The Information-Theoretic Limits of Hiding Among Decoys in the Shuffle Model

Pengcheng Su [a], Haibo Cheng* [b] and Ping Wang* [c]

[a] *School of Computer Science, Peking University*
[b] *Peking University & Minzu University of China*
[c] *National Engineering Research Center for Software Engineering, Peking University*
{pcs, hbcheng, pwang}@pku.edu.cn

Abstract

Hiding a genuine message among plausible decoys is a recurring principle in cryptography and privacy, underlying honeywords, symmetric encryption, and anonymous data collection. We study the information-theoretic security of mechanisms that release a uniformly shuffled sequence containing one genuine message and $n-1$ decoys over a d-symbol alphabet. We characterize how restrictions on decoy generation determine achievable secrecy.

We distinguish four classes of decoy mechanisms: i.i.d., heterogeneous independent, correlated message-oblivious, and message-dependent. For every fixed $d \geq 2$, the optimal worst-case mutual information over message priors is $\Theta_d(n^{-1})$ for both independent classes and $\Theta_d(n^{-2})$ for correlated message-oblivious mechanisms. A discrete spectral construction achieves the latter rate without knowledge of the prior. Message-dependent mechanisms achieve perfect secrecy when $n \geq d$. Moreover, even with $n = 3$, heterogeneous decoys can improve upon i.i.d. decoys by an arbitrarily large multiplicative factor. These results quantify the distinct benefits of heterogeneity, correlation, and dependence on the genuine message.

A key-symmetrization reduction shows that every perfectly correct symmetric encryption scheme with a key drawn uniformly from a key space of size n admits a message-dependent shuffle construction without increasing plaintext leakage. This connects optimal message secrecy to preimage-constrained channel design and yields an exact characterization of key leakage for all message-leakage-optimal ciphers when the number of keys does not exceed the plaintext support size.

For canonical mechanisms, whose decoys are drawn i.i.d. from a blanket distribution Q independently of the genuine message, a unified likelihood-ratio representation yields exact expressions for mutual information, total-variation information, message recovery advantage, and expected posterior variance, together with explicit finite-sample remainder bounds for the mutual-information expansion. For a fixed non-degenerate message prior P, the leading asymptotic coefficient of message mutual information is minimized by $Q^*(y) \propto \sqrt{P(y)(1-P(y))}$, which generally differs from P. For binary alphabets, quasi-uniform blankets exactly minimize message total-variation information for every $n \geq 2$, independently of the prior.

Contents

1 Introduction **3**
1.1 A Hierarchy of Decoy Mechanisms and the Canonical Shuffle-Only Model 4
1.2 Main Results 6
1.3 Technical Overview 11

1.4 Related Work . . . 12
1.5 Paper Organization . . . 14

2 Likelihood Ratios and Posteriors in the Shuffle-Only Model 14

3 Separations in the Hierarchy of Decoy Generation 15
3.1 Separation between Message-Dependent and Message-Oblivious Decoys . . . 15
3.2 Separation between Message-Oblivious Correlated and Independent Decoys . . . 16
3.3 Separation between i.i.d. and Heterogeneous Decoys . . . 18

4 Connections to Information-Theoretic Security of Symmetric Encryption 19
4.1 Reduction from General Encryption to the Shuffle-Only Model . . . 19
4.2 Optimal Message Secrecy and the Induced Key Leakage . . . 20

5 Mutual Information Leakage in the Canonical Shuffle-Only Model 23
5.1 Information Leakage of Position . . . 24
5.2 Information Leakage of Message . . . 24

6 Total-Variation Information in the Canonical Shuffle-Only Model 26
6.1 Information Leakage of Position . . . 27
6.2 Information Leakage of Message . . . 27

7 Quantitative Information Flow in the Canonical Shuffle-Only Model 29
7.1 Analysis of Message Recovery Advantage . . . 29
7.2 Analysis of Expected Posterior Variance . . . 29

8 Conclusion 31

A Background 35
A.1 Mutual-Information-Based Privacy . . . 35
A.2 Quantitative Information Flow . . . 36
A.3 Proof of Property 2.1 . . . 37
A.4 Blanket Decomposition and Reduction . . . 38

B Missing Proofs in Section 3 39
B.1 Proof of Theorem 3.1 . . . 39
B.2 Proof of Lemma 3.1 . . . 40
B.3 Proof of Theorem 3.2 . . . 40
B.4 Proof of Theorem 3.3 . . . 41
B.5 Proof of Lemma 3.2 . . . 42
B.6 Proof of Lemma 3.3 . . . 42
B.7 Proof of Theorem 3.4 . . . 43
B.8 Proof of Theorem 3.5 . . . 43
B.9 Prior-Independent Analysis of the Decoy Hierarchy . . . 45

C Missing Proofs in Section 5 48
C.1 Proof of Lemma C.1 . . . 48
C.2 Proof of Theorem 5.1 . . . 48
C.3 Proof of Theorem 5.2 . . . 49

D Missing Proofs in Section 6 **51**
D.1 Proof of Theorem 6.1 51
D.2 Proof of Theorem 6.2 51
D.3 Full Proof of Theorem 6.3 52
D.4 Proof of Lemma 6.1 53

E Missing Proofs in Section 7 **54**
E.1 Proof of Theorem 7.1 54
E.2 Expression for the Expected Posterior Variance 54
E.3 Proof of Theorem 7.2 56

F Analysis of the Canonical Shuffle-Only Model when $P \not\ll Q$ **58**

1 Introduction

Hiding a genuine secret among plausible decoys is a common approach in cryptography and privacy, aiming to provide information-theoretic security even against computationally unbounded adversaries. These mechanisms expose multiple candidate values while obscuring which one is genuine. We quantify the information these mechanisms reveal. We illustrate this principle through three examples from computer security, cryptography, and privacy, and place them within a unified framework that we call the *shuffle-only model*.

Honeyword. The honeyword mechanism was introduced by Juels and Rivest [JR13] to detect password database compromises. Instead of storing only a user's genuine password, a honeyword-based system maintains a collection of n candidate passwords, consisting of the genuine password and $n-1$ decoys, called *honeywords*. The candidates are stored without revealing which one is genuine, while a separate verification mechanism enables the system to detect authentication attempts using a honeyword.

Symmetric encryption. Consider a symmetric encryption scheme $\Pi = (\mathsf{Enc}, \mathsf{Dec})$ with message space $\mathcal{M}$ and key space $\mathcal{K}$. Given a ciphertext c, a computationally unbounded adversary can exhaustively enumerate all possible keys, obtaining a vector of candidate plaintexts $\big(\mathsf{Dec}_k(c)\big)_{k\in\mathcal{K}}$. Under perfect correctness, the candidate corresponding to the correct key is guaranteed to be the genuine plaintext. Consequently, exhaustive key search can be viewed as a message-identification problem: the adversary must distinguish the genuine plaintext from the spurious decryptions under incorrect keys [JR14, JRT16].

Single-message shuffle model. The single-message shuffle model is a prominent paradigm for privacy-preserving data analysis. In this model, n users, each holding an input X_i, independently generate messages $Y_i = g(X_i)$ and submit them to a shuffler. Here g may be either a deterministic mapping or a randomized mechanism. The shuffler applies a uniformly random permutation $\sigma : [n] \to [n]$ and releases $\boldsymbol{Z} = (Y_{\sigma(i)})_{i=1}^n$. We denote by $K = \sigma^{-1}(1)$ the position of the first user's message in the shuffled output[1].

[1] Under our convention, $\sigma(i) = j$ means that the message at output position i originates from user j.

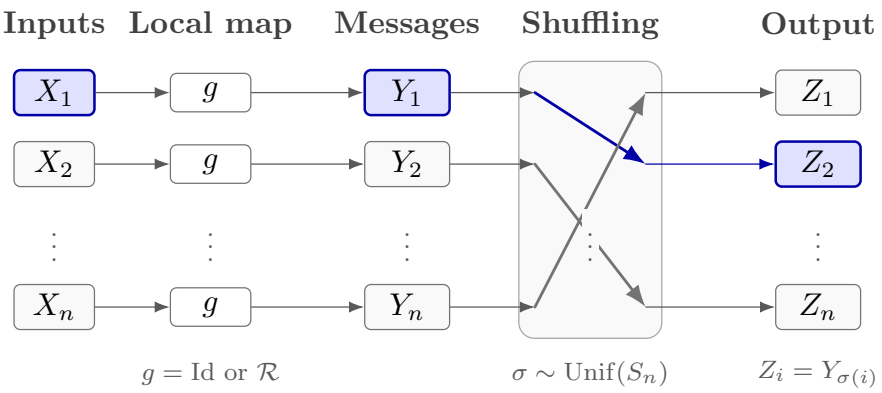


Figure 1: Illustration of the single-message shuffle model. Each user applies a local mapping $g = \mathrm{Id}$ (shuffle-only) or $g = \mathcal{R}$ (shuffle-DP) to generate a message Y_i. The shuffler applies a uniformly random permutation $\sigma \in S_n$ and releases $\boldsymbol{Z} = (Y_{\sigma(i)})_{i=1}^n$. The highlighted path illustrates the target user's message Y_1 appearing at position $K = \sigma^{-1}(1)$.

This simple architecture encompasses two important settings. When g is the identity mapping, we obtain the *shuffle-only model* [SCW26a], which captures anonymous voting and other anonymous data-collection mechanisms. When g is an ε_0-locally differentially private (LDP) randomizer, we obtain the *single-message shuffle-DP model* [BEM+17, BBGN19, CSU+19], which has attracted considerable attention for its ability to provide *privacy amplification via shuffling* [EFM+19, BBGN19, FMT22, SCW26b].

An earlier version of this work [SCW26d] analyzed both the shuffle-only and shuffle-DP models. The present paper focuses on the shuffle-only model; we plan to present the shuffle-DP analysis in a separate manuscript.

Throughout the paper, we typically denote the target message by M in decoy-generation and encryption contexts, and by Y_1 in the shuffle-only model to align with the notation for other users' messages.

1.1 A Hierarchy of Decoy Mechanisms and the Canonical Shuffle-Only Model

Let $M \sim P$ take values in $[d] = \{1, 2, \ldots, d\}$. The mechanism releases a uniformly random permutation of M together with $n-1$ decoys $\boldsymbol{H} = (H_2, \ldots, H_n)$. We consider four nested classes, distinguished by their restrictions on decoy generation.

1. **Canonical decoys.** The decoys H_i are drawn i.i.d. from Q and are jointly independent of M. We refer to Q as the *privacy blanket distribution*, following [BBGN19, SCW26a].

2. **Heterogeneous decoys.** The decoys $H_i \sim Q_i$ are mutually independent and jointly independent of M, but their distributions may differ.

3. **Message-oblivious decoys.** The decoy vector $\boldsymbol{H}$ may have arbitrary internal correlations, but must satisfy $\boldsymbol{H} \perp M$.

4. **Message-dependent decoys.** The conditional distribution $\mathrm{Law}(\boldsymbol{H} \mid M = m)$ may depend on the realized value m. This class contains the preceding three classes.

Write these classes as

$$\mathcal{C}_{\mathrm{can}} \subseteq \mathcal{C}_{\mathrm{het}} \subseteq \mathcal{C}_{\mathrm{obl}} \subseteq \mathcal{C}_{\mathrm{dep}}.$$

An equivalent histogram channel. When the permutation is uniform, these mechanisms admit a representation as constrained additive histogram noise. Let $e_m \in \mathbb{Z}^d_{\geq 0}$ be the unit count vector for message m, let $S = \sum_{i=2}^{n} e_{H_i}$ be the decoy histogram, and let $T = e_M + S$ be the output histogram. Conditional on T, every ordering of the multiset is equally likely and its distribution is independent of M. Thus $\boldsymbol{Z}$ and T are equivalent statistical experiments for inferring message content:

$$\begin{aligned} I_P(M;\boldsymbol{Z}) = I_P(M;T), \qquad T = e_M + S, \\ S \in \left\{ s \in \mathbb{Z}^d_{\geq 0} : \|s\|_1 = n-1 \right\}. \end{aligned} \tag{1}$$

The four classes correspond, respectively, to multinomial noise, Poisson multinomial noise [LWH22], arbitrary message-independent noise supported on this finite set, and input-dependent conditional noise.

Interpretation. The hierarchy directly describes alternative strategies for generating decoys in applications such as honeywords: decoys may be drawn independently from a common distribution, drawn independently from different distributions, generated jointly but independently of the genuine secret, or generated conditional on that secret.

In anonymous voting, the same hierarchy describes statistical relationships among votes. Canonical shuffling models independent votes from a common population distribution Q, which may differ from the target voter's distribution P. Heterogeneous shuffling allows voters to have different preference distributions while retaining independence. Message-oblivious shuffling allows arbitrary dependence among the other voters' votes, provided that their joint distribution remains independent of the target's vote. Message-dependent shuffling further allows the other votes to be correlated with the target's vote, without imposing a causal order.

Canonical shuffle-only model. Among the four classes, canonical shuffling is a natural and widely studied model in prior work. For honeyword generation, a common approach is to sample i.i.d. from a learned password distribution [WAMG09, DAC+15, MUS+16, WCW+18, WZD+22, CLL+22].

