Optimal Suspicion Functions for Tardos Traitor Tracing Schemes

Jan-Jaap Oosterwijk
Eindhoven Univ. of Tech.
P.O. Box 513
5600 MB Eindhoven
The Netherlands
J.Oosterwijk@TUe.nl

Boris Škorić
Eindhoven Univ. of Tech.
P.O. Box 513
5600 MB Eindhoven
The Netherlands
B.Skoric@TUe.nl

Jeroen Doumen
Irdeto BV
P.O. Box 3047
2130 KA Hoofddorp
The Netherlands
JDoumen@Irdeto.com

ABSTRACT

We investigate alternative suspicion functions for Tardos traitor tracing schemes. In the simple decoder approach (computation of a score for every user independently) we derive suspicion functions that optimize a performance indicator related to the sufficient code length ℓ in the limit of large coalition size c. Our results hold for the Restricted-Digit Model as well as the Combined-Digit Model. The scores 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 scaling $\ell \propto c^2$ is replaced by a lower power of c, e.g. $c^{3/2}$. We find that the interleaving strategy is an especially powerful attack, and the suspicion function tailored against interleaving is effective against all considered attacks.

Categories and Subject Descriptors

E.4 [**Data**]: Coding and Information Theory;
G.1 [**Mathematics of Computing**]: Numerical Analysis;
G.1.6 [**Numerical Analysis**]: Optimization—*Constrained optimization, Stochastic programming*

General Terms

Design, Measurement, Security, Theory

Keywords

Traitor tracing, collusion-resistance

1. INTRODUCTION

1.1 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*, 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 [21] was followed by a flurry of activity, e.g. improved analyses [2, 6, 7, 10, 16, 20], code modifications [8, 14, 15], decoder modifications [1, 5, 12] and various generalizations [4, 17, 18, 22]. 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*, which assign a level of suspicion to single users and (ii) *joint decoders* [1, 5, 12], 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 [17] and the scheme was generalized to q-ary alphabets. However, it turns out [19] that the suspicion function yields sub-optimal fingerprinting rates for $q > 3$, i.e. rather far below the fingerprinting capacity [3, 9] and far below the dynamic code rate [11].

Alternative suspicion functions for the binary case were introduced [5], 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.

1.2 Contributions

In this paper we further study suspicion functions.

- We generalize the work of Charpentier et al. [5] to q-ary alphabets. Using functional derivation methods we obtain suspicion functions that asymptotically ($c \gg 1$) maximize the expected score for the coalition, allowing the tracer to distinguish best between them and the innocent users. 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 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 cases we have 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 smaller than with the Tardos suspicion function), and in many cases we find $\ell \propto c^{3/2}$ even for a mismatch.

In Sections 3.1 and 7 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.

This paper contains a large number of lemmas and theorems. Full proofs are given in the appendix.

2. PRELIMINARIES

2.1 General notation

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

2.2 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 $|\boldsymbol{P}_i| = 1$. In [17] a symmetric Dirichlet distribution was taken, with concentration parameter $\kappa > 0$,

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

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

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

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 $\boldsymbol{P}_i$, 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 2.3), 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 suspect.

Tardos [21] introduced a (simple decoder) score system for he RDM at $q = 2$ that was later [17] 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, $f_{\boldsymbol{m}|\boldsymbol{p}} = \binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}}$. The attack strategy may be probabilistic.

2.3 Combined-Digit Model (CDM)

The CDM [18] allows colluders to mix symbols and to introduce noise. 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 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 [18, 22].

$$\begin{tikzcd} \boldsymbol{P} \arrow[r, "\text{code generation}", "f_{\boldsymbol{M}|\boldsymbol{P}}"'] & \boldsymbol{M} \arrow[r, "\text{colluder mix}", "f_{\boldsymbol{\Psi}|\boldsymbol{M}}"'] & \boldsymbol{\Psi} \arrow[r, "\text{tracer detection}", "f_{\boldsymbol{\Phi}|\boldsymbol{\Psi}}"'] & \boldsymbol{\Phi} \end{tikzcd}$$

Figure 1: A schematic depiction of the CDM.

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 $f_{y|\boldsymbol{m}}$.

2.4 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. If the Gaussian approximation holds, then the sufficient code length is proportional to $(\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}})^{-2}c^2$ [20]. We will use the fraction $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ as a performance indicator.

3. 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

$$L(h, \lambda_1, \lambda_2) = \tilde{\mu}_{\mathcal{C}} - \lambda_1\tilde{\mu}_{\text{inn}} - \tfrac{1}{2}\lambda_2(\tilde{\sigma}^2_{\text{inn}} - 1), \tag{8}$$

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 $\boldsymbol{P}_i$.

3.1 ... 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. Full proofs are given in the appendix. 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 $\boldsymbol{\psi}$. □

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*

$$T_{\boldsymbol{\Phi}}(x, \boldsymbol{\phi}, \boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{c p_x f_{\boldsymbol{\phi}|\boldsymbol{p}}} = \frac{1}{c}\frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x}\bigg|_{|\boldsymbol{p}|=1} + 1. \tag{9}$$

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

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.

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

$$T_{\boldsymbol{\Psi}}(x, \boldsymbol{\psi}, \boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\psi}|\boldsymbol{M}}]}{c p_x f_{\boldsymbol{\psi}|\boldsymbol{p}}} = \frac{1}{c}\frac{\partial \ln f_{\boldsymbol{\psi}|\boldsymbol{p}}}{\partial p_x}\bigg|_{|\boldsymbol{p}|=1} + 1. \tag{10}$$

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

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

$$T_{\boldsymbol{M}}(x, \boldsymbol{m}, \boldsymbol{p}) := \frac{m_x}{c p_x} = \frac{1}{c}\frac{\partial \ln f_{\boldsymbol{m}|\boldsymbol{p}}}{\partial p_x}\bigg|_{|\boldsymbol{p}|=1} + 1. \tag{11}$$

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

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

$$h = (T - \mathbb{E}[T]) / \sqrt{\text{Var}[T]} \tag{12}$$

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

PROPOSITION 5. *For the function T in all three cases above it holds that*

$$T(x, \square, \boldsymbol{p}) \propto \frac{\mathbb{P}[j \in \mathcal{C}|x, \square, \boldsymbol{p}]}{\mathbb{P}[j \notin \mathcal{C}|x, \square, \boldsymbol{p}]}, \tag{13}$$

and thus T is a Neyman-Pearson score.

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 Section 5.

(iii) We can think of two ways in which the $\boldsymbol{m}$-dependent result of Lemma 4, $h(x,y,\boldsymbol{p}) = (\frac{m_x}{cp_x}-1)\sqrt{\frac{c}{q-1}}$, can be used in practice. First, it could be employed in the EM algorithm [5]. 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 $(\frac{m_x}{cp_x}-1)\sqrt{\frac{c}{q-1}}$ 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 5 is different from the Neyman-Pearson score in [12], where the whole sequence $(Y_i)_{i\in[\ell]}$ was considered.

3.2 ... in the Restricted-Digit Model

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 $Y \in \mathcal{A}$.

Corollary 6. *Let h be of the form $h_Y(x,y,\boldsymbol{p})$ and define*

$$T_Y(x,y,\boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{y|\boldsymbol{M}}]}{cp_x f_{y|\boldsymbol{p}}} = \frac{1}{c}\left.\frac{\partial \ln f_{y|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1. \quad (14)$$

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

In the RDM, Lemma 4 and Theorem 1 hold without change. Note that the Marking Assumption is not invoked to obtain Corollary 6. Hence Corollary 6 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 (14) with $q=2$ matches the expression given by Charpentier et al. [5] (which only considered the binary case).

4. MATCHES

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"). Mismatches are discussed in Section 5.

4.1 Arbitrary alphabets

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

Proposition 7. *Against the interleaving attack, the quantity T is given by $T(x,y,\boldsymbol{p}) = 1 + (1/c)(\delta_{x,y}/p_y - 1)$, and the optimal suspicion function is*

$$h(x,y,\boldsymbol{p}) = \frac{1}{\sqrt{q-1}}\left(\frac{\delta_{x,y}}{p_y} - 1\right). \quad (15)$$

In case of a match it holds that $\tilde{\mu}_{\mathcal{C}} = \sqrt{q-1}$ for any $f_{\boldsymbol{P}}$.

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.

All-high attack. The all-high attack

$$f_{y|\boldsymbol{m}} = \begin{cases} 1 & \text{if } m_y > 0 \text{ and } m_{y+1} = \cdots = m_{q-1} = 0 \\ 0 & \text{else} \end{cases} \quad (16)$$

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 preferred-sequence attack, in which the colluders have a predetermined ranking of the symbols. The results below generalize to the preferred-sequence attack. We will use the shorthand notation $a_k := (p_0 + \cdots + p_{k-1})$ and $a_{\mathcal{B}} = \sum_{\beta\in\mathcal{B}} p_\beta$.

Proposition 8. *Against the all-high attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with*

$$T(x,y,\boldsymbol{p}) = \begin{cases} (a_{y+1}^{c-1} - a_y^{c-1})/(a_{y+1}^c - a_y^c) \ \textit{if } x < y \\ a_{y+1}^{c-1}/(a_{y+1}^c - a_y^c) \ \textit{if } x = y \\ 0 \ \textit{if } x > y. \end{cases} \quad (17)$$

In case of a match, it holds that

$$\tilde{\mu}_{\mathcal{C}} = c\sqrt{-1 + \mathbb{E}_{\boldsymbol{P}}\left[\sum_{y=0}^{q-1} \frac{A_{y+1}^{2c-1} - 2A_y^c A_{y+1}^{c-1} + A_y^{2c-1}}{A_{y+1}^c - A_y^c}\right]}. \quad (18)$$

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 9. *Let $\boldsymbol{P}$ be distributed according to the symmetric Dirichlet distribution without cutoff. The joint distribution for the pair $(A_{y+1}, A_y/A_{y+1})$ is then given by*

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

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

Proposition 10. *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,*

$$\tilde{\mu}_{\mathcal{C}} = c^{1-\kappa}\frac{\kappa\Gamma(q\kappa)\zeta(1+\kappa)}{\Gamma([q-1]\kappa)}\left[1+\mathcal{O}(c^{-\min(1,\kappa)})\right], \tag{19}$$

where ζ is the Riemann zeta function.

