}[1+P_y^{c-1}] \tag{72}$$
$$= 1 + \tfrac{1}{2}\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1} + P_1^{c-1}]. \tag{73}$$

and, with (169),

$$\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1}{P_Y}\right] = \tfrac{1}{2}\sum_{y\in\mathcal{A}}\mathbb{E}_{\boldsymbol{P}}\frac{1+P_y^c - P_{1-y}^c}{P_y}. \tag{74}$$

■

Note the similarity between the coin-flip attack and the all-1 attack. For the Dirichlet distribution, this can be analytically shown:

Proposition 7: Let $q = 2$ *and* $f_{\boldsymbol{P}}$ *be the symmetric Dirichlet distribution with parameter* $\kappa = \frac{1}{2}$ *without cutoff. If the tracer uses the interleaving defense and the coalition uses either the all-1 or the coin-flip attack, then*

$$\tilde{\mu}_{\mathcal{C}} = c\cdot B(\kappa,\kappa+c-1)/B(\kappa,\kappa)\ \ \textit{and} \tag{75}$$

$$\tilde{\sigma}^2_{\text{inn}} = -1 + \frac{c}{1-\kappa}\frac{\Gamma(2\kappa)}{\Gamma(\kappa)}\frac{\Gamma(c+\kappa-1)}{\Gamma(c+2\kappa-1)} + \frac{1-2\kappa}{1-\kappa}. \tag{76}$$

For large c *these behave as* $\tilde{\mu}_{\mathcal{C}} \propto c^{1-\kappa}$ *and* $\tilde{\sigma}^2_{\text{inn}} \propto c^{1-\kappa}$.

Proof: In the case of the coin-flip attack we have

$$\tilde{\mu}_{\mathcal{C}} = \tfrac{1}{2}c\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1} + P_1^{c-1}] = c\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1}] \tag{77}$$
$$= cB(\kappa,\kappa+c-1)/B(\kappa,\kappa) \tag{78}$$

since $f_{\boldsymbol{P}}$ is symbol-symmetric.

Also, $f_{y|\boldsymbol{p}} = \frac{1}{2} + \frac{1}{2}p_y^c - \frac{1}{2}(1-p_y)^c$. When the interleaving suspicion function is used, (54) tells us that $\tilde{\sigma}^2_{\text{inn}} = -1 + \mathbb{E}[1/P_Y]$.

We have

$$\mathbb{E}\left[\frac{1}{P_Y}\right] = \sum_{y\in\{0,1\}}\mathbb{E}_{\boldsymbol{P}}\left[\frac{f_{y|\boldsymbol{P}}}{P_y}\right] \tag{79}$$
$$= \frac{1}{2}\sum_{y\in\{0,1\}}\mathbb{E}_{\boldsymbol{P}}\left[\frac{1}{P_y} + P_y^{c-1} - \frac{(1-P_y)^c}{P_y}\right] \tag{80}$$
$$= \mathbb{E}_{\boldsymbol{P}}\left[\frac{1}{P_y} + P_y^{c-1} - \frac{(1-P_y)^c}{P_y}\right] \tag{81}$$
$$= \frac{B(\kappa-1,\kappa) + B(c+\kappa-1,\kappa) - B(\kappa-1,c+\kappa)}{B(\kappa,\kappa)}. \tag{82}$$

In the third line we used the fact that $f_{\boldsymbol{P}}$ is symbol-symmetric. Re-expressing the Beta functions in terms of Gamma functions, followed by some simplification, yields

$$\mathbb{E}\left[\frac{1}{P_Y}\right] = \frac{c}{1-\kappa}\frac{\Gamma(2\kappa)}{\Gamma(\kappa)}\frac{\Gamma(c+\kappa-1)}{\Gamma(c+2\kappa-1)} + \frac{1-2\kappa}{1-\kappa}. \tag{83}$$

Due to the symbol symmetry of $f_{\boldsymbol{P}}$, the derivations for the all-1 attack are the same. ■

B. Does the Interleaving Defense Achieve Capacity?

The main question we ask ourselves in this section is whether the interleaving defense can be used as a generic suspicion function against all attacks. Proposition 3 shows that it asymptotically yields a rate equal to capacity in the saddle point [23]. However, we have to exclude the possibility that there is a better attack than interleaving against the interleaving defense. We do this in two steps: first we will show the existence of a saddlepoint in the attack vs. distribution space, and then we will argue that for a fixed distribution this saddle point attains the minimum value of the performance indicator-in other words, that there is no better attack.

In the first step we do a saddlepoint analysis of the performance indicator $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$, in the following setting. We fix the employed suspicion function h to be the 'Interleaving defense' as specified in Proposition 2. The tracer has to tune the bias distribution $f_{\boldsymbol{p}}$ and at the same time the coalition has to find the best possible attack $f_{y|\boldsymbol{m}}$ against the combination $f_{\boldsymbol{p}}, h$. This simultaneous counter-acting optimization leads to a *saddle point* solution for $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ which is a minimum as a function of the attack strategy and a maximum as a function of $f_{\boldsymbol{p}}$. A similar analysis was done by Huang and Moulin [23] in the context of the q-ary fingerprinting capacity, abstracting away the exact suspicion function to be employed. They found the saddlepoint at (attack = interleaving, $f_{\boldsymbol{p}}$ = Dirichlet with $\kappa = \frac{1}{2}$), consistent with the asymptotic ($c \to \infty$) fingerprinting capacity $(q-1)/(2c^2 \ln q)$ known earlier [22].

We use the Lagrangian approach, with functional L given by

$$L = \frac{\tilde{\mu}_{\mathcal{C}}^2}{\tilde{\sigma}_{\text{inn}}^2} + \sum_{\boldsymbol{m}} \lambda_{\boldsymbol{m}} \Big(\sum_{y \in \mathcal{Q}} f_{y|\boldsymbol{m}} - 1 \Big) + \Lambda \Big(\int \mathrm{d}^q p \; \delta(|\boldsymbol{p}| - 1) f_{\boldsymbol{p}} - 1 \Big). \tag{84}$$

Here the $\lambda_{\boldsymbol{m}}$ and Λ are constraint multipliers: $\lambda_{\boldsymbol{m}}$ multiplies the constraint that, for every $\boldsymbol{m}$, $f_{y|\boldsymbol{m}}$ is a probability mass function for y, and Λ multiplies the constraint that $f_{\boldsymbol{p}}$ is a probability density function.

Proposition 8: Let the tracer use the interleaving defense. When the interleaving attack strategy is used, and the bias distribution is the Dirichlet distribution with $\kappa = \frac{1}{2}$, a saddlepoint occurs. Also, the asymptotic fingerprinting rate in the saddlepoint is equal to the asymptotic fingerprinting capacity $\frac{q-1}{2c^2 \ln q}$.

Proof: The proof consists of two parts: (i) showing that we have a stationary point in the $(f_{y|\boldsymbol{m}}, f_{\boldsymbol{p}})$-space which corresponds to capacity; (ii) showing that the stationary point is a maximum as a function of $f_{\boldsymbol{p}}$ and a minimum as a function of $f_{y|\boldsymbol{m}}$.

Part 1: For the interleaving defense we have from Lemma 6 and Corollary 1 that

$$\tilde{\mu}_{\mathcal{C}} \sqrt{q-1} = \sum_{y \in \mathcal{Q}} \mathbb{E}_{\boldsymbol{P}} \left[\left. \frac{\partial f_{y|\boldsymbol{P}}}{\partial P_y} \right|_{|\boldsymbol{P}|=1} \right] \tag{85}$$

$$= -c + \sum_{y \in \mathcal{Q}} \sum_{\boldsymbol{m}} \binom{c}{\boldsymbol{m}} f_{y|\boldsymbol{m}} m_y \mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right], \tag{86}$$

and

$$\tilde{\sigma}_{\text{inn}}^2 (q-1) = -1 + \sum_{y \in \mathcal{Q}} \mathbb{E}_{\boldsymbol{P}} \left[\frac{f_{y|\boldsymbol{P}}}{P_y} \right] \tag{87}$$

$$= -1 + \sum_{y \in \mathcal{Q}} \sum_{\boldsymbol{m}} \binom{c}{\boldsymbol{m}} f_{y|\boldsymbol{m}} \mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right]. \tag{88}$$

The functional derivatives of $\tilde{\mu}_{\mathcal{C}}$ and $\tilde{\sigma}_{\text{inn}}^2$ are

$$\sqrt{q-1} \frac{\delta \tilde{\mu}_{\mathcal{C}}}{\delta f_{y|\boldsymbol{m}}} = \binom{c}{\boldsymbol{m}} m_y \mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right]; \tag{89}$$

$$\sqrt{q-1} \frac{\delta \tilde{\mu}_{\mathcal{C}}}{\delta f_{\boldsymbol{p}}} = \delta(|\boldsymbol{p}| - 1) \sum_{y \in \mathcal{Q}} \left. \frac{\partial f_{y|\boldsymbol{p}}}{\partial p_y} \right|_{|\boldsymbol{p}|=1}; \tag{90}$$

$$(q-1) \frac{\delta \tilde{\sigma}_{\text{inn}}^2}{\delta f_{y|\boldsymbol{m}}} = \binom{c}{\boldsymbol{m}} \mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right]; \tag{91}$$

$$(q-1) \frac{\delta \tilde{\sigma}_{\text{inn}}^2}{\delta f_{\boldsymbol{p}}} = \delta(|\boldsymbol{p}| - 1) \sum_{y \in \mathcal{Q}} \frac{f_{y|\boldsymbol{p}}}{p_y}. \tag{92}$$

With these ingredients, the stationarity equations become

$$0 = \frac{\delta L}{\delta f_{y|\boldsymbol{m}}} = \lambda_{\boldsymbol{m}} + \frac{2\tilde{\mu}_{\mathcal{C}}}{\tilde{\sigma}_{\text{inn}}^2} \frac{\delta \tilde{\mu}_{\mathcal{C}}}{\delta f_{y|\boldsymbol{m}}} - \frac{\tilde{\mu}_{\mathcal{C}}^2}{\tilde{\sigma}_{\text{inn}}^4} \frac{\delta \tilde{\sigma}_{\text{inn}}^2}{\delta f_{y|\boldsymbol{m}}} \tag{93}$$

$$= \lambda_{\boldsymbol{m}} + \frac{2\tilde{\mu}_{\mathcal{C}}}{\tilde{\sigma}_{\text{inn}}^2} \binom{c}{\boldsymbol{m}} \mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right] \left[\frac{m_y}{\sqrt{q-1}} - \frac{\tilde{\mu}_{\mathcal{C}}}{2\tilde{\sigma}_{\text{inn}}^2 (q-1)} \right]; \tag{94}$$

$$0 = \frac{\delta L}{\delta f_{\boldsymbol{p}}} = \Lambda + \frac{2\tilde{\mu}_{\mathcal{C}}}{\tilde{\sigma}_{\text{inn}}^2} \frac{\delta \tilde{\mu}_{\mathcal{C}}}{\delta f_{\boldsymbol{p}}} - \frac{\tilde{\mu}_{\mathcal{C}}^2}{\tilde{\sigma}_{\text{inn}}^4} \frac{\delta \tilde{\sigma}_{\text{inn}}^2}{\delta f_{\boldsymbol{p}}} \tag{95}$$

$$= \Lambda + \frac{2\tilde{\mu}_{\mathcal{C}}}{\tilde{\sigma}_{\text{inn}}^2 \sqrt{q-1}} \sum_{y \in \mathcal{Q}} \left. \frac{\partial f_{y|\boldsymbol{p}}}{\partial p_y} \right|_{|\boldsymbol{p}|=1} - \frac{\tilde{\mu}_{\mathcal{C}}^2}{\tilde{\sigma}_{\text{inn}}^4 (q-1)} \sum_{y \in \mathcal{Q}} \frac{f_{y|\boldsymbol{p}}}{p_y}. \tag{96}$$