For symmetric encryption, the canonical decoys naturally appear in the *ideal cipher* (IC) model [BPR00, BRS02, CPS08, BS23]. In this model, the encryption algorithm $\mathsf{Enc}_k^{\mathrm{IC},\mathcal{M}}$ associated with each key $k \in \mathcal{K}$ is modeled as an independent, uniformly random permutation over the message space $\mathcal{M}$. This idealization abstracts away the internal structure of specific symmetric-key primitives[2].

Consequently, decryptions under incorrect keys behave as independent uniform samples from $\mathcal{M}$. When the genuine plaintext follows a non-uniform distribution P, an adversary can exploit the statistical discrepancy between genuine and spurious decryptions to identify the plaintext or the correct key. Such exhaustive attacks are particularly devastating when the key entropy is low, such as in password-based encryption [CBJR15, GBD16, CLW+21, CHY+25].

Honey encryption (HE) [JR14, JRT16] addresses this issue by introducing a *distribution-transforming encoder* (DTE), designed to make incorrect-key decryptions resemble plausible plaintexts rather than uniformly random messages. Specifically, a DTE consists of a randomized encoder $\mathsf{encode} : \mathcal{M} \to \mathcal{M}'$ and a deterministic decoder $\mathsf{decode} : \mathcal{M}' \to \mathcal{M}$ satisfying[3]

$$\Pr[\mathsf{decode}(\mathsf{encode}(m)) = m] = 1, \forall m \in \mathcal{M}.$$

[2] It can also be viewed as sampling uniformly from the family of all permutation-based ciphers over $\mathcal{M}$.

[3] More precisely, $\mathrm{encode}(m)$ samples uniformly from the preimage $\mathrm{decode}^{-1}(m)$. It can be shown that the security analysis of such an HE scheme is equivalent to that of the canonical shuffle-only model; see [JR14, JRT16] for details.

Honey encryption combines the DTE with encryption over the encoded message space:

$$\mathsf{Enc}_k^{\mathrm{HE}}(m) = \mathsf{Enc}_k^{\mathrm{IC},\mathcal{M}'}(\mathsf{encode}(m)), \qquad \mathsf{Dec}_k^{\mathrm{HE}}(c) = \mathsf{decode}\Big(\mathsf{Dec}_k^{\mathrm{IC},\mathcal{M}'}(c)\Big).$$

Under the ideal-cipher model, decryptions under incorrect keys follow the distribution $Q = \mathsf{decode}(U_{\mathcal{M}'})$, where $U_{\mathcal{M}'}$ denotes the uniform distribution over the encoded message space. By appropriately constructing the DTE, this distribution can approximate a desired plaintext distribution, with the approximation error decreasing as $|\mathcal{M}'|$ becomes sufficiently large.

Consequently, under uniformly distributed keys, the idealized exhaustive-key-search experiments for IC and HE correspond to instances of the canonical shuffle-only model with $n = |\mathcal{K}|$: the genuine plaintext $Y_1 \sim P$ is concealed among $n-1$ independent decoys drawn from Q. For an ideal cipher, Q is uniform over $\mathcal{M}$; for honey encryption, Q is induced by the DTE.

The central design principle of honey encryption is to make Q closely match the genuine-message distribution P, thereby rendering correct- and incorrect-key decryptions statistically indistinguishable as individual candidate messages.

From heterogeneous to canonical reductions. The significance of the canonical model is not restricted to the case of i.i.d. background messages. Privacy blanket decomposition [BBGN19, SCW26b] provides a general method for reducing heterogeneous shuffle mechanisms to canonical experiments, for the purpose of upper-bounding information leakage.

To illustrate, suppose that the background messages are mutually independent and their distributions admit a common-component decomposition

$$P_i = \gamma Q^{\mathrm{com}} + (1-\gamma)\mathrm{LO}_i, \qquad i = 2, \ldots, n,$$

where $\gamma \in [0,1]$, Q^{com} is shared by all background distributions, and LO_i denotes the corresponding residual distribution. Introducing a dummy symbol $\perp$ and defining

$$\overline{Q} = \gamma Q^{\mathrm{com}} + (1-\gamma)\delta_\perp,$$

we obtain a canonical reduction in which the target message is shuffled with $n-1$ i.i.d. samples from $\overline{Q}$. The original mechanism can be simulated by post-processing this reduced experiment, so the data-processing inequality provides upper bounds on its information leakage [BBGN19, SCW26b]. Details are provided in Appendix A.4.

We now present our main results.

1.2 Main Results

1.2.1 Separations in the Hierarchy of Decoy Mechanisms

Our first contribution is a quantitative characterization of how structural restrictions on decoy generation limit achievable secrecy. We establish separations between independent, correlated message-oblivious, and message-dependent mechanisms, both for a specified message prior and under prior-independent design. For fixed finite alphabets, allowing correlations among message-oblivious decoys improves the optimal leakage rate from $\Theta(n^{-1})$ to $\Theta(n^{-2})$. Allowing the decoys to depend on the genuine message further permits perfect secrecy once sufficiently many candidates are available. Even within the independent classes, heterogeneity can yield an arbitrarily large multiplicative improvement over i.i.d. decoys.

To formalize these comparisons, we distinguish two optimization problems:

$$\begin{aligned} L_{\mathcal{C}}(P,n) &:= \inf_{\mathsf{A}\in\mathcal{C}} I_P(M;\boldsymbol{Z}_{\mathsf{A}}), \\ \mathfrak{L}_{\mathcal{C}}(d,n) &:= \inf_{\mathsf{A}\in\mathcal{C}} \sup_{P\in\Delta_d} I_P(M;\boldsymbol{Z}_{\mathsf{A}}). \end{aligned} \tag{2}$$

The first captures optimal design for a specified message prior P, allowing the mechanism to depend on that prior. The second minimizes worst-case leakage over all priors on $[d]$: the mechanism must be fixed before the prior is chosen.

Prior-independent limits of secrecy. For every fixed $d \geq 2$, as $n \to \infty$, we prove

$$\begin{aligned} \mathfrak{L}_{\mathrm{can}}(d,n) &= \Theta_d(n^{-1}), \\ \mathfrak{L}_{\mathrm{het}}(d,n) &= \Theta_d(n^{-1}), \\ \mathfrak{L}_{\mathrm{obl}}(d,n) &= \Theta_d(n^{-2}). \end{aligned} \tag{3}$$

For message-dependent mechanisms, we obtain the exact characterization

$$\mathfrak{L}_{\mathrm{dep}}(d,n) = [\log(d/n)]_+, \qquad [x]_+ := \max\{x,0\}.$$

In particular, perfect secrecy is achievable if and only if $n \geq d$.

These guarantees preserve the order of quantifiers in (2). The discrete spectral construction attaining $O_d(n^{-2})$ leakage depends only on the public parameters d, n and satisfies the bound simultaneously for every prior $P \in \Delta_d$. Conversely, the $\Omega(n^{-1})$ lower bound for heterogeneous independent decoys holds uniformly over all choices of the distributions Q_i, including choices that vary with n. Thus, allowing different independent decoy distributions cannot improve the optimal asymptotic decay rate, whereas allowing correlations can.

The distinction arises from the scale and shape of count fluctuations. Under independent decoy generation, each symbol count has standard deviation at most $O(\sqrt{n})$. Our correlated construction instead produces smooth count noise spread over a scale of $\Theta(n)$, reducing the distinguishability of the unit shifts induced by different genuine messages. This yields $O(n^{-2})$ mutual information.

Prior-dependent limits of secrecy. For every fixed non-degenerate prior P on a finite alphabet, we establish the corresponding rate characterizations

$$\begin{aligned} L_{\mathrm{can}}(P,n) &= \Theta_P(n^{-1}), \\ L_{\mathrm{het}}(P,n) &= \Theta_P(n^{-1}), \\ L_{\mathrm{obl}}(P,n) &= \Theta_P(n^{-2}). \end{aligned}$$

Thus, knowledge of the prior can improve the constants and the choice of mechanism, but does not change these optimal decay rates.

The exact expression for $L_{\mathrm{dep}}(P,n)$ follows from the results of [KM16, KM19]; see Section 4.2. The separation between message-oblivious and message-dependent mechanisms holds at every finite candidate count: for every non-degenerate finite prior P and every $n \geq 2$,

$$L_{\mathrm{obl}}(P,n) > L_{\mathrm{dep}}(P,n).$$

An unbounded gap within the independent classes. Although canonical and heterogeneous mechanisms have the same optimal leakage order for every fixed non-degenerate prior, their optimal values need not be comparable by a universal constant. This gap already occurs with only three candidates. Specifically, let $P_p = \mathrm{Bern}(p)$ and $\ell_p = \log(1/p)$. As $p \to 0$, we prove

$$L_{\mathrm{can}}(P_p, 3) = \left(\frac{1}{2} + o(1)\right) p \log \ell_p,$$
$$L_{\mathrm{het}}(P_p, 3) = O(p \log\log \ell_p).$$

Consequently,

$$\frac{L_{\mathrm{het}}(P_p, 3)}{L_{\mathrm{can}}(P_p, 3)} \longrightarrow 0.$$

Thus, heterogeneity can provide an arbitrarily large multiplicative improvement for rare-message priors, despite leaving the optimal n^{-1} decay rate unchanged when the prior is fixed.

1.2.2 Symmetric Encryption as Shuffle-Only Mechanisms

Our second contribution establishes a fundamental connection between shuffle-only mechanisms and information-theoretically secure symmetric encryption. We show that shuffle-only constructions are sufficiently general to capture the optimal message secrecy achievable by arbitrary symmetric encryption schemes. This connection allows us to bring existing results on information-theoretically optimal channels into the study of symmetric encryption, while revealing an unavoidable relationship between message secrecy and key secrecy.

Universality of message-dependent shuffling. Consider an arbitrary symmetric encryption scheme $\Pi = (\mathsf{Enc}, \mathsf{Dec})$ with finite message space $\mathcal{M}$ and key space $\mathcal{K} = [n]$. We assume that the secret key K is uniformly distributed over $[n]$, independently of the plaintext $M \sim P$, and that the scheme satisfies perfect correctness. The encryption algorithm may be randomized.

Given a ciphertext C, a computationally unbounded adversary can enumerate all keys and construct the *decryption vector*

$$\mathsf{Dec}(C) = \big(\mathsf{Dec}_1(C), \ldots, \mathsf{Dec}_n(C)\big).$$

Perfect correctness guarantees that $\mathsf{Dec}_K(C) = M$. Since the decryption vector is a deterministic function of the ciphertext, it reveals no more information about the plaintext than C itself. Thus, without loss of generality for minimizing message leakage, we may restrict attention to encryption schemes with ciphertext space $\mathcal{C} = \mathcal{M}^{|\mathcal{K}|}$ and decryption rule $\mathsf{Dec}_k(c) = c[k]$.

The decryption vector of an arbitrary encryption scheme, however, need not be exchangeable and thus does not necessarily admit a shuffled-message representation. We overcome this difficulty through a simple *key-symmetrization reduction.*

Consider a perfectly correct encryption scheme with $\mathcal{K} = [n]$, $\mathcal{C} = \mathcal{M}^n$, and $\mathsf{Dec}_k(c) = c[k]$. We independently sample a uniformly random permutation $\sigma \in S_n$ and define

$$\widetilde{\mathsf{Enc}}_k(m) = \sigma^{-1}\big(\mathsf{Enc}_{\sigma(k)}(m)\big), \qquad \sigma \leftarrow_{\$} S_n,$$
$$\widetilde{\mathsf{Dec}}_k(c) = c[k],$$

where σ^{-1} acts by permuting ciphertext coordinates.

The transformed scheme encrypts under the original key $\sigma(k)$ and then randomly permutes the ciphertext coordinates, without revealing σ. Its ciphertext is therefore a uniformly shuffled vector

containing the genuine plaintext and $n-1$ possibly correlated, message-dependent decoys. Since the original key is uniform, the transformed ciphertext has the same distribution as an independent uniform permutation of the original ciphertext. Thus, it is a randomized post-processing of the original ciphertext and cannot increase any plaintext information leakage measure satisfying the data-processing inequality.

Consequently, optimizing information-theoretic message secrecy over perfectly correct symmetric encryption schemes is equivalent to optimizing over message-dependent shuffle-only ciphers.

Optimal message secrecy and key leakage. Our reduction connects optimal encryption secrecy to the pre-image-constrained channel-design problem studied in [KM16, KM19]. Their optimal construction simultaneously maximizes a broad class of generalized conditional entropies. For a fixed plaintext prior P with support size d and $2 \le n \le d$, let $\boldsymbol{\pi}^* = \boldsymbol{\pi}^*(P, n)$ denote their optimal posterior probability vector. By realizing their construction as a shuffle-only cipher, we obtain

$$\inf_{\Pi} I(M; C_\Pi) = H(P) - H(\boldsymbol{\pi}^*),$$

where the infimum ranges over perfectly correct symmetric encryption schemes with a key drawn uniformly from a key space of size n. We further prove that every cipher attaining this minimum satisfies

$$I(K; C_\Pi) = \log n - H(\boldsymbol{\pi}^*).$$

Optimality forces each ciphertext to decrypt to distinct plaintexts under different keys, so the posterior key and message distributions coincide up to permutation. Section 4.2 defines $\boldsymbol{\pi}^*$ and presents the shuffle-only realization and the proofs.