Random-symbol attack. 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 $f_{y|\boldsymbol{m}} = (1-\delta_{m_y,0})/|\{\alpha\in\mathcal{A} : m_\alpha > 0\}|$.

Proposition 11. *For the random-symbol attack we find*

$$|\boldsymbol{p}|^c f_{y|\boldsymbol{p}} = \frac{a^c_{\mathcal{A}} - a^c_{\mathcal{A}\setminus\{y\}}}{q} + \sum_{\mathcal{B}\subsetneq\mathcal{A}:\, y\in\mathcal{B}} \frac{a^c_{\mathcal{B}} - a^c_{\mathcal{B}\setminus\{y\}}}{|\mathcal{B}|(|\mathcal{B}|+1)}. \tag{20}$$

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

$$T(x,y,\boldsymbol{p}) = \frac{1}{c}\frac{\partial \ln(|\boldsymbol{p}|^c f_{y|\boldsymbol{p}})}{\partial p_x}\bigg|_{|\boldsymbol{p}|=1} = \tag{21}$$

$$\begin{cases} \dfrac{1}{f_{y|\boldsymbol{p}}}\left(\dfrac{1}{q} + \displaystyle\sum_{\mathcal{B}\subsetneq\mathcal{A}:\, y\in\mathcal{B}} \dfrac{a^{c-1}_{\mathcal{B}}}{|\mathcal{B}|(|\mathcal{B}|+1)}\right) & \text{if } x=y \\ \dfrac{1}{f_{y|\boldsymbol{p}}}\left(\dfrac{1-(1-p_y)^{c-1}}{q} + \displaystyle\sum_{\substack{\mathcal{B}\subsetneq\mathcal{A}\\ x,y\in\mathcal{B}}} \dfrac{a^{c-1}_{\mathcal{B}} - a^{c-1}_{\mathcal{B}\setminus\{y\}}}{|\mathcal{B}|(|\mathcal{B}|+1)}\right) & \text{if } x\neq y \end{cases} \tag{22}$$

4.2 Binary alphabet ($q=2$)

All-1 attack. 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 12. *Against the all-1 attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with*

$$T(x,y,\boldsymbol{p}) = \begin{cases} (1-p_0^{c-1})/(1-p_0^c) & \text{if } (x,y)=(0,1) \\ 1/(1-p_0^c) & \text{if } (x,y)=(1,1) \\ 1/p_0 & \text{if } (x,y)=(0,0) \\ 0 & \text{if } (x,y)=(1,0). \end{cases} \tag{23}$$

In case of a match it holds that

$$\tilde{\mu}_{\mathcal{C}} = c\sqrt{\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1}(1-P_0)/(1-P_0^c)]}. \tag{24}$$

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

Corollary 13. *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.*

Coin-flip attack. 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 14. *Against the coin-flip attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with*

$$T(x,y,\boldsymbol{p}) = \begin{cases} (1+p_y^{c-1})/(1+p_y^c - p_{1-y}^c) & \text{if } x=y \\ (1-p_{1-y}^{c-1})/(1+p_y^c-p_{1-y}^c) & \text{if } x\neq y. \end{cases} \tag{25}$$

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 $1-p_{1-y}^{c-1} < 1+p_y^c-p_{1-y}^c$.

Majority-vote attack. The majority-vote attack outputs the symbol with the highest tally. In case of a tie, a uniform choice is made from the winners.

$$f_{y|\boldsymbol{m}} = \begin{cases} 1 & \text{if } m_y > \frac{1}{2}c \\ \frac{1}{2} & \text{if } m_y = \frac{1}{2}c \\ 0 & \text{if } m_y < \frac{1}{2}c. \end{cases} \tag{26}$$

Minority-vote attack. The minority-vote attack outputs the symbol with the lowest *nonzero* tally. In case of a tie, a uniform choice is made from the winners.

$$f_{y|\boldsymbol{m}} = \begin{cases} 1 & \text{if } 0 < m_y < \frac{1}{2}c \text{ or } m_y = c \\ \frac{1}{2} & \text{if } m_y = \frac{1}{2}c \\ 0 & \text{if } m_y = 0 \text{ or } \frac{1}{2}c < m_y < c. \end{cases} \tag{27}$$

We have analytical expressions for the majority-vote and minority-vote attack, but we do not write them down here because of lack of space.

5. MISMATCHES

In this section, we analyze what happens when the coalition mounts a different attack than the tracer expected. We call the "optimal suspicion function against strategy A" the A-defense. 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. The main results of this section are analytical expressions for the performance indicator in case of a mismatch for the Tardos defense, the interleaving defense, and the interleaving attack.

Recall that $\tilde{\mu}_{\mathcal{C}} = c\cdot\mathbb{E}[T\cdot h]$. This expression remains valid in the case of a mismatch, where T is for the actual attack and h is the function that is used as defense.

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 $\mathbb{E}_{X|\boldsymbol{p}}[h^2(X,y,\boldsymbol{p})] = 1$.

Lemma 15. *Each optimal suspicion function (see Theorem 1) is strongly centered. So is the symmetric Tardos function.*

5.1 Tardos Suspicion Function

We start by considering the traditional symmetric Tardos suspicion function.

Lemma 16. *If the tracer uses the symmetric Tardos suspicion function, then*

$$\tilde{\mu}_{\mathcal{C}} = c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_Y\left(\sqrt{\frac{1-P_Y}{P_Y}} - \sqrt{\frac{P_Y}{1-P_Y}}\right)T(Y,Y,\boldsymbol{P}) - \sqrt{\frac{P_Y}{1-P_Y}}\right]. \tag{28}$$

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

PROPOSITION 17. *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* $\boldsymbol{P}$ *has a symmetric Dirichlet distribution with concentration parameter* $\kappa = \frac{1}{q}$ *and no cutoff is used,*

$$\tilde{\mu}_{\mathcal{C}} = \begin{cases} \frac{2}{\pi} & \textit{for } q = 2 \\ \frac{1}{2}(q-2)\tan(\frac{\pi}{q}) & \textit{for } q > 2 \\ \frac{\pi}{2} & \textit{as } q \to \infty \end{cases} \tag{29}$$

We see that $\tilde{\mu}_{\mathcal{C}}$ is a slowly increasing function of q, which is bad for the code rate.

PROPOSITION 18. *If the tracer uses the symmetric Tardos suspicion function and the coalition uses the all-high attack, then*

$$\tilde{\mu}_{\mathcal{C}} = c\sum_{y=0}^{q-1} \mathbb{E}_{\boldsymbol{P}}\left[P_y \left(\sqrt{\frac{1-P_y}{P_y}} - \sqrt{\frac{P_y}{1-P_y}} \right) A_{y+1}^{c-1} \right. \tag{30}$$
$$\left. - \sqrt{\frac{P_y}{1-P_y}} (A_{y+1}^{c} - A_y^{c}) \right].$$

PROPOSITION 19. *If the tracer uses the symmetric Tardos suspicion function and the coalition uses the random-symbol attack, then*

$$\tilde{\mu}_{\mathcal{C}} = c\sum_{y=0}^{q-1} \mathbb{E}_{\boldsymbol{P}}\Bigg[$$
$$P_y \left[\sqrt{\frac{1-P_y}{P_y}} - \sqrt{\frac{P_y}{1-P_y}} \right] \left[\frac{1}{q} + \sum_{\mathcal{B}\subset\mathcal{A}\colon\, y\in\mathcal{B}} \frac{a_{\mathcal{B}}^{c-1}}{|\mathcal{B}|(|\mathcal{B}|+1)} \right]$$
$$- \sqrt{\frac{P_y}{1-P_y}} \left(\frac{1-(1-P_y)^c}{q} + \sum_{\mathcal{B}\subset\mathcal{A}\colon\, y\in\mathcal{B}} \frac{a_{\mathcal{B}}^{c} - a_{\mathcal{B}\setminus\{y\}}^{c}}{|\mathcal{B}|(|\mathcal{B}|+1)} \right) \Bigg]. \tag{31}$$

In the binary case the Tardos defense has a constant $\tilde{\mu}_{\mathcal{C}}$:

PROPOSITION 20. *[17] 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.*

5.2 Interleaving defense

We now turn our attention to the interleaving defense.