Equation (94) has to hold for all symbols y. This means that the expression $\mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right] \left[\frac{m_y}{\sqrt{q-1}} - \frac{\tilde{\mu}_{\mathcal{C}}}{2\tilde{\sigma}_{\text{inn}}^2 (q-1)} \right]$ has to be *independent* of y. This is a very complicated requirement on $f_{\boldsymbol{p}}$ and the attack; in general there is no easy way of solving it. However, if we take the Dirichlet distribution for $f_{\boldsymbol{p}}$, with parameter κ, then

$$\mathbb{E}_{\boldsymbol{P}} \left[\frac{\boldsymbol{P}^{\boldsymbol{m}}}{P_y} \right] = \frac{B(\kappa \mathbf{1}_q + \boldsymbol{m})}{B(\kappa \mathbf{1}_q)} \frac{\kappa q + c - 1}{m_y - (1-\kappa)} \tag{97}$$

and it becomes possible to satisfy the independence requirement by demanding

$$\tilde{\mu}_{\mathcal{C}} = 2\tilde{\sigma}_{\text{inn}}^2 (1-\kappa) \sqrt{q-1}. \tag{98}$$

With this special relation between $\tilde{\mu}_{\mathcal{C}}$ and $\tilde{\sigma}_{\text{inn}}^2$, (96) becomes

$$\forall \boldsymbol{p} : 0 = \Lambda + 4(1-\kappa) \sum_{y \in \mathcal{Q}} \left. \frac{\partial f_{y|\boldsymbol{p}}}{\partial p_y} \right|_{|\boldsymbol{p}|=1} - 4(1-\kappa)^2 \sum_{y \in \mathcal{Q}} \frac{f_{y|\boldsymbol{p}}}{p_y}. \tag{99}$$

We have to find strategy parameters $f_{y|\boldsymbol{m}}$ that give rise to a function $f_{y|\boldsymbol{p}}$ that satisfies (99). We happen to know from (46) that the interleaving attack satisfies

$$\sum_{y \in \mathcal{Q}} \left. \frac{\partial f_{y|\boldsymbol{p}}}{\partial p_y} \right|_{|\boldsymbol{p}|=1} = q - 1 \quad \text{and} \quad \sum_{y \in \mathcal{Q}} \frac{f_{y|\boldsymbol{p}}}{p_y} = q. \tag{100}$$

Thus (99) is satisfied if we take the interleaving attack and $\Lambda = 4(1-\kappa)^2 q - 4(1-\kappa)(q-1)$.

Next, we know that $\tilde{\sigma}_{\text{inn}}^2 = 1$ in case of a match since our suspicion function is normalized, and Proposition 3 tells us that $\tilde{\mu}_{\mathcal{C}} = \sqrt{q-1}$ for the interleaving match. The relationship (98) can only hold if $\kappa = 1/2$.

Thus, we have a stationary point in which the interleaving attack is used and

$$\begin{cases} \tilde{\mu}_{\mathcal{C}} & = \sqrt{q-1}; \\ \tilde{\sigma}_{\text{inn}}^2 & = 1; \\ \kappa & = \frac{1}{2}; \\ \Lambda & = -(q-2); \\ \lambda_{\boldsymbol{m}} & = -\frac{1}{2} \binom{c}{\boldsymbol{m}} \frac{B(\kappa \mathbf{1}_q + \boldsymbol{m})}{B(\kappa \mathbf{1}_q)} \left(\frac{q}{2} + c - 1 \right). \end{cases} \tag{101}$$

We have a match with $\tilde{\mu}_{\mathcal{C}}^2 / \tilde{\sigma}_{\text{inn}}^2 = q - 1$, which corresponds to asymptotic capacity as described in Proposition 3.

Part 2: We take an arbitrary point in $(f_{y|\boldsymbol{m}}, f_{\boldsymbol{p}})$-space and consider the infinitesimal steps

$$\begin{cases} \hat{f}_{y|\boldsymbol{m}} = f_{y|\boldsymbol{m}} + \Delta_{y\boldsymbol{m}} \\ \hat{f}_{\boldsymbol{p}} = f_{\boldsymbol{p}} + \beta(\boldsymbol{p}) \end{cases} \tag{102}$$

with $\sum_y \Delta_{y\boldsymbol{m}} = 0$ and $\int \mathrm{d}^q p\, \delta(|\boldsymbol{p}| - 1)\beta(\boldsymbol{p}) = 0$. In the new point we write

$$\hat{\tilde{\mu}}_{\mathcal{C}} = \tilde{\mu}_{(0)} + \tilde{\mu}_{(1)} + \tilde{\mu}_{(2)}; \quad \hat{\tilde{\sigma}}^2_{\text{inn}} = \tilde{\sigma}^2_{(0)} + \tilde{\sigma}^2_{(1)} + \tilde{\sigma}^2_{(2)} \tag{103}$$

where $\tilde{\mu}_{(0)} = \tilde{\mu}_{\mathcal{C}}$ and $\tilde{\sigma}^2_{(0)} = \tilde{\sigma}^2_{\text{inn}}$ refer to values in the original point $(f_{y|\boldsymbol{m}}, f_{\boldsymbol{p}})$, and we have defined

$$\tilde{\mu}_{(1)}\sqrt{q-1} = \sum_{y \in \mathcal{Q}} \sum_{\boldsymbol{m}} \int \mathrm{d}^q p\, \delta(|\boldsymbol{p}| - 1) \binom{c}{\boldsymbol{m}} \frac{\boldsymbol{p}^{\boldsymbol{m}}}{p_y} \cdot m_y \left[f_{\boldsymbol{p}} \Delta_{y\boldsymbol{m}} + \beta(\boldsymbol{p}) f_{y|\boldsymbol{m}} \right]; \tag{104}$$

$$\tilde{\sigma}^2_{(1)}(q-1) = \sum_{y \in \mathcal{Q}} \sum_{\boldsymbol{m}} \int \mathrm{d}^q p\, \delta(|\boldsymbol{p}| - 1) \binom{c}{\boldsymbol{m}} \frac{\boldsymbol{p}^{\boldsymbol{m}}}{p_y} \cdot \left[f_{\boldsymbol{p}} \Delta_{y\boldsymbol{m}} + \beta(\boldsymbol{p}) f_{y|\boldsymbol{m}} \right]; \tag{105}$$

$$\tilde{\mu}_{(2)}\sqrt{q-1} = \sum_{y \in \mathcal{Q}} \sum_{\boldsymbol{m}} \int \mathrm{d}^q p\, \delta(|\boldsymbol{p}| - 1) \binom{c}{\boldsymbol{m}} \frac{\boldsymbol{p}^{\boldsymbol{m}}}{p_y} \cdot m_y \beta(\boldsymbol{p}) \Delta_{y\boldsymbol{m}}; \tag{106}$$

$$\tilde{\sigma}^2_{(2)}(q-1) = \sum_{y \in \mathcal{Q}} \sum_{\boldsymbol{m}} \int \mathrm{d}^q p\, \delta(|\boldsymbol{p}| - 1) \binom{c}{\boldsymbol{m}} \frac{\boldsymbol{p}^{\boldsymbol{m}}}{p_y} \cdot \beta(\boldsymbol{p}) \Delta_{y\boldsymbol{m}}. \tag{107}$$

The subscript indicates the order of the small step. The maximum order is 2, since the expressions for $\tilde{\mu}_{\mathcal{C}}$ and $\tilde{\sigma}^2_{\text{inn}}$ are linear in both $f_{\boldsymbol{p}}$ and $f_{y|\boldsymbol{m}}$. We investigate the fraction $\hat{\tilde{\mu}}^2_{\mathcal{C}} / \hat{\tilde{\sigma}}^2_{\text{inn}}$.

$$\frac{\hat{\tilde{\mu}}^2_{\mathcal{C}}}{\hat{\tilde{\sigma}}^2_{\text{inn}}} = \frac{\left[\tilde{\mu}_{(0)} + \tilde{\mu}_{(1)} + \tilde{\mu}_{(2)}\right]^2}{\tilde{\sigma}^2_{(0)} + \tilde{\sigma}^2_{(1)} + \tilde{\sigma}^2_{(2)}} \tag{108}$$

$$= \frac{\tilde{\mu}^2_{(0)}}{\tilde{\sigma}^2_{(0)}} \cdot \frac{1 + 2\frac{\tilde{\mu}_{(1)}}{\tilde{\mu}_{(0)}} + 2\frac{\tilde{\mu}_{(2)}}{\tilde{\mu}_{(0)}} + \frac{[\tilde{\mu}_{(1)}]^2}{\tilde{\mu}^2_{(0)}} + \cdots}{1 + \frac{\tilde{\sigma}^2_{(1)}}{\tilde{\sigma}^2_{(0)}} + \frac{\tilde{\sigma}^2_{(2)}}{\tilde{\sigma}^2_{(0)}}} \tag{109}$$

$$= \frac{\tilde{\mu}^2_{(0)}}{\tilde{\sigma}^2_{(0)}} \cdot \left[1 + 2\frac{\tilde{\mu}_{(1)}}{\tilde{\mu}_{(0)}} + 2\frac{\tilde{\mu}_{(2)}}{\tilde{\mu}_{(0)}} + \frac{\left[\tilde{\mu}_{(1)}\right]^2}{\left[\tilde{\mu}_{(0)}\right]^2} + \cdots \right] \cdot \left[1 - \frac{\tilde{\sigma}^2_{(1)}}{\tilde{\sigma}^2_{(0)}} - \frac{\tilde{\sigma}^2_{(2)}}{\tilde{\sigma}^2_{(0)}} + \frac{\left[\tilde{\sigma}^2_{(1)}\right]^2}{\left[\tilde{\sigma}^2_{(0)}\right]^2} + \cdots \right] \tag{110}$$

where the dots stand for higher order terms. In the last line we did a Taylor expansion of the denominator. By collecting equal order terms in (110) we obtain the first and second order components

$$\left[\frac{\tilde{\mu}^2_{\mathcal{C}}}{\tilde{\sigma}^2_{\text{inn}}}\right]_{(1)} = \frac{\tilde{\mu}^2_{(0)}}{\tilde{\sigma}^2_{(0)}} \cdot \left[2\frac{\tilde{\mu}_{(1)}}{\tilde{\mu}_{(0)}} - \frac{\tilde{\sigma}^2_{(1)}}{\tilde{\sigma}^2_{(0)}} \right] \tag{111}$$

and

$$\left[\frac{\tilde{\mu}^2_{\mathcal{C}}}{\tilde{\sigma}^2_{\text{inn}}}\right]_{(2)} = \frac{\tilde{\mu}^2_{(0)}}{\tilde{\sigma}^2_{(0)}} \left[2\frac{\tilde{\mu}_{(2)}}{\tilde{\mu}_{(0)}} + \frac{\left[\tilde{\mu}_{(1)}\right]^2}{\left[\tilde{\mu}_{(0)}\right]^2} - \frac{\tilde{\sigma}^2_{(2)}}{\tilde{\sigma}^2_{(0)}} + \frac{\left[\tilde{\sigma}^2_{(1)}\right]^2}{\left[\tilde{\sigma}^2_{(0)}\right]^2} - 2\frac{\tilde{\mu}_{(1)}}{\tilde{\mu}_{(0)}}\frac{\tilde{\sigma}^2_{(1)}}{\tilde{\sigma}^2_{(0)}} \right] \tag{112}$$

$$= \frac{\tilde{\mu}^2_{(0)}}{\tilde{\sigma}^2_{(0)}} \left[2\frac{\tilde{\mu}_{(2)}}{\tilde{\mu}_{(0)}} - \frac{\tilde{\sigma}^2_{(2)}}{\tilde{\sigma}^2_{(0)}} \right] + \frac{\tilde{\mu}^2_{(0)}}{\tilde{\sigma}^2_{(0)}} \left[\frac{\tilde{\mu}_{(1)}}{\tilde{\mu}_{(0)}} - \frac{\tilde{\sigma}^2_{(1)}}{\tilde{\sigma}^2_{(0)}} \right]^2. \tag{113}$$

We take the stationary point as our starting point and first make a step in the $f_{\boldsymbol{p}}$-direction only, i.e. $\Delta_{y\boldsymbol{m}} = 0$.