1.2.3 Exact Leakage and Optimal Design in the Canonical Shuffle-Only Model

Our third contribution is a systematic information-theoretic characterization of the canonical shuffle-only model. We investigate both the leakage about the target message Y_1 and that about its shuffled position K, under several complementary privacy measures.

Our analysis encompasses Shannon mutual information and total-variation information, together with operational measures such as message recovery advantage (MRA) and expected posterior variance (EPV). These metrics capture different aspects of privacy, ranging from the expected information gained by an observer to the success of an optimal message-recovery or estimation attack. Throughout the paper, all logarithms are natural, and mutual information is measured in nats.

The tractable structure of the canonical model allows us to derive exact finite-sample expressions for information leakage. These identities provide a common starting point for deriving explicit non-asymptotic bounds, asymptotic expansions, and optimal blanket distributions.

Information leakage of message. Consider the canonical model, in which $M \sim P$ is shuffled with $n-1$ i.i.d. decoys from Q. For a finite alphabet and $P \ll Q$, writing $p_y = P(y)$ and $q_y = Q(y)$, we prove the mutual information

$$I(M; \boldsymbol{Z}) = \frac{1}{2n} \sum_{y \in \mathrm{supp}(P)} \frac{p_y(1-p_y)}{q_y} + O_{P,Q}(n^{-3/2}),$$

together with an explicit finite-sample remainder bound. If P is non-degenerate, the leading coefficient is uniquely minimized, among blankets supported on $\mathrm{supp}(P)$, by

$$q_y^*(P) = \frac{\sqrt{p_y(1-p_y)}}{\sum_{x\in\mathrm{supp}(P)} \sqrt{p_x(1-p_x)}}, \qquad y \in \mathrm{supp}(P).$$

For binary messages, $Q^* = (1/2, 1/2)$, regardless of the bias of the true prior.

We further characterize the exact minimax value of the leading coefficient when the blanket cannot depend on P. Let Δ_d° denote the interior of the probability simplex. The optimization problem

$$\min_{Q\in\Delta_d^\circ} \max_{P\in\Delta_d} \frac{1}{2} \sum_{y=1}^{d} \frac{p_y(1-p_y)}{q_y}$$

has the unique solution $Q = U_d$, with optimal value $(d-1)/2$. This distinguishes two design principles: with a known prior, the optimal leading coefficient generally requires a prior-dependent blanket; with an unknown prior, a uniform blanket minimizes the worst-case leading coefficient.

Position anonymity. For fixed $P \ll Q$, we also prove

$$I(K; \boldsymbol{Z}) = D_{\mathrm{KL}}(P\|Q) - \frac{\chi^2(P\|Q)}{2n} + O_{P,Q}(n^{-3/2}),$$
$$\lim_{n\to\infty} I_{\mathrm{TV}}(K; \boldsymbol{Z}) = D_{\mathrm{TV}}(P, Q).$$

Thus, for a fixed $Q \neq P$, message mutual information decays as $1/n$, while position mutual information generally converges to a nonzero constant. When $Q = P$, the position posterior is uniform and $I(K; \boldsymbol{Z}) = 0$ for every n.

A trade-off between position and message secrecy. For a prior P that is non-uniform on its support, position secrecy and message secrecy favor different blanket distributions. Let $d = |\mathrm{supp}(P)|$. Choosing $Q = P$ achieves perfect position secrecy, $I(K; \boldsymbol{Z}) = 0$, but yields a leading message-leakage coefficient of $(d-1)/2$. In contrast, Q^* strictly improves this coefficient while incurring positive limiting position leakage: $\lim_{n\to\infty} I(K; \boldsymbol{Z}) = D_{\mathrm{KL}}(P\|Q^*) > 0$.

Other inference objectives. We also analyze secrecy under several operational criteria. For total-variation information, we obtain an exact count-based representation on any finite alphabet. For every non-degenerate binary prior and every $n \geq 2$, the quasi-uniform blanket

$$Q^* = \left(\frac{\lfloor n/2 \rfloor}{n}, \frac{\lceil n/2 \rceil}{n} \right),$$

and its coordinate permutation are exact finite-sample optimizers, independently of the prior.

For one-shot message recovery advantage, we derive an exact expectation over multinomial counts. For expected posterior variance, when messages are numerical and $P \ll Q$, we establish

$$\mathbb{E}[\mathrm{Var}(M \mid \boldsymbol{Z})] = \mathrm{Var}(M) - \frac{1}{n} \sum_{y\in\mathrm{supp}(P)} \frac{p_y^2(y-\mu)^2}{q_y} + O_{P,Q}(n^{-3/2}),$$
$$\mu = \mathbb{E}M.$$

When $\mu \notin \mathrm{supp}(P)$, the leading coefficient of the variance reduction is minimized by $q_y^* \propto p_y|y-\mu|$. If μ has positive prior probability, the boundary case can instead be stated as an infimum over blankets satisfying the support constraint.

These results show that the optimal blanket generally depends on the chosen secrecy metric.

1.3 Technical Overview

Smooth distributions under unit shifts. For message-oblivious correlated decoys, we use the histogram-noise representation in (1). In the binary case, $T = M + S$, so changing the secret shifts the same noise distribution by one unit. If r_j is the probability mass function of S, the squared Hellinger distance, under the convention without the factor $1/2$, between the two conditional output distributions is

$$\sum_{j=0}^{n} \left(\sqrt{r_j} - \sqrt{r_{j-1}}\right)^2, \qquad r_{-1} = r_n = 0.$$

Writing $v_j = \sqrt{r_j}$ expresses this energy as $v^\top L_n v$ for a unit vector v, where L_n is the discrete Laplacian with zero boundary conditions. Its minimum is the smallest eigenvalue of L_n, attained by a sine eigenvector and hence by a sine-squared probability distribution [NPR13]. A continuous analogue appears in Fisher-information-based privacy design, where bounded-support optimal densities have a squared-sine profile [Far19, FS19]. We bound mutual information through pairwise squared Hellinger distances and extend the construction to the integer simplex in fixed dimension d, obtaining an $O_d(n^{-2})$ leakage bound uniformly over all message priors.

A matching lower bound from count shifts. For any message-oblivious mechanism, fix a symbol y with $p = P(y) \in (0,1)$ and write $B = \mathbf{1}_{\{M=y\}}$. The released count of y is $T_y = B + S_y$, where S_y is independent of B. The conditional laws of T_y therefore differ by a unit shift, and their means differ by one. Since $T_y \in \{0, \dots, n\}$, the variational characterization of total variation gives

$$D_{\mathrm{TV}}\big(\mathrm{Law}(T_y \mid B = 0), \mathrm{Law}(T_y \mid B = 1)\big) \geq \frac{1}{n}.$$

Data processing reduces the analysis to this single count coordinate, and Pinsker's inequality yields

$$I(M;T) \geq I(B;T_y) \geq \frac{2p(1-p)}{n^2}.$$

Thus, no message-oblivious mechanism can improve on the n^{-2} leakage order for a fixed non-degenerate prior.

A variance barrier for independent decoys. Independence imposes a stronger restriction. Under heterogeneous independent decoy generation, each symbol count is a sum of independent Bernoulli variables and has variance at most $(n-1)/4$. We prove that an integer-valued noise variable with standard deviation σ has total variation distance $\Omega((\sigma+1)^{-1})$ from its unit shift. Applying this bound to a count coordinate, followed by data processing and Pinsker's inequality, gives an $\Omega_P(n^{-1})$ mutual-information lower bound uniformly over all heterogeneous decoy distributions.

Separating message-oblivious and message-dependent decoys. For $n < d = |\operatorname{supp}(P)|$, strict separation follows from the optimal-channel characterization of [KM16, KM19]. Its Shannon equality conditions require every positive-probability output to induce a posterior supported on exactly n messages. A shuffle output attaining this optimum must therefore contain n distinct candidates. Message-oblivious decoys cannot satisfy this condition: independence from the genuine message forces a collision with positive probability. After restricting to the finite message support, compactness and continuity ensure that the message-oblivious minimum is attained, turning this obstruction into a strict gap between the optimal values. For $n \geq d$, message-dependent mechanisms achieve perfect secrecy, whereas the count-shift lower bound keeps message-oblivious leakage strictly positive.

Separating independent decoy mechanisms. For $n = 3$ and a rare binary prior $M \sim \mathrm{Bern}(p)$, write $L = \log(1/p)$ and let T count the output ones. Under canonical decoys with parameter q, two effects constrain secrecy: the boundary event $T = 3$ reveals the rare message and contributes at least pq^2L to mutual information, while small q makes $T = 1$ strong evidence for that message. A two-regime lower bound, matched by the choice $q = L^{-1/2}$, yields optimal canonical leakage $(1/2+o(1))p \log L$. Heterogeneous decoys can control interior overlap and boundary mass separately. Choosing Bernoulli parameters $q_1 = L^{-1}$ and $q_2 = (\log L)^{-1}$ gives $\Pr(S = 1) = \Theta((\log L)^{-1})$ and $\Pr(S = 2) = 1/(L \log L)$, where S is the decoy count. A reference-distribution KL bound then yields leakage $O(p \log \log L)$. Thus, heterogeneity breaks the coupling between overlap and boundary exposure imposed by identical decoy distributions.

Symmetric encryption through the lens of the shuffle-only model. Our connection to symmetric encryption rests on the conceptual reduction outlined in Section 1.2.2, which uses post-processing arguments to preserve correctness without increasing plaintext leakage.

Likelihood-ratio representation. For canonical mechanisms, the central tool is a likelihood-ratio representation relative to the reference experiment $Q^{\otimes n}$. For $P \ll Q$, the density of the output obtained by uniformly shuffling one sample from P with $n - 1$ independent samples from Q is

$$L(\boldsymbol{z}) = \frac{1}{n} \sum_{i=1}^{n} \frac{P(z_i)}{Q(z_i)}.$$

This representation yields exact identities for the secrecy metrics we consider. On finite alphabets, a global remainder inequality for $f(t) = t \log t$, together with second- and fourth-moment calculations for independent averages, gives explicit $O_{P,Q}(n^{-3/2})$ remainder bounds for the mutual-information expansions. Cauchy–Schwarz optimizes the leading MI coefficient, while exact finite-sample optimality for binary TV-information follows from the combinatorial structure of binomial mean absolute deviation.

1.4 Related Work

Appendix A reviews mutual-information-based privacy and quantitative information flow (QIF). Here we focus on the literature most directly related to our conceptual and technical contributions.

Information hiding in the shuffle model. Shuffling has long been used to obscure the association between individuals and messages collected in batches [DSC09]. Our focus is on how restrictions on the joint distribution of these messages determine achievable information-theoretic

secrecy. In the honeyword literature, [WCW+18] analyzed the setting we call canonical shuffle-only. The distinction between identifying the genuine message's position and recovering its value, which need not coincide when candidates can repeat, is explicitly examined in [SCW26a].

To the best of our knowledge, we are the first to formulate a hierarchy of decoy-generation mechanisms and establish mutual-information separations between its classes. This hierarchy distinguishes the benefits of heterogeneity, correlation among decoys, and dependence on the genuine message. Within the canonical class, we derive exact leakage expressions and characterize optimal blanket distributions under several secrecy metrics. These results also clarify why message secrecy and position anonymity can favor different decoy distributions.

Information-theoretically secure symmetric encryption. A central question in information-theoretic encryption is what secrecy guarantees remain achievable with short keys. Shannon's perfect-secrecy bound requires the key space to be at least as large as the message space [Sha49]. Entropic security obtains secrecy with shorter keys by restricting attention to message distributions with sufficiently high min-entropy [RW06, DS05]. Honey encryption instead studies security tailored to a specified message distribution, including settings with nonuniform keys, under notions such as message recovery advantage (MRA) [JR14] and target-distribution semantic security (TDSS) [JRT16, LTZ21]. Classical constructions combine an ideal cipher with a distribution-transforming encoder, although this paradigm does not generally achieve optimal MRA security [CSL+26].

We connect this literature to the shuffle-only model. Ideal-cipher and honey-encryption experiments instantiate canonical decoy generation, while our key-symmetrization reduction transforms any perfectly correct symmetric encryption scheme with a uniform key into a message-dependent shuffle cipher without increasing plaintext leakage. This establishes an equivalence at the level of optimal information-theoretic message secrecy.

The resulting optimization is closely related to the pre-image-constrained channel-design problem studied by Khouzani and Malacaria [KM16, KM19]. Their construction simultaneously maximizes a broad class of generalized conditional entropies. We realize their optimal channel as a shuffle cipher and use the Shannon equality conditions to characterize the key leakage of every message-leakage-optimal cipher when the number of keys does not exceed the plaintext support size.

Building on the encryption–shuffle connection, our companion work [SCW26c] studies TDSS for canonical shuffle-only ciphers. It uses shuffle-only analysis to characterize the key-space complexity of TDSS and establish connections to entropic security. The present paper instead focuses on the decoy hierarchy, mutual-information separations, and exact leakage analysis.