LEMMA 21. *If the tracer uses the interleaving defense, then, no matter what attack is used,*

$$\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}} \left(-1 + \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{Y|\boldsymbol{P}} \left[T(Y,Y,\boldsymbol{P}) \right] \right) \tag{32}$$

and

$$\tilde{\sigma}_{inn}^2 = \frac{1}{q-1} \left(-1 + \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{Y|\boldsymbol{P}} \left[\frac{1}{P_Y} \right] \right). \tag{33}$$

where T *belongs to the attack.*

We can explicitly calculate the performance against the all-high attack:

PROPOSITION 22. *If the tracer uses the interleaving defense, but the coalition uses the all-high attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \frac{c}{\sqrt{q-1}} \sum_{y=0}^{q-2} \mathbb{E}_{\boldsymbol{P}} \left[A_{y+1}^{c-1} \right], \ \textit{and} \tag{34}$$

$$\tilde{\sigma}_{\mathrm{inn}}^2 = \frac{1}{q-1} \left(-1 + \sum_{y=0}^{q-1} \mathbb{E}_{\boldsymbol{P}} \left[\frac{A_{y+1}^{c} - A_y^{c}}{P_y} \right] \right). \tag{35}$$

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

PROPOSITION 23. *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*

$$\tilde{\mu}_{\mathcal{C}} = \frac{\Gamma(q\kappa)}{\Gamma([q-1]\kappa)} \frac{c^{1-\kappa}}{\sqrt{q-1}} [1 + \mathcal{O}(1/c)]. \tag{36}$$

We now investigate the binary case $q = 2$. We can then rephrase Proposition 22 as

COROLLARY 24. *For* $q = 2$*, if the tracer uses the interleaving defense, but the coalition uses the all-1 attack, then* $\tilde{\mu}_{\mathcal{C}} = c\,\mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1}\right]$ *and* $\tilde{\sigma}_{\mathrm{inn}}^2 = -1 + \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + \frac{1-P_0^c}{P_1}\right]$*.*

In the binary case, we obtain explicit results for the coin-flip attack against the interleaving defense:

PROPOSITION 25. *For* $q = 2$*, if the tracer uses the interleaving defense, but the coalition uses the random coin-flip attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \tfrac{1}{2} c\, \mathbb{E}_{\boldsymbol{P}} \left[P_0^{c-1} + P_1^{c-1} \right] \ \textit{and} \tag{37}$$

$$\tilde{\sigma}_{\mathrm{inn}}^2 = -1 + \mathbb{E}_{\boldsymbol{P}} \left[\frac{1 + P_0^c - P_1^c}{2P_0} + \frac{1 + P_1^c - P_0^c}{2P_1} \right]. \tag{38}$$

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

PROPOSITION 26. *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 random coin-flip attack, then*

$$\tilde{\mu}_{\mathcal{C}} = c \cdot B(\kappa, \kappa + c - 1)/B(\kappa,\kappa) \quad \textit{and} \tag{39}$$

$$\tilde{\sigma}_{\mathrm{inn}}^2 = -1 + \frac{c}{1-\kappa} \frac{\Gamma(2\kappa)}{\Gamma(\kappa)} \frac{\Gamma(c+\kappa-1)}{\Gamma(c+2\kappa-1)} + \frac{1-2\kappa}{1-\kappa}. \tag{40}$$

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

5.3 Interleaving attack

Finally, we will analyze the interleaving attack.

LEMMA 27. *If the tracer uses a strongly centered score function and the coalition uses the interleaving attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \sum_{y\in\mathcal{A}} \mathbb{E}_{\boldsymbol{P}}[P_y\ h(y,y,\boldsymbol{P})]. \tag{41}$$

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

PROPOSITION 28. *If the tracer uses the all-high defense but the coalition uses the interleaving attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}} \mathbb{E}_{\boldsymbol{P}} \left[\sum_{y=1}^{q-1} \frac{P_y A_{y+1}^{c-1}}{A_{y+1}^{c} - A_y^{c}} \right] \tag{42}$$

where T *belongs to the all-high defense.*

	interleaving	all-1	coin-flip	majority vote	minority vote
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 vote defense	0.91	$0.66c^{0.22}$	$0.66c^{0.22}$	$0.77c^{0.25}$	$0.90c^{0.23}$
minority vote defense	-0.08	$3.2c^{-0.51}$	$3.2c^{-0.51}$	$-1.9c^{-0.52}$	$1.4c^{0.25}$

Table 1: Numerical trends for the performance indicator $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ in the binary case $q = 2$ for large c.

Proof.

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\left(-1+\mathbb{E}_{\boldsymbol{P}}\left[\sum_{y\in\mathcal{A}}\frac{P_y A_{y+1}^{c-1}}{A_{y+1}^{c}-A_{y}^{c}}\right]\right) \tag{43}$$

$$= \frac{1}{\sqrt{\mathrm{Var}[T]}}\,\mathbb{E}_{\boldsymbol{P}}\left[\sum_{y=1}^{q-1}\frac{P_y A_{y+1}^{c-1}}{A_{y+1}^{c}-A_{y}^{c}}\right]. \tag{44}$$

□

In the binary case this reduces to

Proposition 29. *For $q = 2$, if the tracer uses the all-1 defense, but the coalition uses the interleaving attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\mathbb{E}_{\boldsymbol{P}}\left[P_1\sum_{k=0}^{\infty}P_0^{kc}\right]. \tag{45}$$

Proof.

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\left(-1+\mathbb{E}_{\boldsymbol{P}}\left[1+\frac{P_1}{1-P_0^c}\right]\right) \tag{46}$$

□

The scaling behaviour for large c is

Lemma 30. *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*

$$\tilde{\mu}_{\mathcal{C}} = \frac{\Gamma(\kappa+1)}{B(\kappa,\kappa)\sqrt{\mathrm{Var}[T]}}\sum_{t=0}^{\infty}\frac{\Gamma(tc+\kappa)}{\Gamma(tc+2\kappa+1)}. \tag{47}$$

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

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

Lemma 31. *For $q = 2$, if the tracer uses the coin-flip defense, but the coalition uses the interleaving attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \frac{1}{\sqrt{\mathrm{Var}[T]}}\left[-1+\mathbb{E}_{\boldsymbol{P}}\left[\frac{P_0(1+P_0)^{c-1}}{1+P_0^c-P_1^c}+\frac{P_1(1+P_1)^{c-1}}{1+P_1^c-P_0^c}\right]\right]. \tag{48}$$

6. NUMERICAL RESULTS

To verify our theory and its practical applicability, we ran simulations for the binary case and the arcsine distribution (without cut-off). We chose to simulate the five described attacks (interleaving, all-1, coin-flip, majority-vote, and minority-vote) and their optimal defenses. Both the cases where the defense matches the colluder strategy and the mismatches were simulated to obtain the $\tilde{\mu}_{\mathcal{C}}$ and the $\tilde{\sigma}_{\text{inn}}$ for $1 \le c \le 200$. We then analyzed this data to obtain the leading-order term in c. The results can be found in Table 1. The matching cases for each considered attack are shown in Figure 2. The interleaving defense and attack values are depicted in Figures 3 and 4. Since for mismatches the innocent score is no longer normalised ($\tilde{\sigma}_{\text{inn}} \neq 1$), we present the numeric results for $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ to make a fair comparison.

Looking at the diagonal elements of the table above, only the interleaving attack keeps a constant $\tilde{\mu}_{\mathcal{C}}$. For the other four attacks analysed, $\tilde{\mu}_{\mathcal{C}}$ seems to grow as $c^{1/4}$. We were able to prove this only for the all-1 attack. The mismatches bring even more surprises. As expected, in some cases the defense fails completely against different attacks with $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ tending to 0 (majority-vote attack against coin-flip defense) or even negative (interleaving and majority-vote attacks against the minority vote defense). Other cases tend to a constant $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ and many cases even still grow as $c^{1/4}$.

Surprisingly, we often see $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}} > 2/\pi$, the value for the Tardos score function. There are two exceptions seen in the defenses: firstly, the minority vote defense is an exception to this, as it only seems to work well against the minority vote attack, and is of little or no use against other attacks. Secondly, the coin-flip defense also fails against the majority-vote attack, and seems to scale as approximately $c^{-3/4}$ against the interleaving attack. We do stress that these five attacks are by no means exhaustive, so we do expect more exceptions to this observation.

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 $\tilde{\mu}_{\mathcal{C}}/\tilde{\sigma}_{\text{inn}}$ values are again the same. We have proven this against the interleaving defense in Proposition 26. This similarity can be explained by realizing that after the collusion attack is performed, the tracer can flip all symbols in the positions 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 **0** vector. Naturally, this does apply to the all-1 defense, as this score function is not symbol-symmetric.

We do stress that in the case of mismatches, $\tilde{\sigma}_{\text{inn}} \neq 1$. This means that to use our optimal score functions in a practical Tardos traitor tracing system, we need to add an additional step to the accusation phase. After calculating the user scores, we will need to estimate $\tilde{\sigma}_{\text{inn}}$ and normalize the scores before we can check them against the Tardos threshold.

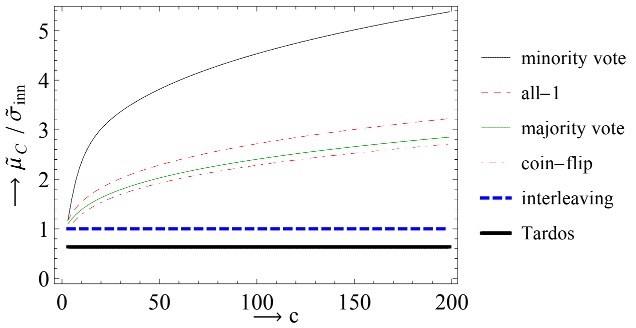

Figure 2: Matches in the binary case.

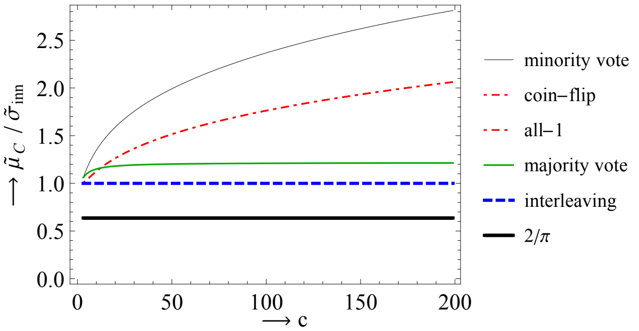

Figure 3: Optimal interleaving defense against various attacks in the binary case.