The fact that $\Delta_{y\boldsymbol{m}} = 0$ yields $\tilde{\mu}_{(2)} = 0$ and $\tilde{\sigma}^2_{(2)} = 0$ from (106) and (107). Furthermore, in (104) and (105) note that the sum over y yields a constant as in (100), and then integrating β gives zero. So $\tilde{\mu}_{(1)} = 0$ and $\tilde{\sigma}^2_{(1)} = 0$ as well. Thus we conclude that from the stationary point, changing only $f_{\boldsymbol{p}}$ does not change the performance indicator $\tilde{\mu}^2_{\mathcal{C}}/\tilde{\sigma}^2_{\text{inn}}$. Note that this is consistent with Proposition 3.

Secondly we fix $f_{\boldsymbol{p}}$ to be the Dirichlet distribution with $\kappa = \frac{1}{2}$ and vary the attack slightly from interleaving. Now $\beta = 0$, which yields $\tilde{\mu}_{(2)} = 0$ and $\tilde{\sigma}^2_{(2)} = 0$. Equation (113) with $\beta = 0$ then reduces to a square, which is non-negative. Thus the performance indicator is minimized when the interleaving attack is used, and the found stationary point is indeed a saddlepoint. ■

This saddlepoint leads to a global minimum:

Theorem 2: Assume the tracer uses the interleaving defense and the Dirichlet distribution with $\kappa = \frac{1}{2}$. Then the interleaving attack minimizes the performance indicator $\frac{\tilde{\mu}_{\mathcal{C}}}{\tilde{\sigma}_{\text{inn}}}$.

Proof: From Proposition 8 we know that the interleaving attack is a local minimum in this setting. Also, when the distribution is fixed as the Dirichlet distribution with $\kappa = \frac{1}{2}$, the proof of Proposition 8 states that the second derivative (113) is non-negative for any strategy, as $\beta(\boldsymbol{p}) = 0$ implies that $\tilde{\mu}_{(2)} = \tilde{\sigma}_{(2)} = 0$. Since $\frac{\tilde{\mu}_{\mathcal{C}}}{\tilde{\sigma}_{\text{inn}}}$ is a rational function of $\Delta_{y\boldsymbol{m}}$, we can conclude that the interleaving attack is a *global* minimum for this setting. ■

C. Relation to the Tardos Suspicion Function

The interleaving defense is closely related to the Tardos suspicion function:

Proposition 9: The symmetric Tardos function is the strongly normalized optimal suspicion function against the interleaving attack.

Proof: We know from (51) that $\text{Var}_{X|\boldsymbol{p}}[T] = \frac{1-p_y}{c^2 p_y}$. So by Theorem 2, the strongly normalized optimal suspicion function against the interleaving attack is

$$h(x, y, \boldsymbol{p}) = \sqrt{\frac{p_y}{1-p_y}} \left(\frac{\delta_{x,y}}{p_y} - 1 \right), \tag{114}$$

which equals the symmetric Tardos function (3). ■

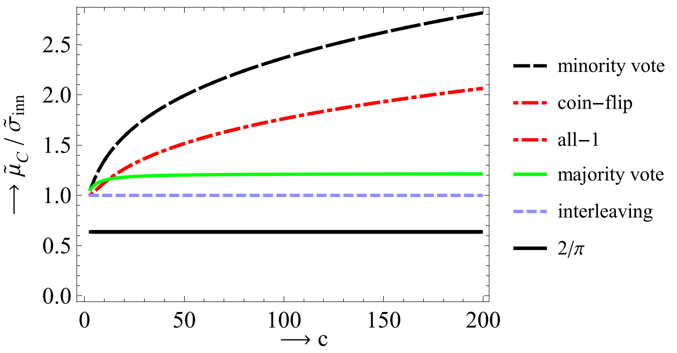

Fig. 2. Interleaving defense against various attacks in the binary case.

TABLE I
NUMERICAL TRENDS FOR THE PERFORMANCE INDICATOR $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\rm inn}$ OF THE INTERLEAVING DEFENSE IN THE BINARY CASE $q = 2$ FOR LARGE c

interleaving attack	1.0
all-1 attack	$0.61c^{0.23}$
coin-flip attack	$0.61c^{0.23}$
majority voting attack	1.2
minority voting attack	$0.75c^{0.25}$

D. Interleaving Defense Numerics

To verify our analytic results and their practical applicability, we ran simulations for the binary case and the arcsine distribution (without cut-off), which is equal to the Dirichlet distribution with $\kappa = \frac{1}{2}$. We simulated the five described attacks (interleaving, all-1, coin-flip, majority voting, and minority voting) against the interleaving defense. We stress that these five attacks are by no means exhaustive.

We ran simulations for $1 \leq c \leq 200$ to obtain the $\tilde{\mu}_{\mathcal{C}}$ and the $\tilde{\sigma}_{\rm inn}$ in these five cases as depicted in Fig.2. We then analyzed this data to obtain the leading-order term in c. The results can be found in Table I. Since for mismatches the innocent score is no longer normalized ($\tilde{\sigma}_{\rm inn} \neq 1$), we present the results for $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\rm inn}$ to make a fair comparison.

As predicted by Theorem 2, the interleaving defense attains capacity ($\tilde{\mu} = 1$) against the interleaving attack. We also observe that the majority voting attack has a constant $\tilde{\mu}$. For the other three attacks, $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\rm inn}$ seems to grow as $c^{1/4}$. We were able to prove this for the all-1 and coin-flip attacks in Proposition 7.

VI. ALL-HIGH DEFENSE

A. Optimal Defense

The all-high attack

$$f_{y|\boldsymbol{m}} = \begin{cases} 1 & \text{if } m_y > 0 \text{ and } m_{y+1} = \cdots = m_{q-1} = 0 \\ 0 & \text{else} \end{cases} \tag{115}$$

outputs the highest symbol among those received by the coalition.

Note that this is the only attack we consider that breaks symbol symmetry and assumes an ordering of the alphabet. This is a special case of the so-called *preferred-sequence attack*, in which the colluders have a predetermined ranking of the symbols. The results below generalize to the preferred-sequence attack. Recall our shorthand notation $a_k := (p_0 + \cdots + p_{k-1})$ and $a_{\mathcal{B}} = \sum_{\beta \in \mathcal{B}} p_\beta$.

Proposition 10: Against the all-high attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with

$$T(x, y, \boldsymbol{p}) = \begin{cases} (a_{y+1}^{c-1} - a_y^{c-1})/(a_{y+1}^c - a_y^c) & \textit{if } \; x < y \\ a_{y+1}^{c-1}/(a_{y+1}^c - a_y^c) & \textit{if } \; x = y \\ 0 & \textit{if } \; x > y. \end{cases} \tag{116}$$

In case of a match, it holds that

$$\tilde{\mu}_{\mathcal{C}} = c\sqrt{-1 + \mathbb{E}_{\boldsymbol{P}}\left[\sum_{y=0}^{q-1} \frac{A_{y+1}^{2c-1} - 2A_y^c A_{y+1}^{c-1} + A_y^{2c-1}}{A_{y+1}^c - A_y^c}\right]}. \tag{117}$$

Proof: We find

$$\begin{aligned} f_{y|\boldsymbol{p}} &= \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{y|\boldsymbol{M}}] && (118) \\ &= \mathbb{P}[M_y > 0, M_{y+1} = \cdots = M_{q-1} = 0] && (119) \\ &= \mathbb{P}[M_{y+1} = \cdots = M_{q-1} = 0] \\ &\quad - \mathbb{P}[M_y = \cdots = M_{q-1} = 0] && (120) \\ &= \frac{a_{y+1}^c}{|\boldsymbol{p}|^c} - \frac{a_y^c}{|\boldsymbol{p}|^c}. && (121) \end{aligned}$$

So

$$\begin{aligned} T &= \frac{1}{c} \left. \frac{\partial \ln(|\boldsymbol{p}|^c f_{y|\boldsymbol{p}})}{\partial p_x} \right|_{|\boldsymbol{p}|=1} && (122) \\ &= \begin{cases} \frac{a_{y+1}^{c-1} - a_y^{c-1}}{a_{y+1}^c - a_y^c} & \text{if } x < y \\ \frac{a_{y+1}^{c-1}}{a_{y+1}^c - a_y^c} & \text{if } x = y \\ 0 & \text{if } x > y. \end{cases} && (123) \end{aligned}$$

Also,

$$\begin{aligned} \mathbb{E}[T^2] &= \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{Y|\boldsymbol{P}} \mathbb{E}_{X|\boldsymbol{P}}[T^2(X, Y, \boldsymbol{P})] && (124) \\ &= \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{Y|\boldsymbol{P}} \left[P_Y T^2(Y, Y, \boldsymbol{P}) + A_Y T^2(0, Y, \boldsymbol{P})\right] && (125) \\ &= \mathbb{E}_{\boldsymbol{P}} \sum_{y=0}^{q-1} \left[P_y \frac{A_{y+1}^{2(c-1)}}{A_{y+1}^c - A_y^c} + A_y \frac{\left(A_{y+1}^{c-1} - A_y^{c-1}\right)^2}{A_{y+1}^c - A_y^c}\right] && (126) \\ &= \mathbb{E}_{\boldsymbol{P}} \sum_{y=0}^{q-1} \frac{A_{y+1}^{2c-1} - 2A_y^c A_{y+1}^{c-1} + A_y^{2c-1}}{A_{y+1}^c - A_y^c}. && (127) \end{aligned}$$

We obtain (117) using $\tilde{\mu}_{\mathcal{C}} = c\sqrt{\mathrm{Var}[T]} = c\sqrt{\mathbb{E}[T^2] - 1}$. ■

When $x = y$, the h is positive. When $x > y$, it is negative and constant. When $x < y$, it might be negative or it might not. For instance, for $c = 2$, we find $(a_{y+1} - a_y)/(a_{y+1}^2 - a_y^2) = 1/(a_{y+1} + a_y) = 1/(p_y + 2a_y)$, in which case h is negative if and only if $p_y > 1 - 2a_y$. In particular it is negative if $a_y \geq \frac{1}{2}$. Also, h is the same for all $x < y$.