Connections to classical statistics. When $Q = P$, the shuffled output consists of n i.i.d. samples from P, and the posterior distribution of the genuine message is their empirical distribution $\hat{P}_n$. The secrecy metrics studied here therefore reduce to familiar statistical quantities:

$$I(Y_1; \boldsymbol{Z}) = \mathbb{E}\Big[D_{\mathrm{KL}}(\hat{P}_n \| P)\Big],$$
$$I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) = \mathbb{E}\Big[D_{\mathrm{TV}}(\hat{P}_n, P)\Big],$$
$$\mathrm{Adv}_{\mathrm{MR}}(Y_1; \boldsymbol{Z}) = \mathbb{E}\Big[\max_y \hat{P}_n(y)\Big] - \max_y P(y).$$

For a fixed distribution P with finite support size $m \geq 2$, the classical asymptotic expansion

$$\mathbb{E}\Big[D_{\mathrm{KL}}(\hat{P}_n \| P)\Big] = \frac{m-1}{2n} + o(n^{-1})$$

provides the probability-matching baseline for our MI analysis [PW22]. Likewise, the expected empirical TV distance has order $\Theta_P(n^{-1/2})$; see [HJW15] for broader results on discrete distribution estimation. For uniform P, message recovery reduces to the classical balls-into-bins maximum-load problem, a connection used in [JGAP23]. When $Q = P$, for numerical messages with finite variance, the expected posterior variance is $\left(1 - \frac{1}{n}\right) \mathrm{Var}_P(Y_1)$.

Our canonical analysis extends beyond this empirical-distribution setting to arbitrary target and blanket distributions $P \ll Q$, allowing us to optimize the blanket rather than assume $Q = P$.

Changes from the previous preprint. Relative to our earlier preprint [SCW26d], this version adds a hierarchy of decoy-generation mechanisms and separation results, establishes connections to information-theoretically secure symmetric encryption, and develops a unified likelihood-ratio analysis with explicit finite-sample remainder bounds for the mutual-information expansions. The earlier preprint also analyzed the shuffle-DP model. The present paper focuses on shuffle-only mechanisms; we plan to present the shuffle-DP analysis in a separate manuscript.

1.5 Paper Organization

The remainder of this paper is organized as follows. Section 2 presents the posterior distributions and likelihood-ratio representation that underpin our analysis of the canonical shuffle-only model. Section 3 establishes separations between the levels of the decoy-generation hierarchy. Section 4 establishes connections to information-theoretically secure symmetric encryption. Sections 5, 6, and 7 analyze the canonical shuffle-only model through mutual information, TV-information, and QIF measures (MRA and EPV), respectively. Section 8 concludes the paper.

To emphasize the main ideas, the body of the paper presents proof sketches and key conceptual insights. Complete proofs and additional technical details are provided in the appendices.

2 Likelihood Ratios and Posteriors in the Shuffle-Only Model

The following property expresses the shuffled-output distribution relative to the i.i.d. reference distribution $Q^{\otimes n}$ and gives the posterior distributions of the target position K and message Y_1. These representations underpin our analysis of the canonical shuffle-only model.

Property 2.1 (Likelihood-ratio representation and posteriors)**.** *Consider the canonical shuffle-only model, where $Y_1 \sim P$ is shuffled with $n-1$ i.i.d. samples from Q, jointly independent of Y_1. Assume that P and Q are discrete distributions with $P \ll Q$, and define*

$$r(y) := \frac{P(y)}{Q(y)}, \qquad y \in \mathrm{supp}(Q).$$

Let $M_n = \mathrm{Law}(\boldsymbol{Z})$ denote the shuffled-output distribution, and let K be the position of Y_1 in $\boldsymbol{Z}$. Then

$$\frac{dM_n}{dQ^{\otimes n}}(\boldsymbol{z}) = L(\boldsymbol{z}) := \frac{1}{n}\sum_{i=1}^{n} r(z_i).$$

For every output $\boldsymbol{z} = (z_1, \ldots, z_n)$ of positive probability,

$$\Pr[K = k \mid \boldsymbol{Z} = \boldsymbol{z}] = \frac{r(z_k)}{\sum_{i=1}^{n} r(z_i)},$$

$$\Pr[Y_1 = y \mid \boldsymbol{Z} = \boldsymbol{z}] = \sum_{k=1}^{n} \Pr[K = k \mid \boldsymbol{Z} = \boldsymbol{z}]\, \mathbb{1}[z_k = y].$$

The derivation of the posterior distribution of K follows the approach in prior work [WCW+18, SCW26a]; for completeness, the detailed proof is provided in Appendix A.3.

Our analysis primarily focuses on the regime where P is absolutely continuous with respect to Q. Intuitively, if there exists a symbol y^* such that $P(y^*) > 0$ but $Q(y^*) = 0$, the occurrence of $Y_1 = y^*$ in the shuffled sequence $\boldsymbol{Z}$ would be uniquely attributable to the first user, leading to non-vanishing leakage even as $n \to \infty$. Conversely, when $P \ll Q$, the information leakage of message Y_1 typically vanishes in the large-n limit. To substantiate this intuition, we provide a formal analysis of mutual information leakage for the case where $P \not\ll Q$ in Appendix F.

Interpretation. Property 2.1 demonstrates that, upon observing the shuffled sequence $\boldsymbol{Z}$, the posterior probability that the first user's message occupies position k is proportional to the likelihood ratio $\frac{P(z_k)}{Q(z_k)}$. In particular, when $P = Q$, this ratio equals one for all k, resulting in a uniform posterior distribution that ensures perfect anonymity.

3 Separations in the Hierarchy of Decoy Generation

We now compare the optimal secrecy achievable at different levels of the decoy-generation hierarchy. We first establish a universal strict separation between message-dependent and message-oblivious mechanisms. We then show that correlations among message-oblivious decoys improve the optimal leakage rate from order n^{-1} to order n^{-2}, and finally separate i.i.d. from heterogeneous independent decoys under rare binary priors. The main text considers a fixed message distribution P; the corresponding prior-independent minimax analysis appears in Appendix B.9.

3.1 Separation between Message-Dependent and Message-Oblivious Decoys

We begin by examining whether optimal secrecy requires the decoys to depend on the realized message. Both message-dependent and message-oblivious mechanisms may be designed using the prior P; the distinction is whether the generator can use the actual value of M. The following theorem shows that this additional freedom strictly improves optimal secrecy for every non-degenerate finite prior and every $n \geq 2$.

Theorem 3.1 (Universal strict separation between oblivious and dependent decoys)**.** *Let P be a non-degenerate distribution with finite support of size $d \geq 2$, and write $v(P) := \max_{E \subseteq \mathrm{supp}(P)} P(E)(1 - P(E))$. For every $n \geq 2$,*

$$L_{\mathrm{obl}}(P, n) > L_{\mathrm{dep}}(P, n).$$

Moreover, every message-oblivious mechanism satisfies

$$I(M; \boldsymbol{Z}) \geq \frac{2v(P)}{n^2}. \tag{4}$$

If $n \geq d$, then $L_{\mathrm{dep}}(P, n) = 0$. If $2 \leq n < d$, then $L_{\mathrm{dep}}(P, n) = H(P) - H(\boldsymbol{\pi}^*)$, where $\boldsymbol{\pi}^* = \boldsymbol{\pi}^*(P, n)$ is defined in (17)*.*

The equality conditions for optimal message-dependent secrecy require collision-free outputs when $n < d$, whereas message-oblivious decoys collide with the target message with positive probability. For $n \geq d$, message-dependent decoys achieve zero leakage, while a count statistic and Pinsker's inequality give a positive lower bound for every message-oblivious mechanism. Together, these arguments establish strict separation for every $n \geq 2$.

3.2 Separation between Message-Oblivious Correlated and Independent Decoys

We next retain the requirement that the decoy vector be independent of the genuine message and examine the role of correlations among the decoys themselves. For every fixed non-degenerate finite prior, independent decoys have optimal leakage of order n^{-1}, whereas correlated decoys can attain order n^{-2}. We establish this separation through a variance-based lower bound and a construction of smooth correlated count noise. Throughout this comparison, we work with the output histogram (defined in Section 1.1) $T = e_M + S$, using the equivalence $I(M; \boldsymbol{Z}) = I(M; T)$.

3.2.1 Lower Bounds for Independent Decoys

Our lower-bound argument starts from a single symbol count. Changing whether the genuine message equals that symbol shifts the corresponding noise count by one unit. The following lemma relates the distinguishability of an integer-valued noise distribution and its unit shift to its variance. It provides the probabilistic ingredient for the leakage bounds below.

Lemma 3.1 (Total variation under a unit shift)**.** *Let S be an integer-valued random variable with finite variance $\mathrm{Var}(S) = \sigma^2$. Then*

$$D_{\mathrm{TV}}\big(\mathrm{Law}(S), \mathrm{Law}(S+1)\big) \geq \frac{3}{16(\sigma+1)}.$$

The shift bound translates into an information-leakage bound by considering the binary event $\{M = y\}$ and the output count T_y. Conditional on this event or its complement, the count distributions differ by exactly one unit. Combining the lemma with Pinsker's inequality and data processing gives the following result, which applies to arbitrary message-oblivious noise, including correlated noise.

Theorem 3.2 (A variance-dependent leakage lower bound)**.** *Let $M \sim P$ take values in a finite alphabet $\mathcal{Y}$, and let $S = (S_y)_{y\in\mathcal{Y}}$ be an integer-valued random vector independent of M. Set $T = e_M + S$, where e_m denotes the unit vector indexed by m. For any $y \in \mathcal{Y}$ with $p := P(y) \in (0,1)$, suppose that $\sigma_y^2 := \mathrm{Var}(S_y) < \infty$. Then,*

$$I(M;T) \geq \frac{9p(1-p)}{128(\sigma_y+1)^2}.$$

Independence now enters through a restriction on count variance: each decoy count is a sum of independent Bernoulli variables, so its variance is at most $(n-1)/4$. The preceding theorem therefore yields an $\Omega_P(n^{-1})$ leakage lower bound uniformly over all heterogeneous decoy distributions. Choosing every decoy distribution equal to P supplies a matching upper bound.

Theorem 3.3 (Optimal leakage rate for heterogeneous and canonical decoys)**.** *Let P be a fixed non-degenerate distribution on a finite alphabet $\mathcal{Y}$, and let $M \sim P$. For $n \geq 2$, let $H_2, \ldots, H_n$ be mutually independent decoys with respective distributions $Q_2, \ldots, Q_n$, jointly independent of M, and set*

$$T = e_M + \sum_{i=2}^{n} e_{H_i}.$$

For every $y \in \mathcal{Y}$ with $p_y := P(y) \in (0,1)$,

$$I(M;T) \geq \frac{9p_y(1-p_y)}{32(\sqrt{n-1}+2)^2}. \tag{5}$$

This bound holds uniformly over all choices of the decoy distributions, including choices that depend on n. Consequently, as $n \to \infty$,

$$L_{\mathrm{het}}(P, n) = \Theta_P(n^{-1}), \qquad L_{\mathrm{can}}(P, n) = \Theta_P(n^{-1}). \tag{6}$$

3.2.2 Upper Bounds for Correlated Decoys

To improve on the independent-decoy leakage rate, we construct correlated count noise whose distribution changes little under the shifts induced by different messages. We measure this change through distances between square-root probability vectors. The following lemma converts pairwise bounds on these distances into a mutual-information upper bound. Importantly, it remains applicable when the conditional output distributions have different supports.

Lemma 3.2 (A pairwise square-root distance bound on mutual information)**.** *Let M take values in $[d]$ with prior probabilities $p_1, \ldots, p_d > 0$. Let $W_i = \mathrm{Law}(T \mid M = i)$ be probability mass functions on a common discrete output space, and write $W = \sum_{i=1}^{d} p_i W_i$. Then,*

$$I(M; T) \le 2 \sum_{1 \le i < j \le d} \max\{p_i, p_j\} \left\| \sqrt{W_i} - \sqrt{W_j} \right\|_2^2, \tag{7}$$

where

$$\left\| \sqrt{A} - \sqrt{B} \right\|_2^2 := \sum_t \left(\sqrt{A(t)} - \sqrt{B(t)} \right)^2.$$

It remains to construct a valid histogram distribution with small square-root distances under the relevant shifts. We use sine-squared count distributions, whose square-root vectors vary smoothly under a unit shift. A final balancing coordinate enforces the fixed total mass of $n - 1$, producing a message-oblivious decoy histogram. The construction and its shift-affinity identity are given below.

Construction (Correlated noise on a finite simplex). Fix $d \ge 2$ and let $n \ge d$. Define

$$L = \left\lfloor \frac{n-1}{d-1} \right\rfloor + 1, \qquad \theta = \frac{\pi}{L+1},$$

and let

$$r(j) = \frac{2}{L+1} \sin^2((j+1)\theta), \qquad j = 0, \ldots, L-1.$$

Independently of M, draw $X_1, \ldots, X_{d-1}$ independently from r, and set

$$S = \left(X_1, \ldots, X_{d-1}, n - 1 - \sum_{i=1}^{d-1} X_i \right). \tag{8}$$

Since

$$(d-1)(L-1) \le n - 1,$$

the last coordinate is nonnegative, and the total mass is exactly $n - 1$. Interpreting S as the histogram of a decoy multiset therefore defines a valid message-oblivious mechanism: the decoy histogram has the required total mass and is independent of M.