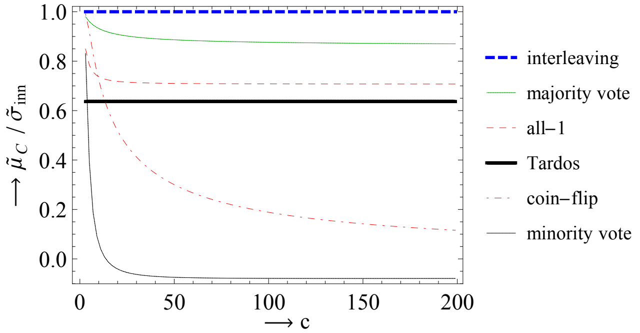

Figure 4: Interleaving attack against various defenses in the binary case.

7. 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-position 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 3.1 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 [3]. For $q = 2$ the results are summarized in Table 1. We observe 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 the power of c is changed from $\ell \propto c^2$ to $c^{3/2}$, which is a huge reduction. It is dangerous to draw general conclusions from the table, however, since not all possible attacks are listed.

The results of Sections 4–6 give us hope that the strategy-dependent suspicion functions can be used advantageously in a practical tracing scheme. We envisage a decoder that runs the Tardos function 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 Tardos function still 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) more precise performance indicators such as false positive and false negative error probability; (b) attacks targeted against the special suspicion functions derived in this paper (c) simulations using multiple suspicion functions in parallel (d) iterative joint decoders employing the m-dependent suspicion functions.

Acknowledgements

This research was supported by the Dutch Technology Foundation STW. We would like to thank Dion Boesten, Thijs Laarhoven, Antonino Simone and Benne de Weger for valuable discussions and comments.

8. REFERENCES

[1] E. Amiri and G. Tardos. High rate fingerprinting codes and the fingerprinting capacity. In *Proc. 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 336–345, 2009.

[2] O. Blayer and T. Tassa. Improved versions of Tardos' fingerprinting scheme. *Designs, Codes and Cryptography*, 48(1):79–103, 2008.

[3] D. Boesten and B. Škorić. Asymptotic fingerprinting capacity for non-binary alphabets. In *Information Hiding 2011*, volume 6958 of *LNCS*, pages 1–13. Springer, 2011.

[4] A. Charpentier, C. Fontaine, T. Furon, and I. Cox. An asymmetric fingerprinting scheme based on Tardos codes. In *Information Hiding*, volume 6958 of *LNCS*, pages 43–58. Springer, 2011.

[5] A. Charpentier, F. Xie, C. Fontaine, and T. Furon. Expectation maximization decoding of Tardos probabilistic fingerprinting code. In *Media Forensics and Security*, volume 7254 of *SPIE Proceedings*, page 72540, 2009.

[6] T. Furon, A. Guyader, and F. Cérou. On the design and optimization of Tardos probabilistic fingerprinting codes. In *Information Hiding*, volume 5284 of *Lecture Notes in Computer Science*, pages 341–356. Springer, 2008.

[7] T. Furon, L. Pérez-Freire, A. Guyader, and F. Cérou. Estimating the minimal length of Tardos code. In *Information Hiding*, volume 5806 of *LNCS*, pages 176–190, 2009.

[8] Y.-W. Huang and P. Moulin. Capacity-achieving fingerprint decoding. In *IEEE Workshop on Information Forensics and Security*, pages 51–55, 2009.

[9] Y.-W. Huang and P. Moulin. On fingerprinting capacity games for arbitrary alphabets and their asymptotics. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2571 –2575, July 2012.

[10] T. Laarhoven and B. de Weger. Optimal symmetric Tardos traitor tracing schemes. *Designs, Codes and Cryptography*, pages 1–21, 2012.

[11] T. Laarhoven, J.-J. Oosterwijk, and J. Doumen. Dynamic traitor tracing for arbitrary alphabets: Divide and conquer. In *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on*, pages 240 –245, dec. 2012.

[12] P. Meerwald and T. Furon. Towards joint Tardos decoding: the 'Don Quixote' algorithm. In *Information Hiding*, volume 6958 of *LNCS*, pages 28–42. Springer, 2011.

[13] J. Neyman and E.S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

[14] K. Nuida. Short collusion-secure fingerprint codes against three pirates. In *Information Hiding*, volume 6387 of *LNCS*, pages 86–102. Springer, 2010.

[15] K. Nuida, S. Fujitsu, M. Hagiwara, T. Kitagawa, H. Watanabe, K. Ogawa, and H. Imai. An improvement of discrete Tardos fingerprinting codes. *Des. Codes Cryptography*, 52(3):339–362, 2009.

[16] A. Simone and B. Škorić. Accusation probabilities in Tardos codes: beyond the Gaussian approximation. *Designs, Codes and Cryptography*, 63(3):379–412, 2012.

[17] B. Škorić, S. Katzenbeisser, and M.U. Celik. Symmetric Tardos Fingerprinting Codes for Arbitrary Alphabet Sizes. *Designs, Codes and Cryptography*, 46(2):137–166, 2008.

[18] B. Škorić, S. Katzenbeisser, H.G. Schaathun, and M.U. Celik. Tardos Fingerprinting Codes in the Combined Digit Model. *IEEE Transactions on Information Forensics and Security*, 6(3):906–919, 2011.

[19] B. Škorić and J.-J. Oosterwijk. Binary and q-ary Tardos codes, revisited. Cryptology ePrint Archive, Report 2012/249, 2012.

[20] B. Škorić, T.U. Vladimirova, M.U. Celik, and J.C. Talstra. Tardos Fingerprinting is Better Than We Thought. *IEEE Transactions on Information Theory*, 54(8):3663–3676, 2008.

[21] G. Tardos. Optimal Probabilistic Fingerprint Codes. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing (STOC '03)*, pages 116–125, 2003.

[22] F. Xie, T. Furon, and C. Fontaine. On-off keying modulation and Tardos fingerprinting. In *Proc. 10th Workshop on Multimedia & Security (MM&Sec)*, pages 101–106. ACM, 2008.

APPENDIX

A. PROOFS

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

$$T_{\boldsymbol{\Phi}}(x, \boldsymbol{\phi}, \boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{c p_x f_{\boldsymbol{\phi}|\boldsymbol{p}}} = \frac{1}{c} \left.\frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1. \tag{49}$$

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

PROOF OF LEMMA 2.

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}} \mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{M}} \sum\nolimits_{x \in \mathcal{A}} M_x h(x, \boldsymbol{\Phi}, \boldsymbol{P}) \tag{50}$$

$$= \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}} \mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{P}} \frac{f_{\boldsymbol{\Phi}|\boldsymbol{M}}}{f_{\boldsymbol{\Phi}|\boldsymbol{P}}} \mathbb{E}_{X|\boldsymbol{P}} \frac{M_X}{P_X} h(X, \boldsymbol{\Phi}, \boldsymbol{P}) \tag{51}$$

$$= \mathbb{E}_{\boldsymbol{P}} \mathbb{E}_{\boldsymbol{\Phi}|\boldsymbol{P}} \mathbb{E}_{X|\boldsymbol{P}} \left[\frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}[M_X f_{\boldsymbol{\Phi}|\boldsymbol{M}}]}{P_X f_{\boldsymbol{\Phi}|\boldsymbol{P}}} \, h(X, \boldsymbol{\Phi}, \boldsymbol{P}) \right] \tag{52}$$

$$= c \, \mathbb{E}[T \cdot h]. \tag{53}$$

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

$$\mathbb{E}_{X|\boldsymbol{p}}[T] = \mathbb{E}_{X|\boldsymbol{p}} \left[\frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_X f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{c p_X f_{\boldsymbol{\phi}|\boldsymbol{p}}} \right] = 1. \tag{54}$$

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

$$f_{\boldsymbol{\phi}|\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{\boldsymbol{\phi}|\boldsymbol{M}}] = \frac{1}{|\boldsymbol{p}|^c}\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}} f_{\boldsymbol{\phi}|\boldsymbol{m}}. \tag{55}$$

$$\frac{\partial \ln f_{\boldsymbol{\phi}|\boldsymbol{p}}}{\partial p_x} = \frac{\frac{1}{p_x}\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}} m_x f_{\boldsymbol{\phi}|\boldsymbol{m}}}{\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}} f_{\boldsymbol{\phi}|\boldsymbol{m}}} - \frac{c}{|\boldsymbol{p}|} \tag{56}$$

$$= \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_X f_{\boldsymbol{\phi}|\boldsymbol{M}}]}{p_X f_{\boldsymbol{\phi}|\boldsymbol{p}}} - \frac{c}{|\boldsymbol{p}|}. \tag{57}$$

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

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

$$T_{\boldsymbol{\Psi}}(x, \boldsymbol{\psi}, \boldsymbol{p}) := \frac{\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[M_x f_{\boldsymbol{\psi}|\boldsymbol{M}}]}{c p_x f_{\boldsymbol{\psi}|\boldsymbol{p}}} = \frac{1}{c}\left.\frac{\partial \ln f_{\boldsymbol{\psi}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1. \tag{58}$$

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

Proof of Lemma 3.