We now analyze the behaviour of $\tilde{\mu}_{\mathcal{C}}$ when the symmetric Dirichlet distribution is employed. Before we can state our result, we will need the following Lemma:

Lemma 7: Let $f_{\boldsymbol{P}}$ be the symmetric Dirichlet distribution without cutoff. The joint distribution for the pair $(A_{y+1}, A_y/A_{y+1})$ is then given by

$$J(a_{y+1}, \frac{a_y}{a_{y+1}}) = \frac{a_{y+1}^{-1+(y+1)\kappa}(1-a_{y+1})^{-1+(q-y-1)\kappa}}{B([y+1]\kappa, [q-y-1]\kappa)} \times \frac{(a_y/a_{y+1})^{-1+y\kappa}(1-a_y/a_{y+1})^{-1+\kappa}}{B(y\kappa, \kappa)}.$$

Proof: We first derive the joint distribution $J(a_y, a_{y+1})$ for A_y and A_{y+1}:

$$\begin{aligned} J(a_y, a_{y+1}) &= \mathbb{E}_{\boldsymbol{P}}\left[\delta\left[A_y - \sum_{i=0}^{y-1} P_i\right]\delta\left[A_{y+1} - \sum_{i=0}^{y} P_i\right]\right] && (128)\\ &\propto \int_{|\boldsymbol{p}|=1} \mathrm{d}^{q-1}p\ \boldsymbol{p}^{\kappa-1}\delta\left[a_y - \sum_{i=0}^{y-1} p_i\right]\delta\left[a_{y+1} - \sum_{i=0}^{y} p_i\right] && (129)\\ &= \int \mathrm{d}^q p\ \boldsymbol{p}^{\kappa-1}\delta\left[a_y - \sum_{i=0}^{y-1} p_i\right] \\ &\quad \times\delta\left[a_{y+1} - \sum_{i=0}^{y} p_i\right]\delta(1-|\boldsymbol{p}|) && (130)\\ &= \int \mathrm{d}^q p\ \boldsymbol{p}^{\kappa-1}\delta\left[a_y - \sum_{i=0}^{y-1} p_i\right] \\ &\quad \times\delta\left[a_{y+1} + \sum_{i=y+1}^{q-1} p_i - 1\right]\delta(1-|\boldsymbol{p}|). && (131) \end{aligned}$$

Here $\delta(x)$ is the Dirac delta function. We perform the following change of variables: for $i < y$ we define $p_i = a_y s_i$; for $i > y$ we define $p_i = (1-a_{y+1})t_i$. This yields $\mathrm{d}^q p\ \boldsymbol{p}^{\kappa-1}= \mathrm{d}^y s\ \mathrm{d}p_y \mathrm{d}^{q-y-1}t\ p_y^{\kappa-1}a_y^{y\kappa}\boldsymbol{s}^{\kappa-1}(1-a_{y+1})^{[q-y-1]\kappa}\boldsymbol{t}^{\kappa-1}$ and

$$\delta(a_y - \textstyle\sum_{i=0}^{y-1} p_i) = a_y^{-1}\delta(1-|\boldsymbol{s}|), \tag{132}$$

$$\delta\left[a_{y+1} + \sum_{i=y+1}^{q-1} p_i - 1\right] = (1-a_{y+1})^{-1}\delta(1-|\boldsymbol{t}|), \tag{133}$$

$$\delta(1-|\boldsymbol{p}|) = \delta\left[1-p_y - a_y|\boldsymbol{s}| - (1-a_{y+1})|\boldsymbol{t}|\right]. \tag{134}$$

The expression (131) becomes

$$\begin{aligned} J(a_y, a_{y+1}) &= \int \mathrm{d}^y s\ \mathrm{d}p_y\ \mathrm{d}^{q-y-1}t\ p_y^{\kappa-1}a_y^{y\kappa-1}\boldsymbol{s}^{\kappa-1} \\ &\quad \times(1-a_{y+1})^{[q-y-1]\kappa-1}\boldsymbol{t}^{\kappa-1}\delta(1-|\boldsymbol{s}|) \\ &\quad \times\delta(1-|\boldsymbol{t}|)\delta\left[p_y + a_y - a_{y+1}\right] && (135)\\ &\propto a_y^{y\kappa-1}(1-a_{y+1})^{[q-y-1]\kappa-1}(a_{y+1}-a_y)^{\kappa-1}. && (136) \end{aligned}$$

Finally we do a last change of variables from a_y to $z = a_y/a_{y+1}$. This gives $\mathrm{d}a_y\,\mathrm{d}a_{y+1} = a_{y+1}\,\mathrm{d}a_{y+1}\,\mathrm{d}z$, and (136) becomes

$$J(a_{y+1}, z) \propto a_{y+1}^{[y+1]\kappa-1}(1-a_{y+1})^{[q-y-1]\kappa-1}z^{y\kappa-1}(1-z)^{\kappa-1}. \tag{137}$$

Inserting the normalization constants yields the result of the lemma. ■

Given this joint distribution, we can now derive our main result for the all-high attack when the symmetric Dirichlet distribution is used.

Proposition 11: Let $f_{\boldsymbol{P}}$ be the symmetric Dirichlet distribution without cutoff. If the attack is the all-high attack and the defense matches it, then, for large c,

$$\tilde{\mu}_{\mathcal{C}} = c^{\frac{1-\kappa}{2}}\sqrt{\frac{\kappa\Gamma(q\kappa)\zeta(1+\kappa)}{\Gamma([q-1]\kappa)}}\left[1+\mathcal{O}(c^{-\min(1,\kappa)})\right], \tag{138}$$

where ζ is the Riemann zeta function.

Proof: We write (117) as

$$\frac{\tilde{\mu}_{\mathcal{C}}^2}{c^2} = -1 + \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol{P}} \times\left[A_{y+1}^{c-1}\frac{1-2(A_y/A_{y+1})^c+(A_y/A_{y+1})^{2c-1}}{1-(A_y/A_{y+1})^c}\right]. \tag{139}$$

The fraction can be expanded as

$$\frac{1}{1-(A_y/A_{y+1})^c} = \sum_{t=0}^{\infty}(A_y/A_{y+1})^{tc}. \tag{140}$$

Then we evaluate the expectation using the joint distribution $J(a_{y+1}, \frac{a_y}{a_{y+1}})$ from Lemma 7. This yields

$$\frac{\tilde{\mu}_{\mathcal{C}}^2}{c^2} = -1 + \sum_{y=0}^{q-1}\frac{B([y+1]\kappa+c-1, [q-y-1]\kappa)}{B([y+1]\kappa, [q-y-1]\kappa)} \times\left[1-\sum_{t=1}^{\infty}\frac{B(y\kappa+tc, \kappa)}{B(y\kappa, \kappa)}+\sum_{t=2}^{\infty}\frac{B(y\kappa+tc-1, \kappa)}{B(y\kappa, \kappa)}\right] \tag{141}$$

noting that $1/B(y\kappa, \kappa)$ vanishes for $y = 0$. Further simplification gives

$$\frac{\tilde{\mu}_{\mathcal{C}}^2}{c^2} = \frac{\kappa\Gamma(q\kappa)}{\Gamma(q\kappa+c-1)} \times\left[\sum_{y=0}^{q-2}\frac{\Gamma([y+1]\kappa+c-1)}{([y+2]\kappa+c-1)\Gamma([y+1]\kappa)} +\sum_{y=1}^{q-1}\frac{\Gamma([y+1]\kappa+c-1)}{\Gamma(y\kappa)}\sum_{t=2}^{\infty}\frac{\Gamma(y\kappa+tc-1)}{\Gamma([y+1]\kappa+tc)}\right]. \tag{142}$$

Finally we use the identity $\Gamma(c+a)/\Gamma(c+b) = c^{a-b}[1+\mathcal{O}(c^{-1})]$ to investigate the asymptotics. In the first summation over y the dominant term occurs at $y = q-2$, thus the summation can be simplified to $c^{-\kappa-1}[1+\mathcal{O}(c^{-\min(1,\kappa)})]\cdot \kappa\Gamma(q\kappa)/\Gamma([q-1]\kappa)$. Similarly, in the second summation over y the dominant term occurs at $y = q - 1$ and thus this summation reduces to $c^{-\kappa-1}[1+\mathcal{O}(c^{-\min(1,\kappa)})][\zeta(1+\kappa)-1]$ $\kappa\Gamma(q\kappa)/\Gamma([q-1]\kappa)$, where ζ is the Riemann zeta function. ■

B. All-1 Defense

The binary all-high attack is known as the all-1 attack. It has $f_{1|\boldsymbol{m}} = 1$ whenever $m_1 > 0$ and $f_{1|\boldsymbol{m}} = 0$ when $m_1 = 0$.

Corollary 4: Against the all-1 attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with

$$T(x, y, \boldsymbol{p}) = \begin{cases} (1-p_0^{c-1})/(1-p_0^c) & \textit{if } (x,y) = (0,1) \\ 1/(1-p_0^c) & \textit{if } (x,y) = (1,1) \\ 1/p_0 & \textit{if } (x,y) = (0,0) \\ 0 & \textit{if } (x,y) = (1,0). \end{cases} \tag{143}$$

In case of a match it holds that

$$\tilde{\mu}_{\mathcal{C}} = c\sqrt{\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1}(1-P_0)/(1-P_0^c)]}. \tag{144}$$

Proof: T follows directly from Proposition 10. Furthermore, (117) gives

$$\mathrm{Var}[T] = -1 + \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + \frac{1-2P_0^c+P_0^{2c-1}}{1-P_0^c}\right] \tag{145}$$

$$= \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + \frac{P_0^{2c-1}-P_0^c}{1-P_0^c}\right] = \mathbb{E}_{\boldsymbol{P}}\left[\frac{P_0^c(1-P_0)}{P_0(1-P_0^c)}\right]. \tag{146}$$

■

When $x < y$, the h is positive for any c, in contrast to the q-ary case.

Corollary 5: Let $f_{\boldsymbol{P}}$ be the symmetric Dirichlet distribution with $\kappa = \frac{1}{2}$ and cutoff $\delta = 0$. Against the all-1 attack, the optimal suspicion function attains $\tilde{\mu}_{\mathcal{C}} \propto c^{1/4}$ for large c.

Before we investigate the behaviour of the all-high defense against an interleaving attack, we first prove a general lemma about the interleaving attack.