Lemma 3.3 (Square-root affinity under a unit shift)**.** *Let r be the sine-squared distribution defined above, extended by zero outside $\{0, \ldots, L-1\}$, and define $r^+(j) = r(j-1)$. Then*

$$\sum_{j \in \mathbb{Z}} \sqrt{r(j) r^+(j)} = \cos\theta, \qquad \left\| \sqrt{r} - \sqrt{r^+} \right\|_2^2 = 2 - 2\cos\theta. \tag{9}$$

Square-root affinity factorizes across independent coordinates. The preceding one-dimensional identity therefore controls every pair of conditional output distributions in our construction. Applying Lemma 3.2 yields the following $O_d(n^{-2})$ leakage bound, uniformly over the message prior.

Theorem 3.4 (Correlated-noise upper bound for finite alphabets)**.** *Fix $d \geq 2$ and let $n \geq d$. For any prior P on $[d]$, let $M \sim P$, generate S according to* (8) *independently of M, and release $T = e_M + S$. Then*

$$I(M;T) \leq \frac{2d(d-1)\pi^2}{(L+1)^2} = O_d(n^{-2}), \qquad L = \left\lfloor \frac{n-1}{d-1} \right\rfloor + 1. \tag{10}$$

In particular, this bound is uniform over P.

The construction attains the same n^{-2} order as the universal message-oblivious lower bound in Theorem 3.1. Combining these bounds determines the optimal leakage rate for the entire message-oblivious class. Together with Theorem 3.3, this establishes an improvement by a factor of order n over the optimal independent-decoy mechanism.

Corollary 3.1 (Optimal leakage rate for message-oblivious decoys)**.** *For every fixed non-degenerate prior P on a finite alphabet of size d,*

$$L_{\mathrm{obl}}(P,n) = \Theta_{P,d}(n^{-2}). \tag{11}$$

Proof. The upper bound follows from Theorem 3.4. For the matching lower bound, choose any symbol y with $0 < p_y < 1$. The general message-oblivious leakage lower bound from Theorem 3.1 gives

$$I(M;T) \geq \frac{2p_y(1-p_y)}{n^2}$$

for every message-oblivious mechanism. Since P is fixed and non-degenerate, this yields $L_{\mathrm{obl}}(P,n) = \Omega_P(n^{-2})$, matching the upper bound. □

3.3 Separation between i.i.d. and Heterogeneous Decoys

We finally ask whether allowing different decoy distributions can improve secrecy while retaining independence. For each fixed prior, canonical and heterogeneous mechanisms both have optimal leakage of order n^{-1}, but this does not preclude a substantial gap between their optimal values. We expose such a gap in a different asymptotic regime: fixing $n = 3$ and letting a binary prior become increasingly rare. The following theorem shows that asymmetric heterogeneous decoys achieve an arbitrarily large multiplicative improvement over the optimal canonical mechanism.

Theorem 3.5 (Asymptotic separation between canonical and heterogeneous shuffling)**.** *Consider the binary message space $\mathcal{M} = \{0,1\}$ with $n = 3$ and target distribution $P_p = \mathrm{Bern}(p)$.*

As $p \to 0$, letting $L = \log(1/p)$, we have

$$L_{\mathrm{can}}(P_p,3) = \left(\frac{1}{2} + o(1)\right) p \log L, \tag{12}$$

whereas

$$L_{\mathrm{het}}(P_p,3) = O(p \log\log L). \tag{13}$$

Consequently,

$$\lim_{p\to 0} \frac{L_{\mathrm{het}}(P_p,3)}{L_{\mathrm{can}}(P_p,3)} = 0.$$

In particular, for every sufficiently small fixed $p > 0$, $L_{\mathrm{het}}(P_p,3) < L_{\mathrm{can}}(P_p,3)$.

The separation arises because identical decoy distributions couple interior overlap with boundary exposure. Writing $T = M + S$ for the number of output ones, canonical decoys with parameter q satisfy $\Pr(S = 2) = q^2$. The event $T = 3$ reveals the rare message and contributes at least pq^2L to leakage, while making q small increases the evidence for that message supplied by $T = 1$. Balancing these effects yields the canonical optimum, attained to leading order by $q = L^{-1/2}$. Heterogeneous decoys can control the two effects separately: choosing Bernoulli parameters $q_1 = L^{-1}$ and $q_2 = (\log L)^{-1}$ retains $\Pr(S = 1) = \Theta((\log L)^{-1})$ while reducing $\Pr(S = 2)$ to $1/(L \log L)$. A reference-distribution KL bound then gives $O(p \log\log L)$ leakage. The full proof appears in Appendix B.8.

4 Connections to Information-Theoretic Security of Symmetric Encryption

Decrypting a ciphertext under every possible key produces a collection of candidate plaintexts, one of which is genuine. This observation connects symmetric encryption to hiding among decoys. We formalize this connection by transforming any perfectly correct symmetric encryption scheme with a uniform key into a message-dependent shuffle-only cipher without increasing plaintext information leakage.

Using the pre-image-constrained channel-design results of [KM16, KM19], we then characterize optimal message secrecy. When the number of keys does not exceed the plaintext support size, we further show that every cipher attaining the Shannon optimum has the same key leakage, determined by its optimal posterior distribution. Throughout this section, we consider a single encryption experiment and measure leakage about its plaintext and key from the resulting ciphertext.

4.1 Reduction from General Encryption to the Shuffle-Only Model

We start with a decryption-vector reduction lemma:

Lemma 4.1 (Decryption-vector reduction)**.** *Let $\Pi = (\mathsf{Enc}, \mathsf{Dec})$ be a perfectly correct symmetric encryption scheme with finite message space $\mathcal{M}$, key space $\mathcal{K} = [n]$, possibly randomized encryption, and deterministic decryption $\mathsf{Dec}_k : \mathcal{C} \to \mathcal{M}$. There exists a perfectly correct scheme $\widehat{\Pi} = (\widehat{\mathsf{Enc}}, \widehat{\mathsf{Dec}})$ with ciphertext space $\mathcal{M}^n$ and decryption rule $\widehat{\mathsf{Dec}}_k(c) = c[k]$ that does not increase any plaintext leakage measure satisfying the data-processing inequality.*

Proof. Define the decryption vector

$$D(c) := \big(\mathsf{Dec}_1(c), \ldots, \mathsf{Dec}_n(c)\big)$$

and set

$$\widehat{\mathsf{Enc}}_k(m) := D\big(\mathsf{Enc}_k(m)\big).$$

Perfect correctness of Π gives, almost surely, $\widehat{\mathsf{Dec}}_k\big(\widehat{\mathsf{Enc}}_k(m)\big) = \mathsf{Dec}_k\big(\mathsf{Enc}_k(m)\big) = m$. Moreover, $\widehat{C} = D(C)$ is a deterministic post-processing of C, so the claimed leakage inequality follows by data processing. □

We now formalize the key-symmetrization reduction outlined in Section 1.2.2.

Theorem 4.1 (Key symmetrization and shuffle reduction)**.** *Let $\Pi = (\mathsf{Enc}, \mathsf{Dec})$ be any perfectly correct symmetric encryption scheme with key space $\mathcal{K} = [n]$, a uniformly distributed key K (independent of M), and plaintext prior $M \sim P$. Suppose its ciphertext space is $\mathcal{M}^n$, with $\mathsf{Dec}_k(c) = c[k]$.*

Construct a new encryption scheme $\widetilde{\Pi} = (\widetilde{\mathsf{Enc}}, \widetilde{\mathsf{Dec}})$ *as follows:*

$$\widetilde{\mathsf{Enc}}_k(m) = \sigma^{-1}\big(\mathsf{Enc}_{\sigma(k)}(m)\big), \qquad \sigma \leftarrow_{\$} S_n,$$
$$\widetilde{\mathsf{Dec}}_k(\boldsymbol{z}) = z[k],$$

where $\sigma^{-1}(c)$ *denotes the coordinate permutation defined by* $(\sigma^{-1}(c))[i] = c[\sigma(i)]$.

Then $\widetilde{\Pi}$ *is perfectly correct and has the form of a message-dependent shuffle-only cipher: its ciphertext is a uniformly shuffled vector consisting of the plaintext and* $n-1$ *possibly correlated, message-dependent decoys.*

Moreover, for every plaintext prior P*, the ciphertext* $\widetilde{C}$ *of* $\widetilde{\Pi}$ *is a randomized post-processing of the original ciphertext* C*. Consequently,* $I(M;\widetilde{C}) \le I(M;C)$*. More generally, the reduction cannot increase any plaintext information leakage measure satisfying the data-processing inequality.*

Proof. Perfect correctness follows immediately: for any k and m, letting $c = \mathsf{Enc}_{\sigma(k)}(m)$, we have

$$\widetilde{\mathsf{Dec}}_k\big(\widetilde{\mathsf{Enc}}_k(m)\big) = c[\sigma(k)] = m.$$

To establish the shuffle-only representation, fix a plaintext m. Consider the original encryption experiment $J \leftarrow_{\$} [n]$ and $C \leftarrow \mathsf{Enc}_J(m)$. By perfect correctness, $C[J] = m$.

Let H be the vector of the remaining $n-1$ coordinates of C, listed in increasing order of their original indices. Its conditional distribution given $M = m$ defines a possibly correlated and message-dependent decoy distribution R_m.

In the transformed experiment, sample $K \leftarrow_{\$} [n]$ and $\sigma \leftarrow_{\$} S_n$ independently, and set $J = \sigma(K)$. Then J is uniform on $[n]$. Moreover, conditioned on $J = j$, the permutation σ is uniform over S_n and $K = \sigma^{-1}(j)$.

Consequently, conditional on $(M = m, J = j, C = c)$, the transformed ciphertext $\widetilde{C} = \sigma^{-1}(c)$ is a uniformly random permutation of the coordinates of c. Equivalently, it is obtained by uniformly shuffling m together with the decoy vector H. Thus $\widetilde{\Pi}$ has the claimed message-dependent shuffle-only form.

Finally, under a uniform key, $J = \sigma(K)$ is independent of σ and has the same distribution as the original key. Hence the joint distribution of $(M, \widetilde{C})$ is identical to that obtained by sampling the original ciphertext C and independently applying a uniform coordinate permutation:

$$\widetilde{C} \overset{d}{=} \sigma^{-1}(C), \qquad \sigma \leftarrow_{\$} S_n.$$

Therefore,

$$M \longrightarrow C \longrightarrow \widetilde{C}$$

forms a Markov chain, and the data-processing inequality gives $I(M;\widetilde{C}) \le I(M;C)$. The same argument applies to any plaintext leakage measure satisfying data processing. □

4.2 Optimal Message Secrecy and the Induced Key Leakage

We use the channel-design results of Khouzani and Malacaria [KM16, KM19] to characterize optimal message secrecy under message-dependent shuffling. We then determine the key leakage of every cipher attaining this optimum under Shannon mutual information.

Throughout this subsection, let $M \sim P = (p_1, \dots, p_d)$, where $p_1 \ge \cdots \ge p_d > 0$, and assume $2 \le n \le d$. Ciphertext spaces are discrete, encryption may be randomized, and decryption is deterministic and perfectly correct. The key K is uniform on $[n]$ and independent of M.

From encryption to a pre-image constraint. For a channel $W(c \mid m)$, define

$$\mathrm{Pre}_W(c) := \{m \in [d] : W(c \mid m) > 0\}.$$

Every perfectly correct cipher with n keys satisfies

$$\mathrm{Pre}_W(c) \subseteq \{\mathrm{Dec}_k(c) : k \in [n]\}, \qquad |\mathrm{Pre}_W(c)| \le n. \tag{14}$$

Thus, the minimum leakage under the pre-image constraint lower-bounds the leakage of every such cipher. Conversely, the optimal channel constructed in [KM16, KM19] can be realized as a message-dependent shuffle-only cipher, so the two optimization problems have the same optimal value.

The Khouzani–Malacaria framework. The generalized conditional entropies considered in [KM16, KM19] have the form

$$\mathsf{H}(P) = \eta(F(P)), \qquad \mathsf{H}(M \mid C) = \eta\left(\sum_c \Pr(C = c) F(P_{M|C=c})\right). \tag{15}$$

Here F is symmetric and expansible: permuting its arguments or appending zero-probability entries does not change its value. Their core-concavity condition requires either that F be concave and η increasing, or that F be convex and η decreasing. This class includes Shannon entropy, several conditional Rényi entropies, and guessing-based uncertainty measures. Leakage is defined as $\mathsf{H}(P) - \mathsf{H}(M \mid C)$.

Define

$$j^* := \min\left\{ j \in [n] : p_j \le \frac{\sum_{i=j}^{d} p_i}{n - j + 1} \right\}, \qquad t^* := \frac{\sum_{i=j^*}^{d} p_i}{n - j^* + 1}, \tag{16}$$

and let

$$\boldsymbol{\pi}^* := \left(p_1, \ldots, p_{j^*-1}, \underbrace{t^*, \ldots, t^*}_{n-j^*+1} \right). \tag{17}$$

The largest prior probabilities are retained, and the remaining mass is distributed equally among the remaining coordinates.