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{\Psi}|\boldsymbol{M}}\sum\nolimits_{x\in\mathcal{A}} M_x h(x, \boldsymbol{\Psi}, \boldsymbol{P}). \tag{59}$$

Note the similarity between equations (59) and (50). The proof proceeds analogously with $\boldsymbol{\Psi}$ instead of $\boldsymbol{\Phi}$. □

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

$$T_{\boldsymbol{M}}(x, \boldsymbol{m}, \boldsymbol{p}) := \frac{m_x}{c p_x} = \frac{1}{c}\left.\frac{\partial \ln f_{\boldsymbol{m}|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1} + 1. \tag{60}$$

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 of Lemma 4.

$$\tilde{\mu}_{\mathcal{C}} = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\sum\nolimits_{x\in\mathcal{A}} M_x h(x, \boldsymbol{M}, \boldsymbol{P}) \tag{61}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{M_X}{P_X}\, h(X, \boldsymbol{M}, \boldsymbol{P})\right] = c\,\mathbb{E}[T \cdot h]. \tag{62}$$

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

$$\mathbb{E}[T] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{M_X}{c P_X}\right] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}[1] = 1. \tag{63}$$

Also

$$\mathrm{Var}[T] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\mathbb{E}_{\boldsymbol{M}|\boldsymbol{P}}\left(\frac{M_X}{c P_X} - 1\right)^2 \tag{64}$$

$$= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\mathrm{Var}_{\boldsymbol{M}|\boldsymbol{P}}\left[\frac{M_X}{c P_X}\right] = \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{p}}\left[\frac{c p_X(1 - p_X)}{c^2 p_X^2}\right] \tag{65}$$

$$= (1/c)\mathbb{E}_{\boldsymbol{P}}\sum\nolimits_{x\in\mathcal{A}}(1 - p_x) = (q-1)/c. \tag{66}$$

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*

$$h = (T - \mathbb{E}[T])\,/\sqrt{\mathrm{Var}[T]} \tag{67}$$

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

Proof of Theorem 1. Define the Lagrangian

$$L(h, \lambda_1, \lambda_2) := c\,\mathbb{E}[Th] - \lambda_1\mathbb{E}[h] - \tfrac{1}{2}\lambda_2(\mathbb{E}[h^2] - 1). \tag{68}$$

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, $\tilde{\mu}_{\mathrm{inn}} = 0$, implies that $\lambda_1 = c\,\mathbb{E}[T]$ and the second constraint, $\tilde{\sigma}^2_{\mathrm{inn}} = 1$, that $\lambda_2^2 = \mathbb{E}(cT - \lambda_1)^2 = c^2\mathrm{Var}[T]$. □

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 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

$$L(h, \lambda_1, \lambda_2) := c\,\mathbb{E}[Th] - \lambda_1\mathbb{E}[h] - \tfrac{1}{2}\lambda_2(\mathbb{E}[h^2] - 1). \tag{69}$$

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
$h(x, \boldsymbol{m}, \boldsymbol{p})$. □

Proposition 5. *For the function T in all three cases above (Lemma 2-4) it holds that*

$$T(x, \square, \boldsymbol{p}) \propto \frac{\mathbb{P}[j \in \mathcal{C} | x, \square, \boldsymbol{p}]}{\mathbb{P}[j \notin \mathcal{C} | x, \square, \boldsymbol{p}]}, \tag{70}$$

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 $\mathbb{P}[H = \mathrm{True}|e]/\mathbb{P}[H = \mathrm{False}|e]$. 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 position. (The proof for all the other cases is analogous.) Then the Neyman-Pearson score is

$$\frac{\mathbb{P}[j\in\mathcal{C}|x\boldsymbol{\phi}\boldsymbol{p}]}{\mathbb{P}[j\notin\mathcal{C}|x\boldsymbol{\phi}\boldsymbol{p}]}=\frac{\mathbb{P}[j\in\mathcal{C},x\boldsymbol{\phi}\boldsymbol{p}]}{\mathbb{P}[j\notin\mathcal{C},x\boldsymbol{\phi}\boldsymbol{p}]}=\frac{\mathbb{P}[j\in\mathcal{C}]F(\boldsymbol{p})f_{x|\boldsymbol{p}}f_{\boldsymbol{\phi}|x\boldsymbol{p},j\in\mathcal{C}}}{\mathbb{P}[j\notin\mathcal{C}]F(\boldsymbol{p})f_{x|\boldsymbol{p}}f_{\boldsymbol{\phi}|\boldsymbol{p}}}$$

$$\propto\frac{f_{\boldsymbol{\phi}|x\boldsymbol{p},j\in\mathcal{C}}}{f_{\boldsymbol{\phi}|\boldsymbol{p}}}=\frac{1}{f_{\boldsymbol{\phi}|\boldsymbol{p}}}\sum_{\boldsymbol{m}:m_x\geq 1}\binom{c-1}{\boldsymbol{m}-\boldsymbol{e}_x}\boldsymbol{p}^{\boldsymbol{m}-\boldsymbol{e}_x}f_{\boldsymbol{\phi}|\boldsymbol{m}}$$

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

Here $\boldsymbol{e}_x$ is a length q vector containing a 1 in position x and zero elsewhere. The a priori probability $\mathbb{P}[j\in\mathcal{C}]$ is a constant. It is equal for all users if the tracer has no prior knowledge about the coalition. □

Proposition 7. *Against the interleaving attack, the quantity T is given by $T(x,y,\boldsymbol{p})=1+(1/c)(\delta_{x,y}/p_y-1)$, and the optimal suspicion function is*

$$h(x,y,\boldsymbol{p})=\frac{1}{\sqrt{q-1}}\left(\frac{\delta_{x,y}}{p_y}-1\right). \tag{71}$$

In case of a match it holds that $\tilde{\mu}_{\mathcal{C}}=\sqrt{q-1}$ for any $f_{\boldsymbol{P}}$.

Proof of Proposition 7.

$$f_{y|\boldsymbol{p}}=\frac{1}{c|\boldsymbol{p}|^c}\sum_{\boldsymbol{m}}\binom{c}{\boldsymbol{m}}\boldsymbol{p}^{\boldsymbol{m}}m_y=\frac{p_y}{c|\boldsymbol{p}|^c}\frac{\partial|\boldsymbol{p}|^c}{\partial p_y}=\frac{p_y}{|\boldsymbol{p}|}. \tag{72}$$

$$\left.\frac{\partial\ln f_{y|\boldsymbol{p}}}{\partial p_x}\right|_{|\boldsymbol{p}|=1}=\frac{\delta_{x,y}}{p_y}-1, \tag{73}$$

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

$$\begin{aligned}\mathrm{Var}[T]&=\mathbb{E}(T-1)^2 &(74)\\ &=\tfrac{1}{c^2}\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[(\delta_{x,y}/p_y-1)^2\right] &(75)\\ &=\frac{1}{c^2}\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_Y\left(\frac{1-P_Y}{P_Y}\right)^2+\sum_{x\neq y}P_x\right] &(76)\\ &=\frac{1}{c^2}\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1-P_Y}{P_Y}\right]=\frac{1}{c^2}\,\mathbb{E}_{\boldsymbol{P}}[q-1] &(77)\\ &=\frac{q-1}{c^2}. &(78)\end{aligned}$$

□

Proposition 8. *Against the all-high attack, the optimal suspicion function is $h=(T-1)/\sqrt{\mathrm{Var}[T]}$, with*

$$T(x,y,\boldsymbol{p})=\begin{cases}(a_{y+1}^{c-1}-a_y^{c-1})/(a_{y+1}^c-a_y^c) & \textit{if } x<y\\ a_{y+1}^{c-1}/(a_{y+1}^c-a_y^c) & \textit{if } x=y\\ 0 & \textit{if } x>y.\end{cases} \tag{79}$$

In case of a match, it holds that

$$\tilde{\mu}_{\mathcal{C}}=c\sqrt{-1+\mathbb{E}_{\boldsymbol{P}}\left[\sum_{y=0}^{q-1}\frac{A_{y+1}^{2c-1}-2A_y^cA_{y+1}^{c-1}+A_y^{2c-1}}{A_{y+1}^c-A_y^c}\right]}. \tag{80}$$

Proof of Proposition 8.

$$\begin{aligned}f_{y|\boldsymbol{p}}&=\mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{y|\boldsymbol{M}}] &(81)\\ &=\mathbb{P}[M_y>0,M_{y+1}=\cdots=M_{q-1}=0] &(82)\\ &=\mathbb{P}[M_{y+1}=\cdots=M_{q-1}=0] &(83)\\ &\quad-\mathbb{P}[M_y=\cdots=M_{q-1}=0]\\ &=\frac{a_{y+1}^c}{|\boldsymbol{p}|^c}-\frac{a_y^c}{|\boldsymbol{p}|^c}. &(84)\end{aligned}$$

$$T=\frac{1}{c}\left.\frac{\partial\ln(|\boldsymbol{p}|^cf_{y|\boldsymbol{p}})}{\partial p_x}\right|_{|\boldsymbol{p}|=1} \tag{85}$$

$$=\begin{cases}\dfrac{a_{y+1}^{c-1}-a_y^{c-1}}{a_{y+1}^c-a_y^c} \text{ if } x<y\\ \dfrac{a_{y+1}^{c-1}}{a_{y+1}^c-a_y^c} \text{ if } x=y\\ 0 \text{ if } x>y.\end{cases} \tag{86}$$

$$\begin{aligned}\mathbb{E}[T^2]&=\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}[T^2(X,Y,\boldsymbol{P})] &(87)\\ &=\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[P_YT^2(Y,Y,\boldsymbol{P})+A_YT^2(0,Y,\boldsymbol{P})\right] &(88)\\ &=\mathbb{E}_{\boldsymbol{P}}\sum_{y=0}^{q-1}\left[P_y\frac{A_{y+1}^{2(c-1)}}{A_{y+1}^c-A_y^c}+A_y\frac{\left(A_{y+1}^{c-1}-A_y^{c-1}\right)^2}{A_{y+1}^c-A_y^c}\right] &(89)\\ &=\mathbb{E}_{\boldsymbol{P}}\sum_{y=0}^{q-1}\frac{A_{y+1}^{2c-1}-2A_y^cA_{y+1}^{c-1}+A_y^{2c-1}}{A_{y+1}^c-A_y^c}. &(90)\end{aligned}$$

We obtain (80) using $\tilde{\mu}_{\mathcal{C}}=c\sqrt{\mathrm{Var}[T]}=c\sqrt{\mathbb{E}[T^2]-1}$. □

Lemma 9. *Let $\boldsymbol{P}$ be distributed according to the symmetric Dirichlet distribution without cutoff. The joint distribution for the pair $(A_{y+1},A_y/A_{y+1})$ is then given by*