Lemma 8: If the tracer uses a strongly centered score function and the coalition uses the interleaving attack, then

$$\tilde{\mu}_{\mathcal{C}} = \sum_{y\in\mathcal{A}} \mathbb{E}_{\boldsymbol{P}}[P_y\, h(y, y, \boldsymbol{P})]. \tag{147}$$

Proof: For the interleaving attack, $cT = \frac{\delta_{x,y}}{p_y} + c - 1$, so

$$\tilde{\mu}_{\mathcal{C}} = c\mathbb{E}[T\cdot h] \tag{148}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\left(\frac{\delta_{X,Y}}{P_Y} + c - 1\right)h(X, Y, \boldsymbol{P})\right] \tag{149}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{\delta_{X,Y}}{P_Y}\, h(X, Y, \boldsymbol{P})\right] \tag{150}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}[h(Y, Y, \boldsymbol{P})]. \tag{151}$$

where (151) holds since $\mathbb{E}[h] = 0$. ■

The performance of the all-high defense against the interleaving attack can be analyzed as follows:

Proposition 12: If the tracer uses the all-high defense but the coalition uses the interleaving attack, then

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\, \mathbb{E}_{\boldsymbol{P}}\left[\sum_{y=1}^{q-1} \frac{P_y A_{y+1}^{c-1}}{A_{y+1}^c - A_y^c}\right] \tag{152}$$

where T belongs to the all-high defense.

Proof: Applying Lemma 8 we obtain

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\left(-1 + \mathbb{E}_{\boldsymbol{P}}\left[\sum_{y\in\mathcal{A}} \frac{P_y A_{y+1}^{c-1}}{A_{y+1}^c - A_y^c}\right]\right) \tag{153}$$

$$= \frac{1}{\sqrt{\mathrm{Var}[T]}}\, \mathbb{E}_{\boldsymbol{P}}\left[\sum_{y=1}^{q-1} \frac{P_y A_{y+1}^{c-1}}{A_{y+1}^c - A_y^c}\right]. \tag{154}$$

■

In the binary case this reduces to

Proposition 13: For $q = 2$, if the tracer uses the all-1 defense, but the coalition uses the interleaving attack, then

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\mathbb{E}_{\boldsymbol{P}}\left[P_1 \sum_{k=0}^{\infty} P_0^{kc}\right]. \tag{155}$$

Proof: Applying Lemma 8 we obtain

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\left(-1 + \mathbb{E}_{\boldsymbol{P}}\left[1 + \frac{P_1}{1-P_0^c}\right]\right). \tag{156}$$

■

The scaling behaviour for large c is

Lemma 9: Let $q = 2$ and $f_{\boldsymbol{P}}$ be the symmetric Dirichlet distribution with parameter $\kappa = \frac{1}{2}$ without cutoff. If the tracer uses the all-1 defense, but the coalition uses the interleaving attack, then

$$\tilde{\mu}_{\mathcal{C}} = \frac{\Gamma(\kappa+1)}{B(\kappa,\kappa)\sqrt{\mathrm{Var}[T]}} \sum_{t=0}^{\infty} \frac{\Gamma(tc+\kappa)}{\Gamma(tc+2\kappa+1)}. \tag{157}$$

For large c, this scales as $c^{(\kappa+1)/2}$.

C. All-1 Defense Numerics

We ran simulations for the binary case and the arcsine distribution (without cut-off) with the same parameters as described in Section V-D. The table looks very similar to that of the interleaving defense. As expected, the all-1 defense performs better against the all-1 attack, but worse against the other four attacks. However, it retains the same scaling behaviour.

We again stress that these five attacks are by no means exhaustive.

VII. Random-Symbol Defense

A. Optimal Defense

The random-symbol attack selects one of the received symbols uniformly at random. Tallies are disregarded, but a symbol can only be chosen if its tally is nonzero. The attack is parametrized by $f_{y|\boldsymbol{m}} = (1-\delta_{m_y,0})/|\{\alpha \in \mathcal{A} : m_\alpha > 0\}|$.

TABLE II

NUMERICAL TRENDS FOR THE PERFORMANCE INDICATOR $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ OF THE ALL-1 DEFENSE IN THE BINARY CASE $q = 2$ FOR LARGE c

interleaving attack	0.71
all-1 attack	$0.86c^{0.25}$
coin-flip attack	$0.44c^{0.23}$
majority voting attack	0.84
minority voting attack	$0.54c^{0.25}$

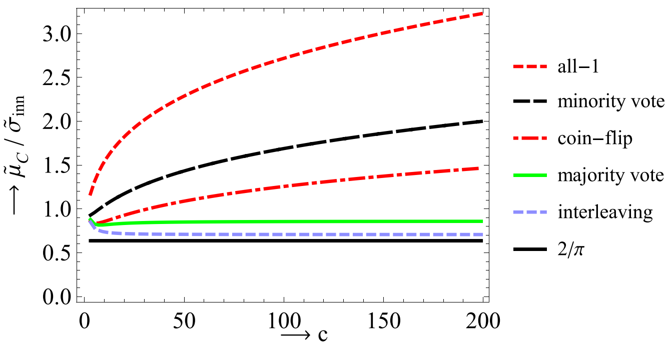

Fig. 3. All-1 defense against various attacks in the binary case.

Proposition 14: For the random-symbol attack we find

$$|\boldsymbol{p}|^c f_{y|\boldsymbol{p}} = \frac{a^c_{\mathcal{A}} - a^c_{\mathcal{A}\setminus\{y\}}}{q} + \sum_{\mathcal{B}\subsetneq\mathcal{A}:\ y\in\mathcal{B}} \frac{a^c_{\mathcal{B}} - a^c_{\mathcal{B}\setminus\{y\}}}{|\mathcal{B}|(|\mathcal{B}|+1)}. \tag{158}$$

The optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, *with*

$$T(x, y, \boldsymbol{p}) = \frac{1}{c}\frac{\partial \ln(|\boldsymbol{p}|^c f_{y|\boldsymbol{p}})}{\partial p_x}\bigg|_{|\boldsymbol{p}|=1} \tag{159}$$

$$= \begin{cases} \frac{1}{f_{y|\boldsymbol{p}}}\left(\frac{1}{q} + \sum\limits_{\mathcal{B}\subsetneq\mathcal{A}:\ y\in\mathcal{B}} \frac{a^{c-1}_{\mathcal{B}}}{|\mathcal{B}|(|\mathcal{B}|+1)}\right) & \text{if } x = y \\ \frac{1}{f_{y|\boldsymbol{p}}}\left(\frac{1-(1-p_y)^{c-1}}{q} + \sum\limits_{\substack{\mathcal{B}\subsetneq\mathcal{A} \\ x,y\in\mathcal{B}}} \frac{a^{c-1}_{\mathcal{B}} - a^{c-1}_{\mathcal{B}\setminus\{y\}}}{|\mathcal{B}|(|\mathcal{B}|+1)}\right) & \text{if } x \neq y \end{cases} \tag{160}$$

Proof: For the random-symbol attack, the probability $f_{y|\boldsymbol{m}}$ that the symbol y is produced, is 0 if $m_y = 0$. It is $\frac{1}{q}$ if for all $\alpha \in \mathcal{A}$, $m_\alpha > 0$. It is $\frac{1}{q-1}$ if $m_y > 0$ and there is exactly one symbol $\alpha_1 \in \mathcal{A}$ for which $m_{\alpha_1} = 0$. It is $\frac{1}{q-2}$ if $m_y > 0$ and there are exactly two distinct symbols $\alpha_1, \alpha_2 \in \mathcal{A}$ for which $m_{\alpha_1} = m_{\alpha_2} = 0$, etc. This can be written in additive form using indicator functions:

$$\begin{aligned} f_{y|\boldsymbol{m}} = {} & \tfrac{1}{q}\mathbf{1}_{\{m_y>0\}} \\ & + \left(\tfrac{1}{q-1} - \tfrac{1}{q}\right)\mathbf{1}_{\{m_y>0\}}\mathbf{1}_{\{\exists\alpha_1:m_{\alpha_1}=0\}} \\ & + \left(\tfrac{1}{q-2} - \tfrac{1}{q-1}\right)\mathbf{1}_{\{m_y>0\}}\mathbf{1}_{\{\exists\alpha_1:m_{\alpha_1}=0\}}\mathbf{1}_{\{\exists\alpha_2\neq\alpha_1:m_{\alpha_2}=0\}} \\ & + \cdots + \left(1 - \tfrac{1}{2}\right)\mathbf{1}_{\{m_y>0\}} \\ & \cdot \mathbf{1}_{\{\exists\alpha_1:m_{\alpha_1}=0\}} \cdots \mathbf{1}_{\{\exists\alpha_{q-1}\neq\alpha_1,\ldots\alpha_{q-2}:m_{\alpha_{q-1}}=0\}}. \end{aligned} \tag{161}$$

Note that

$$\mathbb{P}[M_y > 0] = \mathbb{P}[M_y \geq 0] - \mathbb{P}[M_y = 0] = \frac{A^c_{\mathcal{A}} - A^c_{\mathcal{A}\setminus\{y\}}}{|\boldsymbol{p}|^c} \tag{162}$$

and for each proper subset $\mathcal{B} \subsetneq \mathcal{A}$ with $y \in \mathcal{B}$, it holds that

$$\mathbb{P}[\forall\beta \in \mathcal{B}, M_\beta > 0] = \left(A^c_{\mathcal{B}} - A^c_{\mathcal{B}\setminus\{y\}}\right)/|\boldsymbol{p}|^c. \tag{163}$$

Since $f_{y|\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{y|\boldsymbol{M}}]$, and for all sets $\mathcal{V}, \mathcal{W}$, it holds that $\mathbf{1}_{\mathcal{V}}\mathbf{1}_{\mathcal{W}} = \mathbf{1}_{\mathcal{V}\cap\mathcal{W}}$, and $\mathbb{E}[\mathbf{1}_{\mathcal{V}}] = \mathbb{P}[\mathcal{V}]$, we find

$$\begin{aligned} |\boldsymbol{p}|^c f_{y|\boldsymbol{p}} = {} & \frac{a^c_{\mathcal{A}} - a^c_{\mathcal{A}\setminus\{y\}}}{q} \\ & + \sum_{\mathcal{B}\subsetneq\mathcal{A}:\ y\in\mathcal{B}} \left(\frac{1}{|\mathcal{B}|} - \frac{1}{|\mathcal{B}|+1}\right)\left(a^c_{\mathcal{B}} - a^c_{\mathcal{B}\setminus\{y\}}\right). \end{aligned} \tag{164}$$

which simplifies to equation (158). ■

B. Coin-Flip Defense

The binary random-symbol attack is known as the coin-flip attack, and is parametrized as $f_{y|\boldsymbol{m}} = \frac{1}{2}(1 - \delta_{m_y,0} + \delta_{m_y,c})$.