Theorem 4.2 (Optimal posterior entropy [KM16, KM19])**.** *For every generalized entropy in* (15)*,*

$$\sup_{W : |\mathrm{Pre}_W(c)| \le n \ \forall c} \mathsf{H}(M \mid C) = \mathsf{H}(\boldsymbol{\pi}^*). \tag{18}$$

A single channel attains this bound simultaneously for all these entropies: every output of positive probability has a posterior whose nonzero entries are a permutation of $\boldsymbol{\pi}^*$*.*

Their converse combines sorting, Jensen's inequality, and majorization.

For each output c with $\Pr(C = c) > 0$, let $r_c \in \Delta_n$ consist of its nonzero posterior probabilities, sorted in nonincreasing order and padded with zeros to length n. Write

$$\bar{r} := \sum_{c : \Pr(C=c) > 0} \Pr(C = c) r_c.$$

For every $\ell < n$,

$$\sum_{i=1}^{\ell} \bar{r}_i \geq \sum_{i=1}^{\ell} p_i.$$

To see that $\bar{r} \succeq \boldsymbol{\pi}^*$, write $a = j^* - 1$ and $S_\ell = \sum_{i=1}^{\ell} \bar{r}_i$. For $\ell \leq a$, the preceding inequalities give $S_\ell \geq \sum_{i=1}^{\ell} \pi_i^*$. For $a < \ell < n$, the ordering of $\bar{r}$ gives

$$S_\ell \geq S_a + \frac{\ell - a}{n - a}(1 - S_a) \geq \sum_{i=1}^{a} p_i + (\ell - a)t^* = \sum_{i=1}^{\ell} \pi_i^*.$$

Hence $\bar{r}$ majorizes $\boldsymbol{\pi}^*$.

In the concave case, Jensen's inequality and Schur concavity give

$$\sum_c \Pr(C = c) F(r_c) \leq F(\bar{r}) \leq F(\boldsymbol{\pi}^*).$$

Applying η yields the bound; the convex case follows by replacing F with $-F$ and $\eta(u)$ with $\eta(-u)$.

For the key-leakage result, we also need necessity of the posterior condition. The following equality criterion follows from the same argument. We use H for generalized entropies and reserve H for Shannon entropy.

Proposition 4.1 (Equality at the Shannon optimum)**.** *Among channels W satisfying $|\operatorname{Pre}_W(c)| \leq n$ for every output c, a channel minimizes Shannon message leakage if and only if, for every c with $\Pr(C = c) > 0$,*

$$r_c = \boldsymbol{\pi}^*. \tag{19}$$

The same characterization holds when either F is strictly concave and η is strictly increasing, or F is strictly convex and η is strictly decreasing.

Proof. For Shannon entropy,

$$H(M \mid C) = \sum_c \Pr(C = c) H(r_c) \leq H(\bar{r}) \leq H(\boldsymbol{\pi}^*).$$

At an optimum, both inequalities are equalities. Strict Schur concavity gives $\bar{r} = \boldsymbol{\pi}^*$, since both vectors are sorted. Strict concavity in the first inequality then gives $r_c = \bar{r}$ for every c of positive probability. Conversely, this posterior condition attains the upper bound. The argument extends to the stated strict generalized entropies[4]. □

Realizing the optimal channel as a shuffle cipher. The Khouzani–Malacaria construction [KM16, KM19] outputs an n-element candidate set A containing the input message, with conditional distribution $W(A \mid m)$. Every positive-probability output induces a posterior whose nonzero probabilities are a permutation of $\boldsymbol{\pi}^*$.

To encrypt m under key $K = k$, sample A according to $W(\cdot \mid m)$, place m in coordinate k, and arrange the elements of $A \setminus \{m\}$ uniformly at random in the remaining coordinates. Decryption reads the coordinate indexed by the key, so the scheme is perfectly correct.

[4] Without strictness, the posterior condition remains sufficient but need not be necessary; min-entropy, for example, may admit additional optimal channels with different posterior vectors.

Averaging over the independent uniform key makes the ciphertext $\boldsymbol{Z}$ a uniformly random ordering of A, independently of M conditional on A. Since A is also recoverable from $\boldsymbol{Z}$, we have

$$I(M;\boldsymbol{Z}) = I(M;A) = H(P) - H(\boldsymbol{\pi}^*).$$

Thus the optimal pre-image-constrained channel is realized by a message-dependent shuffle-only cipher.

Theorem 4.3 (Key leakage at optimal message secrecy)**.** *Let $\mathcal{E}_n$ denote the class of perfectly correct symmetric encryption schemes with key space $[n]$ and a uniform key independent of the plaintext. Under the assumptions of this subsection,*

$$\inf_{\Pi \in \mathcal{E}_n} I_P(M; C_\Pi) = L_{\mathrm{dep}}(P, n) = H(P) - H(\boldsymbol{\pi}^*). \tag{20}$$

Every cipher attaining this minimum satisfies

$$I(K; C_\Pi) = \log n - H(\boldsymbol{\pi}^*). \tag{21}$$

In particular, the shuffle cipher constructed above satisfies $I(K;\boldsymbol{Z}) = \log n - H(\boldsymbol{\pi}^*)$.

Proof. Perfect correctness implies the pre-image constraint (14), so Theorem 4.2 gives the lower bound $H(P) - H(\boldsymbol{\pi}^*)$. The shuffle realization described above attains this bound, proving (20).

Now consider any cipher attaining the minimum. Proposition 4.1 implies that every positive-probability ciphertext c has a posterior with exactly n positive entries, since all entries of $\boldsymbol{\pi}^*$ are positive. Consequently, conditional on $C_\Pi = c$, the message $M = \mathrm{Dec}_K(c)$ is a bijective function of K. Their posterior probability vectors therefore differ only by a permutation, giving

$$H(K \mid C_\Pi) = H(M \mid C_\Pi) = H(\boldsymbol{\pi}^*).$$

Since $H(K) = \log n$, the claimed key-leakage identity follows. □

Remark 4.1. *The restriction $n \le d$ is essential to this argument. When $n > d$, perfect message secrecy need not have this property (collision-free decryption). The induced key leakage is zero precisely when $p_1 \le 1/n$, (i.e., $n \le \frac{1}{p_1}$) in which case $\boldsymbol{\pi}^*$ is uniform.*

5 Mutual Information Leakage in the Canonical Shuffle-Only Model

In this section, we investigate information leakage in the canonical shuffle-only model under Shannon mutual information. We distinguish two fundamental privacy objectives: *position privacy*, which concerns the adversary's uncertainty about the position of the target message in the shuffled output, and *message privacy*, which concerns the information revealed about the target message itself.

We derive exact information-theoretic identities and asymptotic characterizations for both objectives, together with explicit finite-sample error bounds. Our results reveal a fundamental distinction between position and message leakage, and lead to a characterization of the asymptotically optimal blanket distribution for message privacy.

Our analysis primarily considers distributions P and Q on a common finite alphabet with $P \ll Q$, although the techniques extend naturally to continuous distributions under suitable regularity conditions.

5.1 Information Leakage of Position

We begin with position leakage, measured by $I(K; \boldsymbol{Z})$, where K denotes the position of the target message in the shuffled output. The following theorem provides an exact decomposition of position leakage in terms of KL divergences. It further shows that, as the number of shuffled messages increases, position leakage approaches $D_{\mathrm{KL}}(P\|Q)$, with a finite-sample error bound.

Theorem 5.1 (Position leakage: Mutual information)**.** *Consider the canonical shuffle-only model, where $Y_1 \sim P$ is shuffled with $n-1$ independent messages drawn from Q. Let M_n denote the distribution of the shuffled output $\boldsymbol{Z}$, and let $Q^{\otimes n}$ be the n-fold product distribution of Q.*

Suppose that $P \ll Q$, i.e., $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$. Then the mutual information between the target position K and the shuffled output satisfies

$$I(K; \boldsymbol{Z}) = D_{\mathrm{KL}}(P\|Q) - D_{\mathrm{KL}}(M_n\|Q^{\otimes n}) \tag{22}$$

$$= D_{\mathrm{KL}}(P\|Q) - \frac{\chi^2(P\|Q)}{2n} + O(n^{-3/2}), \qquad n \to \infty, \tag{23}$$

where $\chi^2(P\|Q) = \sum_{y\in\mathrm{supp}(Q)} \frac{(P(y)-Q(y))^2}{Q(y)}$.

Moreover, letting $C = \max_{y\in\mathrm{supp}(Q)} \left|\frac{P(y)}{Q(y)} - 1\right|$, the approximation admits the following explicit non-asymptotic error bound, valid for every $n \geq 1$:

$$\left| I(K; \boldsymbol{Z}) - D_{\mathrm{KL}}(P\|Q) + \frac{\chi^2(P\|Q)}{2n} \right| \leq \frac{C^3\sqrt{3n-2}}{2n^2}. \tag{24}$$

Theorem 5.1 shows that increasing the number of blanket messages does not necessarily eliminate position leakage. For fixed P and Q, the leakage converges to $D_{\mathrm{KL}}(P\|Q)$, which is strictly positive whenever $P \neq Q$. When $Q = P$, perfect position privacy holds for every n.

The theorem also reveals that $I(K; \boldsymbol{Z})$ is nondecreasing in n. This follows from the fact that $D_{\mathrm{KL}}(M_n\|Q^{\otimes n})$ is nonincreasing: adding an independent report from Q and uniformly reshuffling maps M_n to M_{n+1} and $Q^{\otimes n}$ to $Q^{\otimes(n+1)}$, so the claim follows by data processing.

5.2 Information Leakage of Message

We now consider message leakage, quantified by $I(Y_1; \boldsymbol{Z})$. The next theorem establishes that message leakage vanishes at rate $1/n$ for any fixed blanket distribution Q satisfying $P \ll Q$. More importantly, its leading-order coefficient admits an explicit expression depending on both P and Q, allowing us to identify which blanket distributions provide the strongest asymptotic protection for the target message.

Theorem 5.2 (Message leakage: Mutual information)**.** *Consider the canonical shuffle-only model, where $Y_1 \sim P$ is shuffled with $n-1$ independent messages drawn from Q. Assume $P \ll Q$ over a finite alphabet. Then*

$$I(Y_1; \boldsymbol{Z}) = \frac{1}{2n} \sum_{y\in\mathrm{supp}(P)} \frac{P(y)(1-P(y))}{Q(y)} + O(n^{-3/2}), \qquad n \to \infty.$$

An explicit finite-sample remainder bound is given in Appendix C.3.

Theorem 5.2 identifies the leading-order coefficient of message leakage, which makes it possible to optimize the blanket distribution directly.

Corollary 5.1 (Asymptotically optimal blanket for mutual information)**.** *Consider the canonical shuffle-only model with a fixed target distribution P whose support contains at least two elements. Among blanket distributions supported on* $\mathrm{supp}(P)$*, the leading-order term of $I(Y_1; \boldsymbol{Z})$ is uniquely minimized by*

$$Q^*(i) = \frac{\sqrt{p_i(1-p_i)}}{\sum_{j\in\mathrm{supp}(P)} \sqrt{p_j(1-p_j)}}, \qquad i \in \mathrm{supp}(P). \tag{25}$$

In particular, for a binary target distribution, the asymptotically optimal blanket is uniform over $\mathrm{supp}(P)$*.*

Proof. By Theorem 5.2, the leading-order coefficient of $I(Y_1; \boldsymbol{Z})$ is $\frac{1}{2}\sum_{i\in\mathrm{supp}(P)} \frac{p_i(1-p_i)}{q_i}$. Let $a_i = p_i(1-p_i) > 0$. By the Cauchy–Schwarz inequality,

$$\sum_i \frac{a_i}{q_i} \geq \frac{\left(\sum_i \sqrt{a_i}\right)^2}{\sum_i q_i} = \left(\sum_i \sqrt{a_i}\right)^2,$$

with equality if and only if $q_i \propto \sqrt{a_i}$. Normalizing gives (25).

For a binary target distribution $P = (p, 1-p)$ with $0 < p < 1$, we have $p(1-p) = (1-p)p$. Hence, (25) yields $Q^* = (1/2, 1/2)$, independently of p. □

Corollary 5.1 characterizes the asymptotically optimal blanket for a given message distribution P. The following theorem shows that, when P is unknown, the uniform distribution on the message domain is minimax optimal for the leading asymptotic leakage coefficient.

Theorem 5.3 (Minimax blanket for the asymptotic leakage coefficient)**.** *Fix an integer $d \geq 2$. Let Δ_d denote the probability simplex on $[d]$, and let Δ_d° denote its interior. For $P \in \Delta_d$ and $Q \in \Delta_d^\circ$, define*

$$\Phi(P, Q) := \frac{1}{2}\sum_{y=1}^{d} \frac{p_y(1-p_y)}{q_y}.$$

Then

$$\inf_{Q\in\Delta_d^\circ} \sup_{P\in\Delta_d} \Phi(P, Q) = \frac{d-1}{2},$$

and the infimum is attained uniquely by the uniform blanket $Q^* = U_d$.