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

Proof of Lemma 9. We first derive the joint distribution $J(a_y,a_{y+1})$ for A_y and A_{y+1}:

$$J(a_y,a_{y+1})=\mathbb{E}_{\boldsymbol{P}}\left[\delta\left[A_y-\sum_{i=0}^{y-1}P_i\right]\delta\left[A_{y+1}-\sum_{i=0}^{y}P_i\right]\right] \tag{91}$$

$$\propto\int_{|\boldsymbol{p}|=1}\mathrm{d}^{q-1}p\,\boldsymbol{p}^{\kappa-1}\delta\left[a_y-\sum_{i=0}^{y-1}p_i\right]\delta\left[a_{y+1}-\sum_{i=0}^{y}p_i\right] \tag{92}$$

$$=\int\mathrm{d}^qp\,\boldsymbol{p}^{\kappa-1}\delta\left[a_y-\sum_{i=0}^{y-1}p_i\right]\delta\left[a_{y+1}-\sum_{i=0}^{y}p_i\right]\delta(1-|\boldsymbol{p}|) \tag{93}$$

$$=\int\mathrm{d}^qp\,\boldsymbol{p}^{\kappa-1}\delta\left[a_y-\sum_{i=0}^{y-1}p_i\right]\delta\left[a_{y+1}+\sum_{i=y+1}^{q-1}p_i-1\right]\delta(1-|\boldsymbol{p}|). \tag{94}$$

Here $\delta(x)$ is the Dirac delta function. We perform the following change of variables: for $i<y$ we define $p_i=a_ys_i$;

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

$$\delta(a_y - \textstyle\sum_{i=0}^{y-1} p_i) = a_y^{-1}\delta(1 - |\boldsymbol{s}|), \tag{95}$$

$$\delta\left[a_{y+1} + \sum_{i=y+1}^{q-1} p_i - 1\right] = (1 - a_{y+1})^{-1}\delta(1 - |\boldsymbol{t}|), \tag{96}$$

$$\delta(1 - |\boldsymbol{p}|) = \delta\left[1 - p_y - a_y|\boldsymbol{s}| - (1 - a_{y+1})|\boldsymbol{t}|\ \right]. \tag{97}$$

The expression (94) becomes

$$J(a_y, a_{y+1}) = \int \mathrm{d}^y s\ \mathrm{d}p_y\ \mathrm{d}^{q-y-1}t\ p_y^{\kappa-1} a_y^{y\kappa-1} \boldsymbol{s}^{\kappa-1} \times$$
$$(1 - a_{y+1})^{[q-y-1]\kappa-1} \boldsymbol{t}^{\kappa-1} \delta(1 - |\boldsymbol{s}|)\ \delta(1 - |\boldsymbol{t}|)\ \delta\left[p_y + a_y - a_{y+1}\right] \tag{98}$$

$$\propto a_y^{y\kappa-1}(1 - a_{y+1})^{[q-y-1]\kappa-1}(a_{y+1} - a_y)^{\kappa-1}. \tag{99}$$

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 (99) becomes

$$J(a_{y+1}, z) \propto a_{y+1}^{[y+1]\kappa-1}(1 - a_{y+1})^{[q-y-1]\kappa-1} z^{y\kappa-1}(1 - z)^{\kappa-1}. \tag{100}$$

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

Proposition 10. *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,*

$$\tilde{\mu}_{\mathcal{C}} = c^{1-\kappa} \frac{\kappa\Gamma(q\kappa)\zeta(1+\kappa)}{\Gamma([q-1]\kappa)} \left[1 + \mathcal{O}(c^{-\min(1,\kappa)})\right], \tag{101}$$

where ζ is the Riemann zeta function.

Proof of Proposition 10. We write (80) as

$$\frac{\tilde{\mu}_{\mathcal{C}}^2}{c^2} = -1 + \sum_{y=0}^{q-1} \mathbb{E}_{\boldsymbol{P}}[A_{y+1}^{c-1} \frac{1 - 2(A_y/A_{y+1})^c + (A_y/A_{y+1})^{2c-1}}{1 - (A_y/A_{y+1})^c}]. \tag{102}$$

The fraction can be expanded as

$$\frac{1}{1 - (A_y/A_{y+1})^c} = \sum_{t=0}^{\infty} (A_y/A_{y+1})^{tc}. \tag{103}$$

Then we evaluate the expectation using the joint distribution $J(a_{y+1}, \frac{a_y}{a_{y+1}})$ from Lemma 9. This yields

$$\frac{\tilde{\mu}_{\mathcal{C}}^2}{c^2} = -1 + \sum_{y=0}^{q-1} \frac{B([y+1]\kappa + c - 1, [q-y-1]\kappa)}{B([y+1]\kappa, [q-y-1]\kappa)} \times$$
$$\left[1 - \sum_{t=1}^{\infty} \frac{B(y\kappa + tc, \kappa)}{B(y\kappa, \kappa)} + \sum_{t=2}^{\infty} \frac{B(y\kappa + tc - 1, \kappa)}{B(y\kappa, \kappa)}\right] \tag{104}$$

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

$$\frac{\tilde{\mu}_{\mathcal{C}}^2}{c^2} = \frac{\kappa\Gamma(q\kappa)}{\Gamma(q\kappa + c - 1)} \left[\sum_{y=0}^{q-2} \frac{\Gamma([y+1]\kappa + c - 1)}{([y+2]\kappa + c - 1)\Gamma([y+1]\kappa)}\right.$$
$$\left. + \sum_{y=1}^{q-1} \frac{\Gamma([y+1]\kappa + c - 1)}{\Gamma(y\kappa)} \sum_{t=2}^{\infty} \frac{\Gamma(y\kappa + tc - 1)}{\Gamma([y+1]\kappa + tc)}\right]. \tag{105}$$

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. □

Proposition 11. *For the random-symbol attack we find*

$$|\boldsymbol{p}|^c f_{y|\boldsymbol{p}} = \frac{a_{\mathcal{A}}^c - a_{\mathcal{A}\setminus\{y\}}^c}{q} + \sum_{\mathcal{B}\subsetneq\mathcal{A}:\ y\in\mathcal{B}} \frac{a_{\mathcal{B}}^c - a_{\mathcal{B}\setminus\{y\}}^c}{|\mathcal{B}|(|\mathcal{B}| + 1)}. \tag{106}$$

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

$$T(x, y, \boldsymbol{p}) = \frac{1}{c} \frac{\partial \ln(|\boldsymbol{p}|^c f_{y|\boldsymbol{p}})}{\partial p_x}\bigg|_{|\boldsymbol{p}|=1} = \tag{107}$$

$$\begin{cases} \dfrac{1}{f_{y|\boldsymbol{p}}} \left(\dfrac{1}{q} + \displaystyle\sum_{\mathcal{B}\subsetneq\mathcal{A}:\ y\in\mathcal{B}} \frac{a_{\mathcal{B}}^{c-1}}{|\mathcal{B}|(|\mathcal{B}| + 1)}\right) & \text{if } x = y \\ \dfrac{1}{f_{y|\boldsymbol{p}}} \left(\dfrac{1 - (1 - p_y)^{c-1}}{q} + \displaystyle\sum_{\substack{\mathcal{B}\subsetneq\mathcal{A} \\ x,y\in\mathcal{B}}} \frac{a_{\mathcal{B}}^{c-1} - a_{\mathcal{B}\setminus\{y\}}^{c-1}}{|\mathcal{B}|(|\mathcal{B}| + 1)}\right) & \text{if } x \neq y \end{cases} \tag{108}$$

Proof of Proposition 11. 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:

$$\begin{aligned} f_{y|\boldsymbol{m}} &= \tfrac{1}{q}\mathbf{1}_{\{m_y>0\}} \\ &+ \left(\tfrac{1}{q-1} - \tfrac{1}{q}\right)\mathbf{1}_{\{m_y>0\}}\mathbf{1}_{\{\exists\alpha_1:m_{\alpha_1}=0\}} \\ &+ \left(\tfrac{1}{q-2} - \tfrac{1}{q-1}\right)\mathbf{1}_{\{m_y>0\}}\mathbf{1}_{\{\exists\alpha_1:m_{\alpha_1}=0\}}\mathbf{1}_{\{\exists\alpha_2\neq\alpha_1:m_{\alpha_2}=0\}} \\ &+ \cdots + \\ &\left(1 - \tfrac{1}{2}\right)\mathbf{1}_{\{m_y>0\}}\mathbf{1}_{\{\exists\alpha_1:m_{\alpha_1}=0\}}\cdots\mathbf{1}_{\{\exists\alpha_{q-1}\neq\alpha_1,\ldots\alpha_{q-2}:m_{\alpha_{q-1}}=0\}}. \end{aligned} \tag{109}$$

Note that

$$\mathbb{P}[M_y > 0] = \mathbb{P}[M_y \geq 0] - \mathbb{P}[M_y = 0] = \frac{A_{\mathcal{A}}^c - A_{\mathcal{A}\setminus\{y\}}^c}{|\boldsymbol{p}|^c} \tag{110}$$

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

$$\mathbb{P}[\forall\beta \in \mathcal{B}, M_\beta > 0] = \left(A_{\mathcal{B}}^c - A_{\mathcal{B}\setminus\{y\}}^c\right)/|\boldsymbol{p}|^c. \tag{111}$$