Proposition 15: Against the coin-flip attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, *with*

$$T(x, y, \boldsymbol{p}) = \begin{cases} (1 + p_y^{c-1})/(1 + p_y^c - p_{1-y}^c) & \text{if } x = y \\ (1 - p_{1-y}^{c-1})/(1 + p_y^c - p_{1-y}^c) & \text{if } x \neq y. \end{cases} \tag{165}$$

Proof: Since

$$f_{y|\boldsymbol{m}} = \tfrac{1}{2}(1 - \delta_{m_y,0} + \delta_{m_y,c}). \tag{166}$$

$$f_{y|\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{y|\boldsymbol{M}}] \tag{167}$$

$$= \frac{1}{2|\boldsymbol{p}|^c}\sum_{m_y=0}^{c}\binom{c}{m_y}p_y^{m_y}p_{1-y}^{c-m_y}(1 - \delta_{m_y,0} + \delta_{m_y,c}) \tag{168}$$

$$= \frac{1}{2|\boldsymbol{p}|^c}[(p_y + p_{1-y})^c - p_{1-y}^c + p_y^c]. \tag{169}$$

Thus

$$\begin{aligned} \frac{\partial(|\boldsymbol{p}|^c f_{y|\boldsymbol{p}})}{\partial p_x} = {} & \tfrac{1}{2}c[(p_y + p_{1-y})^{c-1} - (1 - \delta_{x,y})p_{1-y}^{c-1} \\ & + \delta_{x,y}p_y^{c-1}] \end{aligned} \tag{170}$$

So

$$T = \frac{1 - (1 - \delta_{x,y})p_{1-y}^{c-1} + \delta_{x,y}p_y^{c-1}}{1 - p_{1-y}^c + p_y^c}. \tag{171}$$

■

When $x = y$, the h is positive. When $x \neq y$, it is negative, since $-p_{1-y}^{c-1} < p_y^{c-1}$, so $p_{1-y}^{c-1}(p_{1-y} - 1) < p_y^c$, and thus $1 - p_{1-y}^{c-1} < 1 + p_y^c - p_{1-y}^c$.

The interleaving attack against the coin-flip defense behaves as follows in the binary case:

TABLE III

NUMERICAL TRENDS FOR THE PERFORMANCE INDICATOR $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ OF THE COIN-FLIP DEFENSE IN THE BINARY CASE $q = 2$ FOR LARGE c

interleaving attack	$5.1c^{-0.71}$
all-1 attack	$0.72c^{0.25}$
coin-flip attack	$0.72c^{0.25}$
majority voting attack	0.0
minority voting attack	$1.1c^{0.25}$

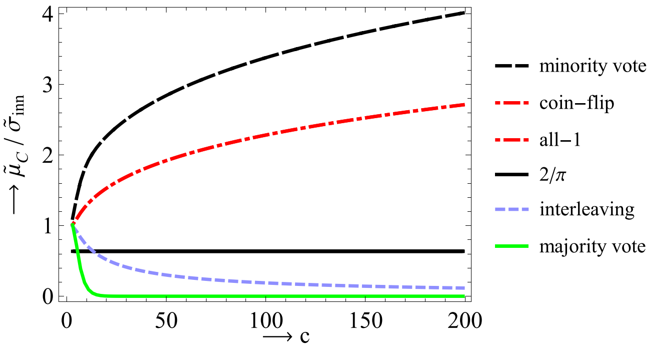

Fig. 4. Coin-flip defense against various attacks in the binary case.

Lemma 10: For $q = 2$, if the tracer uses the coin-flip defense, but the coalition uses the interleaving attack, then

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}} \times \left[-1 + \mathbb{E}_{\boldsymbol{P}}\left[\frac{P_0(1+P_0)^{c-1}}{1+P_0^c-P_1^c} + \frac{P_1(1+P_1)^{c-1}}{1+P_1^c-P_0^c}\right]\right]. \tag{172}$$

Proof: This follows directly from Lemma 8 with (165). ■

C. Coin-Flip Defense Numerics

We ran simulations for the binary case and the arcsine distribution (without cut-off) with the same parameters as described in Section V-D. The results look quite different to those of the interleaving defense. As expected, the coin-flip defense performs better against the coin-flip attack. However, against a majority voting attack this defense fails, as no information on the coalition is gained. There is still a small advantage left against an interleaving attack. However, it retains the same scaling behaviour against the minority voting and all-1 attacks.

We note that the all-1 and coin-flip attacks numerically perform the same against this defense. We could only prove this fact analytically for the interleaving defense.

We again stress that these five attacks are by no means exhaustive.

VIII. MAJORITY VOTING DEFENSE

A. Optimal Defense

The majority voting attack outputs the symbol with the highest tally. In case of a tie, a uniform choice is made from the winners. For the binary case, this can be expressed as:

$$f_{y|\boldsymbol{m}} = \begin{cases} 1 & \text{if } m_y > \frac{1}{2}c \\ \frac{1}{2} & \text{if } m_y = \frac{1}{2}c \\ 0 & \text{if } m_y < \frac{1}{2}c. \end{cases} \tag{173}$$

TABLE IV

NUMERICAL TRENDS FOR THE PERFORMANCE INDICATOR $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ OF THE MAJORITY VOTING DEFENSE IN THE BINARY CASE $q = 2$ FOR LARGE c

interleaving attack	0.91
all-1 attack	$0.66c^{0.22}$
coin-flip attack	$0.66c^{0.22}$
majority voting attack	$0.77c^{0.25}$
minority voting attack	$0.90c^{0.23}$

Lemma 11: Let $q = 2$. For the majority voting attack, we find

$$f_{y|\boldsymbol{p}}|\boldsymbol{p}|^c = \sum_{m_y=(c+1)/2}^{c} \binom{c}{m_y} p_y^{m_y} p_{1-y}^{c-m_y} \tag{174}$$

if c is odd and

$$f_{y|\boldsymbol{p}}|\boldsymbol{p}|^c = \frac{1}{2}\binom{c}{c/2}(p_y p_{1-y})^{c/2} + \sum_{m_y=(c+2)/2}^{c} \binom{c}{m_y} p_y^{m_y} p_{1-y}^{c-m_y} \tag{175}$$

if c is even.

Proof: If c is odd, then

$$\begin{aligned} f_{y|\boldsymbol{p}} &= \mathbb{E}_{M|\boldsymbol{p}}\left[f_{y|\boldsymbol{M}}\right] \\ &= \frac{1}{|\boldsymbol{p}|^c} \sum_{m_y=\lfloor c/2 \rfloor+1}^{c} \binom{c}{m_y} p_y^{m_y} p_y^{c-m_y}. \end{aligned} \tag{176}$$

If instead c is even, the expression receives an additional term $\frac{1}{2}\binom{c}{c/2}(p_y p_{1-y})^{c/2}$. ■

B. Majority Voting Defense Numerics

We ran simulations for the binary case and the arcsine distribution (without cut-off) with the same parameters as described in Section V-D. At first glance, the results look even more promising than those from the interleaving defense. Except against the interleaving attack, the performance of the majority voting defense grows as $c^{0.25}$ against the other 4 considered attacks. However, capacity is not achieved against the interleaving attack.

We note again that the all-1 and coin-flip attacks numerically perform the same against this defense. However, we were unable to show this analytically as we could for the interleaving defense.

We again stress that these five attacks are by no means exhaustive.

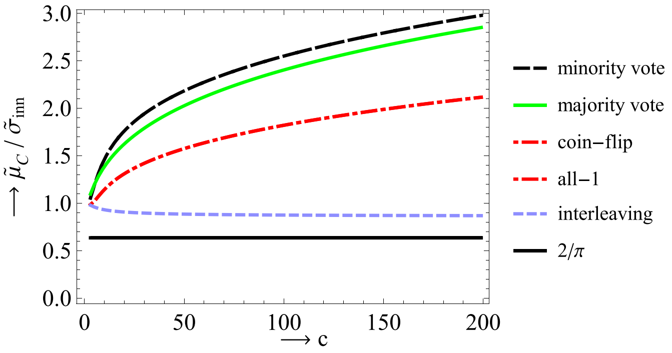

Fig. 5. Majority voting defense against various attacks in the binary case.

IX. Minority Voting Defense

A. Optimal Defense

The minority voting attack outputs the symbol with the lowest *nonzero* tally. In case of a tie, a uniform choice is made from the winners. For the binary case, this can be expressed as:

$$f_{y|\boldsymbol{m}} = \begin{cases} 1 & \text{if } 0 < m_y < \frac{1}{2}c \text{ or } m_y = c \\ \frac{1}{2} & \text{if } m_y = \frac{1}{2}c \\ 0 & \text{if } m_y = 0 \text{ or } \frac{1}{2}c < m_y < c. \end{cases} \tag{177}$$

Lemma 12: Let $q = 2$. For the minority voting attack, we find

$$f_{y|\boldsymbol{p}} |\boldsymbol{p}|^c = p_y^c + \sum_{m_y=1}^{(c-1)/2} \binom{c}{m_y} p_y^{m_y} p_{1-y}^{c-m_y} \tag{178}$$

if c is odd and

$$f_{y|\boldsymbol{p}} |\boldsymbol{p}|^c = \frac{1}{2}\binom{c}{c/2}(p_y p_{1-y})^{c/2} + p_y^c + \sum_{m_y=1}^{(c-2)/2} \binom{c}{m_y} p_y^{m_y} p_{1-y}^{c-m_y} \tag{179}$$

if c is even.

Proof: If c is odd, then

$$f_{y|\boldsymbol{p}} = \mathbb{E}_{M|\boldsymbol{p}}\left[f_{y|\boldsymbol{M}}\right] = \frac{1}{|\boldsymbol{p}|^c}\left(p_y^c + \sum_{m_y=1}^{\lceil c/2\rceil - 1} \binom{c}{m_y} p_y^{m_y} p_y^{c-m_y}\right). \tag{180}$$

If instead c is even, the expression receives an additional term $\frac{1}{2}\binom{c}{c/2}(p_y p_{1-y})^{c/2}$. ■

B. Minority Voting Defense Numerics

We ran simulations for the binary case and the arcsine distribution (without cut-off) with the same parameters as described in Section V-D. The performance against the (targeted) minority voting attack is excellent, and in fact the best when one considers each attack against the matching defense. However, against the other four considered attacks the performance is poor: the minority voting defense fails against the interleaving and majority voting attacks, and only attains a small advantage against the all-1 and coin-flip attacks.

TABLE V

Numerical Trends for the Performance Indicator $\tilde{\mu}_C/\tilde{\sigma}_{\text{inn}}$ of the Minority Voting Defense in the Binary Case $q = 2$ for Large c

interleaving attack	-0.08
all-1 attack	$3.2c^{-0.51}$
coin-flip attack	$3.2c^{-0.51}$
majority voting attack	$-1.9c^{-0.52}$
minority voting attack	$1.4c^{0.25}$

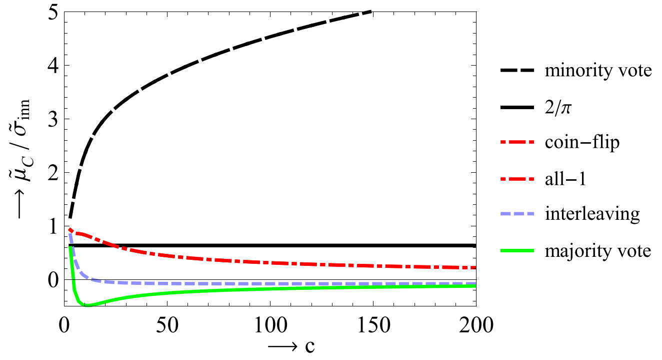

Fig. 6. Minority voting defense against various attacks in the binary case.

We again see that the all-1 and coin-flip attacks numerically perform the same against this defense. However, we were unable to prove this as we could for the interleaving defense.

We again stress that these five attacks are by no means exhaustive.

X. Tardos Suspicion Function

We end by analyzing the performance of the traditional symmetric Tardos suspicion function.