Proof. Lower bound. Fix any $Q \in \Delta_d^\circ$ and choose the uniform prior $P = U_d$. Then

$$\Phi(U_d, Q) = \frac{d-1}{2d^2}\sum_{y=1}^{d} \frac{1}{q_y}.$$

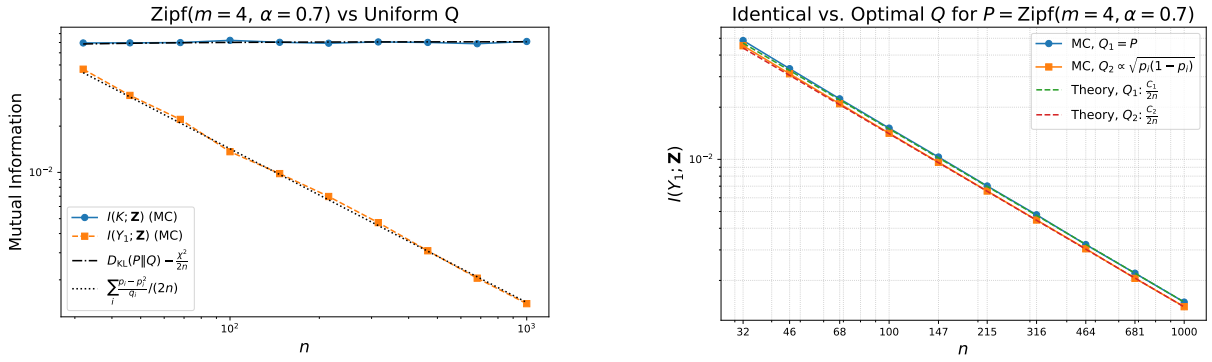
By the Cauchy–Schwarz inequality, $\sum_{y=1}^{d} \frac{1}{q_y} \geq \frac{\left(\sum_{y=1}^{d} 1\right)^2}{\sum_{y=1}^{d} q_y} = d^2$, with equality if and only if $q_y = 1/d$ for every y. Consequently,

$$\sup_{P\in\Delta_d} \Phi(P, Q) \geq \Phi(U_d, Q) \geq \frac{d-1}{2},$$

and the final inequality is strict whenever $Q \neq U_d$.

Upper bound. Choose $Q = U_d$. For every $P \in \Delta_d$,

$$\Phi(P, U_d) = \frac{d}{2}\sum_{y=1}^{d} p_y(1-p_y) = \frac{d}{2}\left(1 - \sum_{y=1}^{d} p_y^2\right) \leq \frac{d}{2}\left(1 - \frac{1}{d}\right) = \frac{d-1}{2},$$



(a) (b)

Figure 2: Concrete vs. asymptotic MI in the canonical shuffle-only setting and empirical validation of the optimal distribution Q.

where we used $\sum_y p_y^2 \geq 1/d$. Thus the uniform blanket attains the global lower bound. The strictness established above for every non-uniform Q proves uniqueness. □

Example 5.1. *We conduct numerical experiments using a Zipf distribution $P = Zipf(m, \alpha)$ with $m = 4$ and $\alpha = 0.7$ ($p_i \propto i^{-\alpha}$), and a uniform distribution Q over $[4] = \{1, 2, 3, 4\}$. Given that closed-form expressions for $I(K; \boldsymbol{Z})$ and $I(Y_1; \boldsymbol{Z})$ are analytically intractable, we estimate these values via Monte Carlo simulations.*[5] *As illustrated in Fig. 2a, the empirical estimates of mutual information exhibit tight alignment with the theoretical asymptotic curves.*

Next, we empirically validate Corollary 5.1 using the same $P = Zipf(4, 0.7)$ setup. We compare the baseline $Q_1 = P$ against the theoretically optimal choice Q_2, where $q_i \propto \sqrt{p_i(1-p_i)}$. The resulting Monte Carlo curves and their corresponding theoretical predictions are displayed in Fig. 2b. For this configuration, the constants are $C_1 = m - 1 = 3$ and $C_2 = \left(\sum_i \sqrt{p_i(1-p_i)}\right)^2 \approx 2.81$. The empirical results are in high agreement with our theoretical findings; notably, the transition from $C_1 = 3$ to $C_2 \approx 2.81$ yields a 6.3% reduction in asymptotic leakage.

6 Total-Variation Information in the Canonical Shuffle-Only Model

Having characterized information leakage under Shannon mutual information, we now turn to TV-information, which measures the statistical dependence between a private variable and the shuffled output in terms of total variation distance:

$$I_{\mathrm{TV}}(X; Z) := D_{\mathrm{TV}}(P_{X,Z}, P_X \otimes P_Z) = \mathbb{E}_Z\left[D_{\mathrm{TV}}(P_{X|Z}, P_X)\right],$$

with $D_{\mathrm{TV}}(A, B) = \frac{1}{2}\|A - B\|_1$.

As in the preceding section, we separately investigate position leakage and message leakage. We establish asymptotic characterizations and finite-sample bounds for both privacy objectives, and further identify the exact optimal blanket distribution for binary message privacy under TV-information.

[5] Specifically, we compute $I(K; \boldsymbol{Z}) = \mathbb{E}_{\boldsymbol{Z}}\left[\sum_{k=1}^{n} \Pr(K = k \mid \boldsymbol{Z}) \log\left(n \Pr(K = k \mid \boldsymbol{Z})\right)\right]$. To ensure statistical convergence, expectations are approximated by the sample mean across 10^5 independent realizations of $\boldsymbol{Z}$.

6.1 Information Leakage of Position

We first examine the information revealed about the target position K. Under TV-information, position leakage quantifies the statistical dependence between K and the shuffled output, providing a counterpart to the KL-based characterization established earlier.

The following theorem shows that position leakage converges to the total variation distance between the target and blanket distributions, with an explicit finite-sample convergence bound.

Theorem 6.1 (Position leakage: TV-information)**.** *Consider the canonical shuffle-only model, where $Y_1 \sim P$ is shuffled with $n-1$ independent messages drawn from Q over a finite alphabet. Assume $P \ll Q$. Then, for every $n \geq 1$,*

$$|I_{\mathrm{TV}}(K; \boldsymbol{Z}) - D_{\mathrm{TV}}(P, Q)| \leq \frac{1}{2}\sqrt{\frac{\chi^2(P\|Q)}{n}}. \tag{26}$$

In particular,

$$\lim_{n\to\infty} I_{\mathrm{TV}}(K; \boldsymbol{Z}) = D_{\mathrm{TV}}(P, Q).$$

Theorem 6.1 establishes a close parallel between TV-information and Shannon mutual information for position privacy. In both cases, the asymptotic leakage is determined by the discrepancy between P and Q: $D_{\mathrm{TV}}(P, Q)$ for TV-information and $D_{\mathrm{KL}}(P\|Q)$ for mutual information. Consequently, choosing $Q = P$ achieves perfect position privacy under either measure.

6.2 Information Leakage of Message

We now turn to message leakage, quantified by $I_{\mathrm{TV}}(Y_1; \boldsymbol{Z})$. We first derive an exact combinatorial expression for the TV-information between the target message and the shuffled output. The key observation is that, under the i.i.d. blanket reference distribution, the shuffled output is completely characterized by its message-frequency vector. This allows us to express TV-information as an expectation over a multinomial distribution.

Theorem 6.2 (Message leakage: TV-information)**.** *Consider the canonical shuffle-only model over a finite alphabet[6] $[d] = \{1, \ldots, d\}$, where $Y_1 \sim P$ is shuffled with $n-1$ independent samples from Q. Assume $P \ll Q$, and write $p_i = P(i)$ and $q_i = Q(i)$. Then the TV-information admits the exact representation*

$$I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) = \frac{1}{2}\sum_{i=1}^{d} p_i \mathbb{E}_{\boldsymbol{N}\sim\mathrm{Mult}(n,Q)}\left[\left|\frac{N_i}{nq_i} - \sum_{j=1}^{d} p_j \frac{N_j}{nq_j}\right|\right]. \tag{27}$$

The proof follows directly from the likelihood-ratio representation of the canonical shuffle model and is deferred to Appendix D.2.

Although the exact expression applies to arbitrary finite alphabets, optimizing it over the blanket distribution Q generally leads to a nontrivial combinatorial problem. Nevertheless, the following numerical experiments show that the optimal Q is also not necessarily equal to P.

Example 6.1. *Consider a ternary alphabet with a prior distribution $P = (0.1, 0.3, 0.6)$. For a fixed sample size $n = 10$, we evaluate the total-variation information $I_{\mathrm{TV}}(Y_1; \boldsymbol{Z})$ by varying the parameters (q_2, q_3) of the blanket noise distribution $Q = (1 - q_2 - q_3, q_2, q_3)$. Figure 3 illustrates the landscape of I_{TV} as a function of the noise configuration. Notably, the smallest value found in our numerical search is at $(q_2, q_3) \approx (0.400, 0.400)$, yielding an I_{TV} of 0.146. This numerically optimized Q provides a strictly stronger privacy guarantee than the standard heuristic of setting $Q = P$ ($q_2 = 0.3, q_3 = 0.6$), which results in a higher information leakage of 0.151.*

[6] Without loss of generality, label the elements of $\mathrm{supp}(Q)$ by $[d]$, so that $q_i > 0$ for every $i \in [d]$.

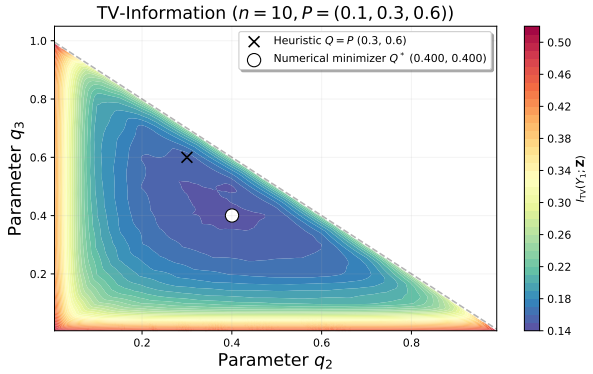


Figure 3: TV-information as a function of the blanket distribution $Q = (1 - q_2 - q_3, q_2, q_3)$.

Non-asymptotic optimality for binary alphabets. Remarkably, the optimal blanket is always quasi-uniform, regardless of the bias in P.

Theorem 6.3 (Optimal blanket for binary TV-information)**.** *Consider the canonical shuffle-only model over a binary alphabet, with a non-degenerate target distribution $P = (p, 1-p)$, where $0 < p < 1$. For every $n \geq 2$, the TV-information $I_{\mathrm{TV}}(Y_1; \boldsymbol{Z})$ is minimized by the quasi-uniform blanket distribution*

$$Q^* = \left(\frac{\lfloor n/2 \rfloor}{n}, \frac{\lceil n/2 \rceil}{n} \right),$$

or its permutation. In particular, the optimal blanket is independent of the target distribution P.

The proof relies on a useful reduction from TV-information to a one-dimensional binomial optimization problem. Writing $Q = (1 - q, q)$, a direct calculation yields

$$I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) = \frac{p(1-p)}{n} \, \phi_n(q), \tag{28}$$

where

$$\phi_n(q) = \frac{1}{q(1-q)} \mathbb{E}_{N \sim \mathrm{Bin}(n,q)} \left[|N - nq| \right].$$

The derivation of (28) is provided in Appendix D.3.

Crucially, the dependence on the target distribution appears only through the multiplicative factor $p(1-p)$, which is independent of q. Consequently, finding the optimal blanket is equivalent to minimizing the normalized mean absolute deviation $\phi_n(q)$ of a binomial random variable.

The following combinatorial lemma establishes that this quantity is minimized when the binomial parameter is as close as possible to $1/2$, subject to the discretization induced by n.

Lemma 6.1 (Optimal normalized binomial deviation)**.** *Let $n \geq 2$ be an integer, and define*

$$\phi_n(q) := \frac{\mathbb{E}_{N \sim \mathrm{Bin}(n,q)}[|N - nq|]}{q(1-q)}, \qquad q \in (0, 1).$$

Then $\phi_n(q)$ attains its global minimum at

$$q^* = \frac{\lfloor n/2 \rfloor}{n} \qquad \text{and} \qquad q^* = \frac{\lceil n/2 \rceil}{n}.$$

The proof of Lemma 6.1, given in Appendix D.4, exploits the combinatorial structure of binomial probabilities. Together with (28), it immediately establishes Theorem 6.3.

7 Quantitative Information Flow in the Canonical Shuffle-Only Model

Beyond mutual information and TV-information, we investigate two operational measures from quantitative information flow (see Appendix A.2): message recovery advantage (MRA) and expected posterior variance (EPV). We derive exact characterizations of both measures in the canonical shuffle-only model, providing a basis for evaluating privacy under different adversarial objectives.

For EPV, we further establish an asymptotic characterization that yields a closed-form expression for the optimal blanket distribution. Remarkably, when the target distribution P is supported on two distinct real values, the asymptotically optimal blanket is uniform over its support, independently of the probabilities assigned by P.

7.1 Analysis of Message Recovery Advantage

Message recovery advantage measures the increase in the optimal one-shot guessing success probability after observing $\boldsymbol{Z}$:

$$\mathrm{Adv}_{\mathrm{MR}}(Y_1; \boldsymbol{Z}) := \mathbb{E}_{\boldsymbol{Z}}\left[\max_y \Pr(Y_1 = y \mid \boldsymbol{Z})\right] - \max_y \Pr(Y_1 = y).$$

The following theorem gives an exact combinatorial expression for MRA in terms of multinomial counts under the blanket distribution. The alphabet is restricted to $\mathrm{supp}(Q)$, so $q_i > 0$ for every indexed symbol.