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

$$\begin{aligned}|\boldsymbol{p}|^c f_{y|\boldsymbol{p}} = {} & \frac{a^c_{\mathcal{A}} - a^c_{\mathcal{A}\setminus\{y\}}}{q} \qquad (112)\\ & + \sum_{\mathcal{B}\subsetneq\mathcal{A}:\; y\in\mathcal{B}} \left(\frac{1}{|\mathcal{B}|} - \frac{1}{|\mathcal{B}|+1}\right)\left(a^c_{\mathcal{B}} - a^c_{\mathcal{B}\setminus\{y\}}\right).\end{aligned}$$

which simplifies to equation 106. □

COROLLARY 12. *Against the all-1 attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with*

$$T(x,y,\boldsymbol{p}) = \begin{cases} (1-p_0^{c-1})/(1-p_0^c) & \textit{if } (x,y) = (0,1)\\ 1/(1-p_0^c) & \textit{if } (x,y) = (1,1)\\ 1/p_0 & \textit{if } (x,y) = (0,0)\\ 0 & \textit{if } (x,y) = (1,0).\end{cases} \qquad (113)$$

In case of a match it holds that

$$\tilde{\mu}_{\mathcal{C}} = c\sqrt{\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1}(1-P_0)/(1-P_0^c)]}. \qquad (114)$$

PROOF OF COROLLARY 12. T follows directly from Proposition 8. Furthermore, equation (80) gives

$$\begin{aligned}\mathrm{Var}[T] &= -1 + \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + \frac{1-2P_0^c+P_0^{2c-1}}{1-P_0^c}\right] \qquad (115)\\ &= \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + \frac{P_0^{2c-1}-P_0^c}{1-P_0^c}\right] = \mathbb{E}_{\boldsymbol{P}}\left[\frac{P_0^c(1-P_0)}{P_0(1-P_0^c)}\right]. \qquad (116)\end{aligned}$$

□

PROPOSITION 14. *Against the coin-flip attack, the optimal suspicion function is $h = (T-1)/\sqrt{\mathrm{Var}[T]}$, with*

$$T(x,y,\boldsymbol{p}) = \begin{cases} (1+p_y^{c-1})/(1+p_y^c-p_{1-y}^c) & \textit{if } x = y\\ (1-p_{1-y}^{c-1})/(1+p_y^c-p_{1-y}^c) & \textit{if } x \neq y.\end{cases} \qquad (117)$$

PROOF OF PROPOSITION 14.

$$f_{y|\boldsymbol{m}} = \tfrac{1}{2}(1-\delta_{m_y,0}+\delta_{m_y,c}). \qquad (118)$$

$$f_{y|\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{M}|\boldsymbol{p}}[f_{y|\boldsymbol{M}}] \qquad (119)$$

$$= \frac{1}{2|\boldsymbol{p}|^c}\sum_{m_y=0}^{c}\binom{c}{m_y}p_y^{m_y}p_{1-y}^{c-m_y}(1-\delta_{m_y,0}+\delta_{m_y,c}) \qquad (120)$$

$$= \frac{1}{2|\boldsymbol{p}|^c}[(p_y+p_{1-y})^c - p_{1-y}^c + p_y^c]. \qquad (121)$$

$$\frac{\partial(|\boldsymbol{p}|^c f_{y|\boldsymbol{p}})}{\partial p_x} = \tfrac{1}{2}c[(p_y+p_{1-y})^{c-1} - (1-\delta_{x,y})p_{1-y}^{c-1} + \delta_{x,y}p_y^{c-1}] \qquad (122)$$

$$T = \frac{1-(1-\delta_{x,y})p_{1-y}^{c-1}+\delta_{x,y}p_y^{c-1}}{1-p_{1-y}^c+p_y^c}. \qquad (123)$$

□

LEMMA 15. *Each optimal suspicion function (see Theorem 1) is strongly centered. So is the symmetric Tardos function.*

PROOF OF LEMMA 15. This follows directly from equation (54). □

LEMMA 16. *If the tracer uses the symmetric Tardos suspicion function, then*

$$\begin{aligned}\tilde{\mu}_{\mathcal{C}} = c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\Bigg[&P_Y\left(\sqrt{\frac{1-P_Y}{P_Y}} - \sqrt{\frac{P_Y}{1-P_Y}}\right)T(Y,Y,\boldsymbol{P})\\ &-\sqrt{\frac{P_Y}{1-P_Y}}\Bigg]. \qquad (124)\end{aligned}$$

PROOF OF LEMMA 16. See equation 3. Since, for fixed y, $h(x,y,\boldsymbol{p})$ is the same for all $x \neq y$, we find

$$\begin{aligned}\tilde{\mu}_{\mathcal{C}} &= c\cdot\mathbb{E}[T\cdot h] \qquad (125)\\ &= c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\Bigg[P_Y\sqrt{\frac{1-P_Y}{P_Y}}T(Y,Y,\boldsymbol{P}) \qquad (126)\\ &\qquad -\sqrt{\frac{P_Y}{1-P_Y}}\sum_{x\neq Y}P_x T(X,Y,\boldsymbol{P})\Bigg]\\ &= c\,\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\Bigg[P_Y\left(\sqrt{\frac{1-P_Y}{P_Y}} - \sqrt{\frac{P_Y}{1-P_Y}}\right)T(Y,Y,\boldsymbol{P}) \qquad (127)\\ &\qquad -\sqrt{\frac{P_Y}{1-P_Y}}\Bigg]\end{aligned}$$

□

PROPOSITION 17. *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 $\boldsymbol{P}$ has a symmetric Dirichlet distribution with concentration parameter $\kappa = \frac{1}{q}$ and no cutoff is used,*

$$\tilde{\mu}_{\mathcal{C}} = \begin{cases} \frac{2}{\pi} & \textit{for } q = 2\\ \frac{1}{2}(q-2)\tan(\frac{\pi}{q}) & \textit{for } q > 2\\ \frac{\pi}{2} & \textit{as } q \to \infty\end{cases} \qquad (128)$$

PROOF OF THEOREM 17. When $q = 2$ and p_1 follows the arcsine distribution on $[\delta, 1-\delta]$ with probability density function (2) then

$$\tilde{\mu}_{\mathcal{C}} = 2\cdot\mathbb{E}_{\mathbf{p}}\sqrt{p_1(1-p_1)} = \frac{1-2\delta}{\arcsin(1-2\delta)}. \qquad (129)$$

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:

$$\tilde\mu_{\mathcal C} = \sum_{y=1}^{q} \mathbb{E}_{\mathbf P}\sqrt{p_y(1-p_y)} \tag{130}$$

$$= \sum_{y=1}^{q} \frac{1}{B(\kappa,(q-1)\kappa)}\int_0^1 p_y^{\kappa+\frac12-1}(1-p_y)^{(q-1)\kappa+\frac12-1}dp_y \tag{131}$$

$$= q\frac{B(\kappa+\frac12,(q-1)\kappa+\frac12)}{B(\kappa,(q-1)\kappa)} \tag{132}$$

$$= q\frac{\Gamma(\kappa+\frac12)\Gamma[(q-1)\kappa+\frac12]\Gamma(\kappa q)}{\Gamma(q\kappa+1)\Gamma(\kappa)\Gamma[(q-1)\kappa]} \tag{133}$$

$$= \frac{1}{\kappa}\cdot\frac{\Gamma(\kappa+\frac12)\Gamma[(q-1)\kappa+\frac12]}{\Gamma(\kappa)\Gamma[(q-1)\kappa]}. \tag{134}$$

Now we set $\kappa=\frac1q$. Using Euler's reflection formula $\Gamma(z)\Gamma(1-z)=\frac{\pi}{\sin(\pi z)}$, we find

$$\tilde\mu_{\mathcal C} = q\cdot\frac{\sin(\frac{\pi}{q})}{\pi}\cdot\Gamma(\tfrac1q+\tfrac12)\Gamma[1-\tfrac1q+\tfrac12] \tag{135}$$

$$= q\cdot\frac{\sin(\frac{\pi}{q})}{\pi}\cdot(\tfrac1q-\tfrac12)\cdot\Gamma(\tfrac1q-\tfrac12)\Gamma[1-\tfrac1q+\tfrac12] \tag{136}$$

$$= (1-\tfrac q2)\cdot\frac{\sin(\frac{\pi}{q})}{\sin[(\frac1q-\frac12)\pi]} = \tfrac12(q-2)\tan(\tfrac{\pi}{q}). \tag{137}$$

□

Lemma 21. *If the tracer uses the interleaving defense, then, no matter what attack is used,*

$$\tilde\mu_{\mathcal C} = \frac{c}{\sqrt{q-1}}\left(-1+\mathbb{E}_{\boldsymbol P}\mathbb{E}_{Y|\boldsymbol P}\left[T(Y,Y,\boldsymbol P)\right]\right) \tag{138}$$

and

$$\tilde\sigma^2_{inn} = \frac{1}{q-1}\left(-1+\mathbb{E}_{\boldsymbol P}\mathbb{E}_{Y|\boldsymbol P}\left[\frac{1}{P_Y}\right]\right). \tag{139}$$

where T belongs to the attack.