Lemma 13: If the tracer uses the symmetric Tardos suspicion function, then

$$\tilde{\mu}_C = c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_Y\left(\sqrt{\frac{1-P_Y}{P_Y}} - \sqrt{\frac{P_Y}{1-P_Y}}\right)T(Y,Y,\boldsymbol{P}) - \sqrt{\frac{P_Y}{1-P_Y}}\right]. \tag{181}$$

Proof: See (3). Since, for fixed y, $h(x, y, \boldsymbol{p})$ is the same for all $x \neq y$, we find

$$\begin{aligned} \tilde{\mu}_C &= c \cdot \mathbb{E}[T \cdot h] && (182) \\ &= c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_Y\sqrt{\frac{1-P_Y}{P_Y}}T(Y,Y,\boldsymbol{P}) - \sqrt{\frac{P_Y}{1-P_Y}}\sum_{x\neq Y} P_x T(X,Y,\boldsymbol{P})\right] && (183) \\ &= c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_Y\left(\sqrt{\frac{1-P_Y}{P_Y}} - \sqrt{\frac{P_Y}{1-P_Y}}\right)T(Y,Y,\boldsymbol{P}) - \sqrt{\frac{P_Y}{1-P_Y}}\right] && (184) \end{aligned}$$

■

Against the interleaving attack, the symmetric Tardos suspicion function does not perform well for large q:

Proposition 16: If the tracer uses the symmetric Tardos suspicion function and the coalition uses the interleaving attack, then $\tilde{\mu}_{\mathcal{C}} = \sum_{y\in\mathcal{A}} \mathbb{E}_{\boldsymbol{P}}[\sqrt{P_y(1-P_y)}]$. When $f_{\boldsymbol{P}}$ is a symmetric Dirichlet distribution with concentration parameter $\kappa = \frac{1}{q}$ and no cutoff is used,

$$\tilde{\mu}_{\mathcal{C}} = \begin{cases} \frac{2}{\pi} & \text{for } q = 2 \\ \frac{1}{2}(q-2)\tan(\frac{\pi}{q}) & \text{for } q > 2 \\ \frac{\pi}{2} & \text{as } q \to \infty \end{cases} \tag{185}$$

Proof: When $q = 2$ and p_1 follows the arcsine distribution on $[\delta, 1-\delta]$ with probability density function (2) then

$$\tilde{\mu}_{\mathcal{C}} = 2 \cdot \mathbb{E}_{\boldsymbol{P}}\sqrt{P_1(1-P_1)} = \frac{1-2\delta}{\arcsin(1-2\delta)}. \tag{186}$$

For $\delta = 0$ we find $\tilde{\mu}_{\mathcal{C}} = \frac{2}{\pi}$.

Since the marginal distribution of the symmetric Dirichlet distribution is the Beta distribution with parameters κ and $(q-1)\kappa$, we find:

$$\tilde{\mu}_{\mathcal{C}} = \sum_{y=1}^{q} \mathbb{E}_{\boldsymbol{P}}\sqrt{P_y(1-P_y)} \tag{187}$$

$$= \sum_{y=1}^{q} \frac{1}{B(\kappa, (q-1)\kappa)} \int_0^1 p_y^{\kappa+\frac{1}{2}-1}(1-p_y)^{(q-1)\kappa+\frac{1}{2}-1} \mathrm{d}p_y \tag{188}$$

$$= q\frac{B(\kappa+\frac{1}{2}, (q-1)\kappa+\frac{1}{2})}{B(\kappa, (q-1)\kappa)} \tag{189}$$

$$= q\frac{\Gamma(\kappa+\frac{1}{2})\Gamma[(q-1)\kappa+\frac{1}{2}]\Gamma(\kappa q)}{\Gamma(q\kappa+1)\Gamma(\kappa)\Gamma[(q-1)\kappa]} \tag{190}$$

$$= \frac{1}{\kappa} \cdot \frac{\Gamma(\kappa+\frac{1}{2})\Gamma[(q-1)\kappa+\frac{1}{2}]}{\Gamma(\kappa)\Gamma[(q-1)\kappa]}. \tag{191}$$

Now we set $\kappa = \frac{1}{q}$. Using Euler's reflection formula $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$, we find

$$\tilde{\mu}_{\mathcal{C}} = q \cdot \frac{\sin(\frac{\pi}{q})}{\pi} \cdot \Gamma(\tfrac{1}{q}+\tfrac{1}{2})\Gamma[1-\tfrac{1}{q}+\tfrac{1}{2}] \tag{192}$$

$$= q \cdot \frac{\sin(\frac{\pi}{q})}{\pi} \cdot (\tfrac{1}{q}-\tfrac{1}{2}) \cdot \Gamma(\tfrac{1}{q}-\tfrac{1}{2})\Gamma[1-\tfrac{1}{q}+\tfrac{1}{2}] \tag{193}$$

$$= (1-\frac{q}{2}) \cdot \frac{\sin(\frac{\pi}{q})}{\sin[(\frac{1}{q}-\frac{1}{2})\pi]} = \tfrac{1}{2}(q-2)\tan(\tfrac{\pi}{q}). \tag{194}$$

■

We see that $\tilde{\mu}_{\mathcal{C}}$ is only a slowly increasing function of q approaching the constant value $\pi/2$, which is far from the optimal code rate.

Proposition 17: If the tracer uses the symmetric Tardos suspicion function and the coalition uses the all-high attack, then

$$\tilde{\mu}_{\mathcal{C}} = c\sum_{y=0}^{q-1} \mathbb{E}_{\boldsymbol{P}}\left[P_y\left(\sqrt{\frac{1-P_y}{P_y}} - \sqrt{\frac{P_y}{1-P_y}}\right)A_{y+1}^{c-1} - \sqrt{\frac{P_y}{1-P_y}}(A_{y+1}^{c} - A_y^{c})\right]. \tag{195}$$

Proof: This follows directly from Lemma 13 with (116) and (121). ■

Proposition 18: If the tracer uses the symmetric Tardos suspicion function and the coalition uses the random-symbol attack, then

$$\tilde{\mu}_{\mathcal{C}} = c\sum_{y=0}^{q-1} \mathbb{E}_{\boldsymbol{P}}\left[P_y\left[\sqrt{\frac{1-P_y}{P_y}} - \sqrt{\frac{P_y}{1-P_y}}\right]\left[\frac{1}{q} + \sum_{\mathcal{B}\subset\mathcal{A}\colon y\in\mathcal{B}} \frac{a_{\mathcal{B}}^{c-1}}{|\mathcal{B}|(|\mathcal{B}|+1)}\right] - \sqrt{\frac{P_y}{1-P_y}}\left(\frac{1-(1-P_y)^c}{q} + \sum_{\mathcal{B}\subset\mathcal{A}\colon y\in\mathcal{B}} \frac{a_{\mathcal{B}}^{c} - a_{\mathcal{B}\setminus\{y\}}^{c}}{|\mathcal{B}|(|\mathcal{B}|+1)}\right)\right]. \tag{196}$$

Proof: This follows directly from Lemma 13 with (158) and (160). ■

It is already known that in the binary case the Tardos defense has a constant $\tilde{\mu}_{\mathcal{C}}$:

Proposition 19 [16]: Let $q = 2$ and $f_{\boldsymbol{P}}$ be the symmetric Dirichlet distribution with parameter $\kappa = \frac{1}{2}$ without cutoff. If the tracer uses the symmetric Tardos defense, then $\tilde{\mu}_{\mathcal{C}} = \frac{2}{\pi}$, no matter what attack the coalition uses.

XI. Discussion

We have investigated the optimization of the performance indicator $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ for bias-based traitor tracing in the simple-decoder setting. A straightforward Lagrangian approach yields a simple expression (Theorem 1) for the optimal suspicion function in a wide variety of contexts, e.g. CDM and RDM, binary and q-ary. The result is a Neyman-Pearson score for the hypothesis $j \in \mathcal{C}$ based on single-location information. It also has the form of a Fisher score, though without a fully understood interpretation.

The h function we obtain with the Lagrangian method depends either on the collusion strategy or on the coalition's symbol tallies $\boldsymbol{m}$. These quantities are usually unknown to the tracer. Our optimization approach does not allow for deriving suspicion functions that are based purely on data known to the tracer.

In Section III-A we speculated on the use of the $\boldsymbol{m}$-dependent suspicion function in the EM algorithm or as a consistency check for candidate coalitions. Further exploration is left for future work.

For several binary and q-ary attacks in the RDM we have derived the optimal suspicion function. We have investigated the performance indicator $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ in many combinations of suspicion function and attack strategy. In some cases analytic results are obtained. Notably, the matching case of the q-ary interleaving attack gives $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}} = \tilde{\mu}_{\mathcal{C}} = \sqrt{q-1}$, asymptotically ($c \to \infty$) yielding a code rate precisely equal to the channel capacity [22].

For $q = 2$ the numerical results for the performance indicator $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ are summarized in Table VI. We observe

TABLE VI
NUMERICAL TRENDS FOR THE PERFORMANCE INDICATOR $\tilde{\mu}_C/\tilde{\sigma}_{\text{inn}}$ IN THE BINARY CASE $q = 2$ FOR LARGE c

	interleaving attack	all-1 attack	coin-flip attack	majority voting attack	minority voting attack
Tardos defense	$2/\pi$	$2/\pi$	$2/\pi$	$2/\pi$	$2/\pi$
interleaving defense	1.0	$0.61c^{0.23}$	$0.61c^{0.23}$	1.2	$0.75c^{0.25}$
all-1 defense	0.71	$0.86c^{0.25}$	$0.44c^{0.23}$	0.84	$0.54c^{0.25}$
coin-flip defense	$5.1c^{-0.71}$	$0.72c^{0.25}$	$0.72c^{0.25}$	0.0	$1.1c^{0.25}$
majority voting defense	0.91	$0.66c^{0.22}$	$0.66c^{0.22}$	$0.77c^{0.25}$	$0.90c^{0.23}$
minority voting defense	−0.08	$3.2c^{-0.51}$	$3.2c^{-0.51}$	$-1.9c^{-0.52}$	$1.4c^{0.25}$

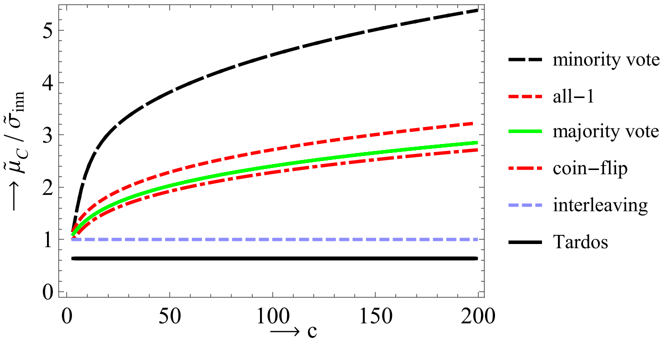

Fig. 7. Performance of optimal suspicion functions against the corresponding attack in the binary case.