Proof of Lemma 21. Using the interleaving defense from (71), we find

$$\tilde\mu_{\mathcal C} = \mathbb{E}[T\cdot h] = \frac{c}{\sqrt{q-1}}\left(\mathbb{E}\left[T(X,Y,\boldsymbol P)\frac{\delta_{X,Y}}{P_Y}\right]-1\right) \tag{140}$$

$$h^2(x,y,\boldsymbol p) = \frac{1}{q-1}\left(\frac{\delta_{x,y}}{p_y}\left(\frac{1}{p_y}-2\right)+1\right). \tag{141}$$

$$\tilde\sigma^2_{\rm inn} = \mathbb{E}[h^2] \tag{142}$$

$$= \frac{1}{q-1}\left(1+\mathbb{E}_{\boldsymbol P}\mathbb{E}_{Y|\boldsymbol P}\left[\frac{1}{P_Y}-2\right]\right). \tag{143}$$

□

Proposition 22. *If the tracer uses the interleaving defense, but the coalition uses the all-high attack, then*

$$\tilde\mu_{\mathcal C} = \frac{c}{\sqrt{q-1}}\sum_{y=0}^{q-2}\mathbb{E}_{\boldsymbol P}\left[A_{y+1}^{c-1}\right],\ \textit{and} \tag{144}$$

$$\tilde\sigma^2_{\rm inn} = \frac{1}{q-1}\left(-1+\sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol P}\left[\frac{A_{y+1}^c-A_y^c}{P_y}\right]\right). \tag{145}$$

Proof of Proposition 22.

$$\mathbb{E}_{\boldsymbol P}\mathbb{E}_{Y|\boldsymbol P}[T(Y,Y,\boldsymbol P)] = \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol P}\left[A_{y+1}^{c-1}\right] = \sum_{y=0}^{q-2}\mathbb{E}_{\boldsymbol P}\left[A_{y+1}^{c-1}\right]+1. \tag{146}$$

$$\mathbb{E}_{\boldsymbol P}\mathbb{E}_{Y|\boldsymbol P}\left[\frac{1}{P_Y}\right] = \sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol P}\left[\frac{A_{y+1}^c-A_y^c}{P_y}\right]. \tag{147}$$

□

Proposition 23. *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*

$$\tilde\mu_{\mathcal C} = \frac{\Gamma(q\kappa)}{\Gamma([q-1]\kappa)}\frac{c^{1-\kappa}}{\sqrt{q-1}}[1+\mathcal{O}(1/c)]. \tag{148}$$

Proof of Proposition 23. Lemma 21 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 8 we get $\tilde\mu_{\mathcal C}=\frac{c}{\sqrt{q-1}}[-1+\sum_{y=0}^{q-1}\mathbb{E}_{\boldsymbol P}A_y^{c-1}]$ which can be simplified to $\tilde\mu_{\mathcal C}=\frac{c}{\sqrt{q-1}}\sum_{y=0}^{q-2}\mathbb{E}_{\boldsymbol P}A_y^{c-1}$. 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

$$\tilde\mu_{\mathcal C} \;=\; \frac{c}{\sqrt{q-1}}\sum_{y=0}^{q-2}\frac{B([q-1-y]\kappa,[y+1]\kappa+c-1)}{B([q-1-y]\kappa,[y+1]\kappa)}$$

$$= \frac{c}{\sqrt{q-1}}\sum_{b=1}^{q-1}\frac{\Gamma(q\kappa)\Gamma(c-1+b\kappa)}{\Gamma(b\kappa)\Gamma(c-1+q\kappa)}. \tag{149}$$

Next we use the property $\Gamma(x+\alpha)/\Gamma(x+\beta)=x^{\alpha-\beta}[1+\mathcal{O}(1/x)]$ which holds if $x\gg1$, $a,b\ll x$, and a, b independent of x. (See e.g. Lemma 7 in [16].) This gives

$$\tilde\mu_{\mathcal C} \;=\; \frac{c}{\sqrt{q-1}}\sum_{b=1}^{q-1}\frac{\Gamma(q\kappa)}{\Gamma(b\kappa)}c^{(b-q)\kappa}[1+\mathcal{O}(\frac1c)]. \tag{150}$$

The dominant term is $b=q-1$, yielding (148). 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 $M(a_y)=\mathbb{E}_{\boldsymbol P}\delta(a_y-\sum_{\alpha=0}^{y-1}p_\alpha)$,

$$M(a_y) = \int_0^1 \mathrm{d}^q p\;\delta(1-|\boldsymbol p|)\frac{\boldsymbol p^{\kappa-1}}{B(\kappa\mathbf{1}_q)}\delta(a_y-\sum_{\alpha=0}^{y-1}p_\alpha), \tag{151}$$

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 (151) gives

$$\begin{aligned} M(a_y) &= \frac{a_y^{y\kappa-1}(1-a_y)^{[q-y]\kappa-1}}{B(\kappa\mathbf{1}_q)}[\int_0^1 \mathrm{d}^y t \delta(1-|\boldsymbol{t}|)\boldsymbol{t}^{\kappa-1}] \\ & \cdot[\int_0^1 \mathrm{d}^{q-y} s \delta(1-|\boldsymbol{s}|)\boldsymbol{s}^{\kappa-1}] \qquad (152) \\ &= \frac{a_y^{y\kappa-1}(1-a_y)^{[q-y]\kappa-1}}{B(\kappa\mathbf{1}_q)} B(\kappa\mathbf{1}_y)B(\kappa\mathbf{1}_{q-y}). (153) \end{aligned}$$

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

Corollary 24. *For $q = 2$, if the tracer uses the interleaving defense, but the coalition uses the all-1 attack, then $\tilde{\mu}_{\mathcal{C}} = c\ \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1}\right]$ and $\tilde{\sigma}^2_{\rm inn} = -1 + \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + \frac{1-P_0^c}{P_1}\right]$.*

Proof of Corollary 24. $A_1 = P_0$ and $A_2 = P_0 + P_1 = 1$. □

Proposition 25. *For $q = 2$, if the tracer uses the interleaving defense, but the coalition uses the random coin-flip attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \tfrac{1}{2}c\ \mathbb{E}_{\boldsymbol{P}}\left[P_0^{c-1} + P_1^{c-1}\right] \textit{ and} \qquad (154)$$

$$\tilde{\sigma}^2_{\rm inn} = -1 + \mathbb{E}_{\boldsymbol{P}}\left[\frac{1+P_0^c - P_1^c}{2P_0} + \frac{1+P_1^c-P_0^c}{2P_1}\right]. \qquad (155)$$

Proof of Proposition 25.

$$\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}[T(Y,Y,\boldsymbol{P})] = \tfrac{1}{2}\textstyle\sum_{y\in\mathcal{A}} \mathbb{E}_{\boldsymbol{P}}[1+p_y^{c-1}] \qquad (156)$$

$$= 1 + \tfrac{1}{2}\mathbb{E}_{\boldsymbol{P}}[p_0^{c-1} + p_1^{c-1}]. \qquad (157)$$

$$\mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\left[\frac{1}{P_Y}\right] = \tfrac{1}{2}\sum_{y\in\mathcal{A}} \mathbb{E}_{\boldsymbol{P}}\frac{1+p_y^c - p_{1-y}^c}{P_y}. \qquad (158)$$

□

Proposition 26. *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 random coin-flip attack, then*

$$\tilde{\mu}_{\mathcal{C}} = c\cdot B(\kappa,\kappa+c-1)/B(\kappa,\kappa) \quad \textit{and} \qquad (159)$$

$$\tilde{\sigma}^2_{\rm inn} = -1 + \frac{c}{1-\kappa}\frac{\Gamma(2\kappa)}{\Gamma(\kappa)}\frac{\Gamma(c+\kappa-1)}{\Gamma(c+2\kappa-1)} + \frac{1-2\kappa}{1-\kappa}. \qquad (160)$$

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

Proof of Proposition 26. In the case of the coin-flip attack we have

$$\begin{aligned} \tilde{\mu}_{\mathcal{C}} &= \tfrac{1}{2}c\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1} + P_1^{c-1}] = c\mathbb{E}_{\boldsymbol{P}}[P_0^{c-1}] & (161) \\ &= cB(\kappa,\kappa+c-1)/B(\kappa,\kappa) & (162) \end{aligned}$$

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, equation (139) tells us that $\tilde{\sigma}^2_{\rm inn} = -1 + \mathbb{E}[1/P_Y]$. We have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{P_Y}\right] &= \sum_{y\in\{0,1\}} \mathbb{E}_{\boldsymbol{P}}\left[\frac{f_{y|\boldsymbol{P}}}{P_y}\right] & (163) \\ &= \frac{1}{2}\sum_{y\in\{0,1\}} \mathbb{E}_{\boldsymbol{P}}\left[\frac{1}{P_y} + P_y^{c-1} - \frac{(1-P_y)^c}{P_y}\right] & (164) \\ &= \mathbb{E}_{\boldsymbol{P}}\left[\frac{1}{P_y} + P_y^{c-1} - \frac{(1-P_y)^c}{P_y}\right] & (165) \\ &= \frac{B(\kappa{-}1,\kappa) + B(c{+}\kappa{-}1,\kappa) - B(\kappa{-}1,c{+}\kappa)}{B(\kappa,\kappa)}. & (166) \end{aligned}$$

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

$$\mathbb{E}\left[\frac{1}{P_Y}\right] = \frac{c}{1-\kappa}\frac{\Gamma(2\kappa)}{\Gamma(\kappa)}\frac{\Gamma(c+\kappa-1)}{\Gamma(c+2\kappa-1)} + \frac{1-2\kappa}{1-\kappa}. \qquad (167)$$

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

Lemma 27. *If the tracer uses a strongly centered score function and the coalition uses the interleaving attack, then*

$$\tilde{\mu}_{\mathcal{C}} = \sum_{y\in\mathcal{A}} \mathbb{E}_{\boldsymbol{P}}[P_y\ h(y,y,\boldsymbol{P})]. \qquad (168)$$

Proof of Lemma 27. For the interleaving attack, $cT = \frac{\delta_{x,y}}{p_y} + c - 1$, so

$$\begin{aligned} \tilde{\mu}_{\mathcal{C}} &= c\mathbb{E}[T\cdot h] & (169) \\ &= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\left(\frac{\delta_{X,Y}}{P_Y} + c - 1\right)h(X,Y,\boldsymbol{P})\right] & (170) \\ &= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}\mathbb{E}_{X|\boldsymbol{P}}\left[\frac{\delta_{X,Y}}{P_Y}\ h(X,Y,\boldsymbol{P})\right] & (171) \\ &= \mathbb{E}_{\boldsymbol{P}}\mathbb{E}_{Y|\boldsymbol{P}}[h(Y,Y,\boldsymbol{P})]. & (172) \end{aligned}$$

where (172) holds since $\mathbb{E}[h] = 0$. □