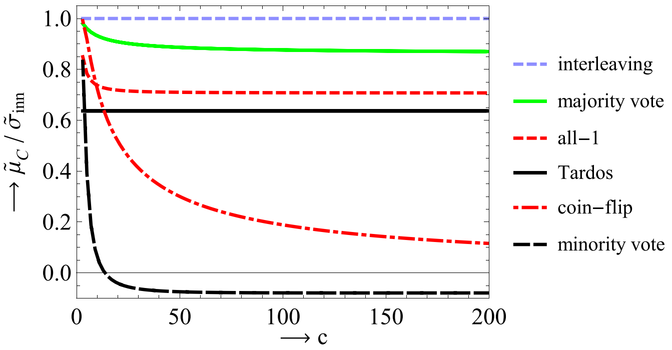

Fig. 8. Interleaving attack against various defenses in the binary case.

that the interleaving defense, all-1 defense and majority voting defense outperform the Tardos suspicion function for all the considered attacks. In many cases even a positive power of c occurs instead of a constant value: $\tilde{\mu}_C/\tilde{\sigma}_{\text{inn}} \propto c^0$ changes to $c^{1/4}$. This is a huge reduction as it leads to a codelength of $\ell \propto c^{3/2}$. Figure 7 depicts the performance of the optimal defenses against the corresponding attacks. This figure shows that the interleaving attack is particularly strong, as it is the only one with a constant value of $\tilde{\mu}_C$. The other attacks all seem to scale as $c^{1/4}$, with minority voting being the attack easiest to defend against. Figure 8 shows the performance of the interleaving defense against the five considered attacks.

Another intriguing pattern from the numerical data is the similarity of the all-1 and coin-flip attacks. Except against the all-1 defense, they have the exact same numerical results. Even though for the all-1 attack against the coin-flip defense $\tilde{\sigma}_{\text{inn}} \neq 1$, the normalized $\tilde{\mu}_C/\tilde{\sigma}_{\text{inn}}$ values are again the same. We have proven this against the interleaving defense in Proposition 7. This similarity can be explained by realizing that after the collusion attack is performed, the tracer can flip all symbols in the locations where the coalition produced a 0. This transforms the coin-flip attack into the all-1 attack, with the caveat that the coalition then never can receive the $\mathbf{0}$ vector. Naturally, this does NOT apply to the all-1 defense, as this score function is not symbol-symmetric.

It is dangerous to draw general conclusions from the table, however, since not all possible attacks are listed.

Proposition 8 and Theorem 2 on the other hand represent a very important general large-c result: the point (interleaving attack, Dirichlet bias distribution with $\kappa = 1/2$) is a saddle-point of the $\tilde{\mu}_C/\tilde{\sigma}_{\text{inn}}$ minimax game when the interleaving defense is used. With the interleaving defense as the simple decoder, the attackers cannot mount a stronger attack than the interleaving attack, and even then they cannot push the rate below the capacity.

With perfect hindsight, this result should not surprise us too much. It was shown by Huang and Moulin [23] that, in the large-c limit, the joint-decoder capacity and simple-decoder capacity coincide. Thus, asymptotically, an optimal simple decoder should automatically achieve capacity.

We now might simply decide to completely switch to the interleaving defense and abandon all other simple decoders. However, the results of Sections IV–X suggest that the other defenses can be used advantageously in a practical decoder scheme at non-asymptotic c. We envisage a decoder that runs the interleaving defense and a small battery of our h functions in parallel (one for every known 'basic' strategy, e.g. the ones discussed in this paper). Whenever the colluders use one of the basic strategies, the associated h function will quickly distinguish them from the innocent users; for other strategies, the interleaving defense does the job. The challenge is to combine the different score systems into an effective decoder. Here it has to be borne in mind that both the computational load and the total false positive probability grow with the number of incorporated h functions.

Future work will focus on (a) investigating which (if any) cutoff to use in a practical traitor tracing scheme, as we expect its scaling behaviour to change; (b) more accurate estimations of σ_{inn} in a practical traitor tracing scheme; (c) efficiency of the interleaving defense at small c, i.e. the non-asymptotic regime where $\tilde{\mu}_C/\tilde{\sigma}_{\text{inn}}$ is no longer the right performance indicator. In this regime the Worst Case Attack (that minimizes $I(Y; X|\boldsymbol{p})$) should also be considered, as this is then no longer equal to the interleaving attack. (d) simulations using multiple suspicion functions in parallel; (e) iterative joint decoders employing the m-dependent suspicion functions.

Acknowledgements

The authors would like to thank Dion Boesten, Thijs Laarhoven, Antonino Simone and Benne de Weger and the anonymous referees for valuable discussions and comments.

References

[1] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 963. Berlin, Germany: Springer-Verlag, 1995, pp. 452–465. [Online]. Available: http://dx.doi.org/10.1007/3-540-44750-4_36

[2] G. Tardos, "Optimal probabilistic fingerprint codes," in *Proc. 35th Annu. ACM Symp. Theory Comput. (STOC)*, 2003, pp. 116–125.

[3] O. Blayer and T. Tassa, "Improved versions of Tardos' fingerprinting scheme," *Designs, Codes Cryptogr.*, vol. 48, no. 1, pp. 79–103, 2008.

[4] T. Furon, A. Guyader, and F. Cérou, "On the design and optimization of Tardos probabilistic fingerprinting codes," in *Information Hiding* (Lecture Notes in Computer Science), vol. 5284. Berlin, Germany: Springer-Verlag, 2008, pp. 341–356.

[5] T. Furon, L. Pérez-Freire, A. Guyader, and F. Cérou, "Estimating the minimal length of Tardos code," in *Information Hiding* (Lecture Notes in Computer Science), vol. 5806. Berlin, Germany: Springer-Verlag, 2009, pp. 176–190.

[6] T. Laarhoven and B. de Weger, "Optimal symmetric Tardos traitor tracing schemes," *Designs, Codes Cryptogr.*, vol. 71, no. 1, pp. 83–103, 2014.

[7] A. Simone and B. Škorić, "Accusation probabilities in Tardos codes: Beyond the Gaussian approximation," *Designs, Codes Cryptogr.*, vol. 63, no. 3, pp. 379–412, Jun. 2012.

[8] B. Škorić, T. U. Vladimirova, M. U. Celik, and J. C. Talstra, "Tardos fingerprinting is better than we thought," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3663–3676, Aug. 2008.

[9] Y.-W. Huang and P. Moulin, "Capacity-achieving fingerprint decoding," in *Proc. 1st IEEE Int. Workshop Inf. Forensics Secur.*, Dec. 2009, pp. 51–55.

[10] K. Nuida, "Short collusion-secure fingerprint codes against three pirates," in *Information Hiding* (Lecture Notes in Computer Science), vol. 6387. Berlin, Germany: Springer-Verlag, 2010, pp. 86–102.

[11] K. Nuida *et al.*, "An improvement of discrete Tardos fingerprinting codes," *Designs, Codes Cryptogr.*, vol. 52, no. 3, pp. 339–362, 2009.

[12] E. Amiri and G. Tardos, "High rate fingerprinting codes and the fingerprinting capacity," in *Proc. 20th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, 2009, pp. 336–345.

[13] A. Charpentier, F. Xie, C. Fontaine, and T. Furon, "Expectation maximization decoding of Tardos probabilistic fingerprinting code," *Proc. SPIE*, vol. 7254, p. 72540, Feb. 2009.

[14] P. Meerwald and T. Furon, "Toward practical joint decoding of binary Tardos fingerprinting codes," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1168–1180, Aug. 2012.

[15] A. Charpentier, C. Fontaine, T. Furon, and I. Cox, "An asymmetric fingerprinting scheme based on Tardos codes," in *Information Hiding* (Lecture Notes in Computer Science), vol. 6958. Berlin, Germany: Springer-Verlag, 2011, pp. 43–58.

[16] B. Škorić, S. Katzenbeisser, and M. U. Celik, "Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes," *Designs, Codes Cryptogr.*, vol. 46, no. 2, pp. 137–166, 2008.

[17] B. Škorić, S. Katzenbeisser, H. G. Schaathun, and M. U. Celik, "Tardos fingerprinting codes in the combined digit model," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 906–919, Sep. 2011.

[18] F. Xie, T. Furon, and C. Fontaine, "On-off keying modulation and Tardos fingerprinting," in *Proc. 10th ACM Workshop Multimedia Secur. (MM&Sec)*, 2008, pp. 101–106.

[19] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM*, vol. 55, no. 2, 2008, Art. ID 10.

[20] P. Meerwald and T. Furon, "Towards joint Tardos decoding: The 'Don Quixote' algorithm," in *Information Hiding* (Lecture Notes in Computer Science), vol. 6958. Berlin, Germany: Springer-Verlag, 2011, pp. 28–42.

[21] B. Škorić and J.-J. Oosterwijk, "Binary and q-ary Tardos codes, revisited," *Designs, Codes Cryptogr.*, vol. 74, no. 1, pp. 75–111, 2015. [Online]. Available: http://dx.doi.org/10.1007/s10623-013-9842-3

[22] D. Boesten and B. Škorić, "Asymptotic fingerprinting capacity for non-binary alphabets," in *Information Hiding* (Lecture Notes in Computer Science), vol. 6958. Berlin, Germany: Springer-Verlag, 2011, pp. 1–13.

[23] Y.-W. Huang and P. Moulin, "On fingerprinting capacity games for arbitrary alphabets and their asymptotics," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2571–2575.

[24] T. Laarhoven, J.-J. Oosterwijk, and J. Doumen, "Dynamic traitor tracing for arbitrary alphabets: Divide and conquer," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2012, pp. 240–245.

[25] J.-J. Oosterwijk, B. Škorić, and J. Doumen, "Optimal suspicion functions for Tardos traitor tracing schemes," in *Proc. 1st ACM Workshop Inf. Hiding Multimedia Secur.*, 2013, pp. 19–28. [Online]. Available: http://doi.acm.org/10.1145/2482513.2482527

Jan-Jaap Oosterwijk was born in Dordrecht, the Netherlands, in 1982. He received both the B.Sc. and M.Sc. degrees in mathematical sciences from Utrecht University, the Netherlands, in 2008 and 2009 respectively. He is currently pursuing the Ph.D. degree in collusion-resistant traitor tracing at Eindhoven University of Technology, the Netherlands.

From 2003 to 2009 he was a Teaching Assistant with the Department of Mathematics at Utrecht University, and has been a Lecturer with Boswell-B‘eta (formerly James Boswell Institute), Utrecht, since 2009. In 2012 he started an Internship at Irdeto Research, Eindhoven. His research interests include traitor tracing, game theory, and ergodic theory.

Mr. Oosterwijk, together with co-authors Thijs Laarhoven and Jeroen Doumen, received the Best Paper Award at the 2012 IEEE International Workshop on Information Forensics and Security (WIFS) for the paper "Dynamic Traitor Tracing for Arbitrary Alphabets: Divide and Conquer."

Boris Škorić received a Ph.D. in theoretical physics from the University of Amsterdam, the Netherlands, in 1999.

From 1999 to 2008 he was a Research Scientist at Philips Research in Eindhoven, working first on display physics and later on security topics. In 2008 he joined the department of Mathematics and Computer Science at Eindhoven University of Technology, the Netherlands, as Assistant Professor.

Jeroen Doumen obtained M.Sc. degrees in the fields of mathematics and physics from Leiden University in 1998 and 1999. In 2003 he received his Ph.D. in the field of cryptography from Eindhoven University of Technology, investigating areas of overlap between cryptography and coding theory.

From 2003 until 2008 he worked at the University of Twente as a Post-Doc and as an Assistant Professor, mainly on the topics of security protocols and searching in encrypted data. Since 2008 he is a senior member of Irdeto's research team, and has (co)authored over 30 papers in international journals and conferences. His current research interests include traitor tracing codes, privacy enhancing technology, white-box cryptography, software security and node-locking technology.