Theorem 7.1 (Message recovery advantage)**.** *Consider the canonical shuffle-only model over a finite alphabet $[d] = \{1, \ldots, d\}$, where $Y_1 \sim P$ is shuffled with $n-1$ independent samples from Q. Assume $P \ll Q$, and write $p_i = P(i)$ and $q_i = Q(i)$.*

Then the expected message recovery advantage admits the exact representation

$$\mathrm{Adv}_{\mathrm{MR}}(Y_1; \boldsymbol{Z}) = \mathbb{E}_{\boldsymbol{N} \sim \mathrm{Mult}(n,Q)}\left[\max_{i \in [d]} \frac{p_i N_i}{n q_i}\right] - \max_{i \in [d]} p_i. \tag{29}$$

The proof follows from Bayes' rule and a change of measure to the i.i.d. blanket distribution $Q^{\otimes n}$, and is deferred to Appendix E.1.

[SCW26a] analyzed the analogous key recovery advantage in the canonical shuffle-only model, defined as the increase in the adversary's optimal probability of guessing K over the prior success probability $1/n$. As $n \to \infty$, this advantage is $(\alpha - 1)/n + o(1/n)$, where $\alpha = \max_y P(y)/Q(y)$. This again illustrates how key secrecy and message secrecy can favor different blanket distributions.

7.2 Analysis of Expected Posterior Variance

We next consider the expected posterior variance, which provides a measure of uncertainty for numerical messages. Rather than evaluating the adversary's probability of exact recovery, this quantity measures the expected uncertainty remaining after observing the shuffled output:

$$\mathrm{EPV}(Y_1; \boldsymbol{Z}) := \mathbb{E}_{\boldsymbol{Z}}\left[\mathrm{Var}(Y_1 \mid \boldsymbol{Z})\right].$$

A larger expected posterior variance indicates greater residual uncertainty about the target message. Appendix E.2 expresses this quantity as expectations involving sums of independent likelihood-ratio weights.

More interestingly, the following theorem characterizes the asymptotic behavior and the optimal blanket under EPV. The proof is provided in Appendix E.3.

Theorem 7.2 (Asymptotically optimal blanket for expected posterior variance)**.** *Consider the canonical shuffle-only model over a finite alphabet $\mathcal{Y} \subset \mathbb{R}$, where $Y_1 \sim P$ is shuffled with $n-1$ independent messages drawn from Q. Assume $P \ll Q$, and let $\mu = \mathbb{E}_P[Y_1]$.*

Then the expected posterior variance satisfies

$$\mathrm{EPV}(Y_1; \boldsymbol{Z}) = \mathrm{Var}_P(Y_1) - \frac{1}{n} \sum_{y \in \mathrm{supp}(P)} \frac{p_y^2 (y-\mu)^2}{q_y} + O(n^{-3/2}), \qquad n \to \infty. \tag{30}$$

Suppose further that P is non-degenerate and $y \neq \mu$ for every $y \in \mathrm{supp}(P)$. Then, among all blanket distributions Q satisfying $P \ll Q$, the leading-order variance reduction is uniquely minimized by

$$q_y^* = \frac{p_y |y-\mu|}{\sum_{z \in \mathrm{supp}(P)} p_z |z-\mu|}, \qquad y \in \mathrm{supp}(P), \tag{31}$$

with zero probability outside $\mathrm{supp}(P)$*.*

Remark 7.1. *If P is non-degenerate but μ belongs to* $\mathrm{supp}(P)$*, the same optimal leading-order coefficient is an infimum over Q satisfying $P \ll Q$, approached by distributions converging to* (31)*.*

Corollary 7.1 (Optimal binary blanket for expected posterior variance)**.** *Consider the canonical shuffle-only model with a non-degenerate target distribution P supported on two distinct real values. Then the asymptotically optimal blanket distribution for maximizing the expected posterior variance is uniform over* $\mathrm{supp}(P)$*, regardless of the probabilities assigned by P.*

Proof. Let $\mathrm{supp}(P) = \{a, b\}$, where $a \neq b$, and write $P(a) = p$ and $P(b) = 1-p$ for some $p \in (0,1)$. Since $\mu = \mathbb{E}_P[Y_1] = pa + (1-p)b$, we have

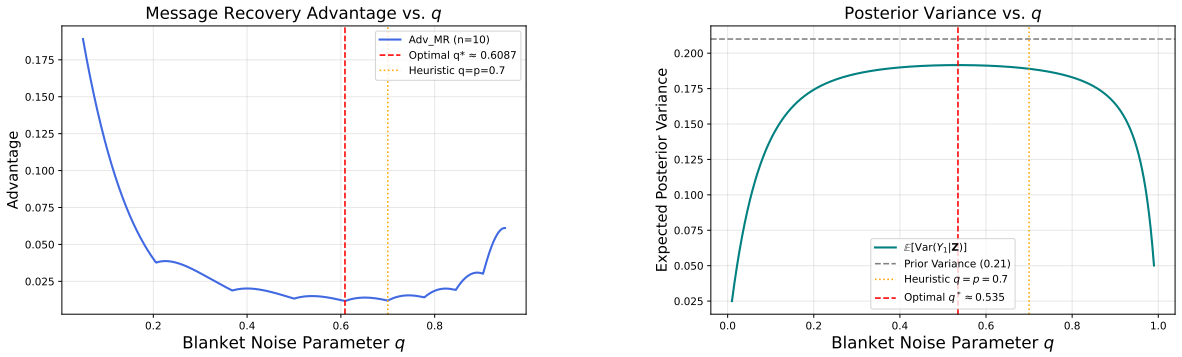
$$p|a-\mu| = p(1-p)|a-b| = (1-p)|b-\mu|.$$

By Theorem 7.2, the asymptotically optimal blanket satisfies $q_y^* \propto p_y |y-\mu|$. Hence $Q^*(a) = Q^*(b) = 1/2$. □

Remark 7.2 (Simultaneous optimality of uniform blankets for binary message distributions)**.** *Combining the results of the preceding sections, for any message distribution supported on two symbols, the uniform blanket is asymptotically optimal under both MI and EPV, while a quasi-uniform blanket is exactly optimal under TV-information at every finite sample size.*

Example 7.1. *To validate the impact of the blanket noise, we consider a binary source $P = (0.3, 0.7)$ with $n = 10$ and evaluate both metrics as functions of the noise parameter q (where $Q = (1-q, q)$). As shown in Fig. 4, the results confirm that the "probability-matching" heuristic ($q = 0.7$) is sub-optimal for both metrics.*

Specifically, the message recovery advantage achieves its minimum at $q \approx 0.6087$, yielding a value of 0.0116 *(compared to* 0.0119 *at $q = 0.7$). Correspondingly, the expected posterior variance attains a numerical maximum of approximately* 0.192 *at $q \approx 0.535$, outperforming the* 0.189 *achieved by the standard heuristic. Moreover, the finite-sample optimal $q^* = 0.535$ for EPV in our numerical search is close to the asymptotic optimal $q^* = 0.5$ for this binary source.*



(a) (b)

Figure 4: Optimal distribution $Q = (1-q, q)$ minimizing MRA and maximizing EPV ($P = (0.3, 0.7)$ on $\{0, 1\}$ and $n = 10$).

8 Conclusion

We developed an information-theoretic framework that characterizes how restrictions on decoy generation constrain achievable secrecy. For fixed finite alphabets, independent decoys have optimal leakage of order n^{-1}, while correlated message-oblivious decoys achieve order n^{-2}, both for fixed non-degenerate priors and under prior-independent minimax design. Message-dependent decoys permit perfect secrecy when $n \geq d$. Even with three candidates, heterogeneity can yield an arbitrarily large multiplicative improvement over i.i.d. decoys.

For canonical shuffling, a unified likelihood-ratio analysis yields exact leakage expressions, explicit finite-sample bounds, and optimal blanket-design principles. The optimal blanket depends on the secrecy metric and generally differs from the message prior, revealing a trade-off between message secrecy and position anonymity.

Finally, key symmetrization connects general perfectly correct symmetric encryption with uniform keys to message-dependent shuffling without increasing plaintext leakage. This connection allows us to apply preimage-constrained channel design to characterize optimal message secrecy and, when the number of keys does not exceed the plaintext support size, the key leakage of every message-leakage-optimal cipher.

References

[AAC+15] Mário S. Alvim, Miguel E. Andrés, Konstantinos Chatzikokolakis, Pierpaolo Degano, and Catuscia Palamidessi. On the information leakage of differentially-private mechanisms. *J. Comput. Secur.*, 23(4):427–469, September 2015.

[ACM+20] Mário S. Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Geoffrey Smith. *The Science of Quantitative Information Flow*. Information Security and Cryptography. Springer, Springer Nature, 2020.

[ACPS12] M'rio S. Alvim, Kostas Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*, pages 265–279, 2012.

[BBGN19] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. The privacy blanket of the shuffle model. In *CRYPTO 2019*, page 638–667, 2019.

[BEM+17] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, page 441–459, 2017.

[BGY+21] Matthieu Bloch, Onur Günlü, Aylin Yener, Frédérique Oggier, H. Vincent Poor, Lalitha Sankar, and Rafael F. Schaefer. An overview of information-theoretic security and privacy: Metrics, limits and applications. *IEEE Journal on Selected Areas in Information Theory*, 2(1):5–22, 2021.

[BK13] Daniel Berend and Aryeh Kontorovich. A sharp estimate of the binomial mean absolute deviation with applications. *Statistics & Probability Letters*, 83(4):1254–1259, 2013.

[BPR00] Mihir Bellare, David Pointcheval, and Phillip Rogaway. Authenticated key exchange secure against dictionary attacks. In *EUROCRYPT 2000*, volume 1807, pages 139–155, 2000.

[BRS02] John Black, Phillip Rogaway, and Thomas Shrimpton. Black-box analysis of the block-cipher-based hash-function constructions from PGV. In *Advances in Cryptology — CRYPTO 2002*, pages 320–335, 2002.

[BS23] Dan Boneh and Victor Shoup. *A Graduate Course in Applied Cryptography*. Online (https://toc.cryptobook.us/), 2023.

[CBJR15] Rahul Chatterjee, Joseph Bonneau, Ari Juels, and Thomas Ristenpart. Cracking-resistant password vaults using natural language encoders. In *IEEE S&P 2015*, pages 481–498, 2015.

[CHY+25] Haibo Cheng, Fugeng Huang, Jiahong Yang, Wenting Li, and Ping Wang. Practically secure honey password vaults: New design and new evaluation against online guessing. In *34th USENIX Security Symposium (USENIX Security 25)*, pages 7781–7798, August 2025.

[CLL+22] Nilesh Chakraborty, Jianqiang Li, Victor C. M. Leung, Samrat Mondal, Yi Pan, Chengwen Luo, and Mithun Mukherjee. Honeyword-based authentication techniques for protecting passwords: A survey. *ACM Comput. Surv.*, 55(8), December 2022.

[CLW+21] Haibo Cheng, Wenting Li, Ping Wang, Chao-Hsien Chu, and Kaitai Liang. Incrementally updateable honey password vaults. In *USENIX Security 2021*, pages 857–874, 2021.

[CPS08] Jean-Sébastien Coron, Jacques Patarin, and Yannick Seurin. The random oracle model and the ideal cipher model are equivalent. In *CRYPTO 2008*, pages 1–20, 2008.

[CSL+26] Haibo Cheng, Haijie Su, Dongyi Li, Wenting Li, and Ping Wang. On optimal information-theoretic security in symmetric encryption under low-entropy keys. Cryptology ePrint Archive, Paper 2026/646, 2026.

[CSU+19] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *EUROCRYPT 2019*, pages 375–403, 2019.

[CY16] Paul Cuff and Lanqing Yu. Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, page 43–54, 2016.

[DAC+15] Markus Dürmuth, Fabian Angelstorf, Claude Castelluccia, Daniele Perito, and Abdelberi Chaabane. OMEN: Faster password guessing using an ordered Markov enumerator. In *International symposium on engineering secure software and systems*, pages 119–132. Springer, 2015.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, page 265–284, 2006.

[dPCF12] Flávio du Pin Calmon and Nadia Fawaz. Privacy against statistical inference. *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1401–1408, 2012.

[DS05] Yevgeniy Dodis and Adam Smith. Entropic security and the encryption of high entropy messages. In *Proceedings of the Second International Conference on Theory of Cryptography*, pages 556–577, 2005.

[DSC09] Josep Domingo-Ferrer, Agusti Solanas, and Jordi Castellà-Roca. h(k)-private information retrieval from privacy-uncooperative queryable databases. *Online Information Review*, 33(4):720–744, 08 2009.

[DZ91] Persi Diaconis and Sandy Zabell. Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre. *Statistical Science*, 6(3):284 – 302, 1991.

[EFM+19] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: from local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, page 2468–2479, 2019.

[Far19] Farhad Farokhi. Taking a lesson from quantum particles for statistical data privacy. arXiv:1908.04954, 2019.

[FMT22] Vitaly Feldman, Audra McMillan, and Kunal Talwar. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 954–964, 2022.

[FS19] Farhad Farokhi and Henrik Sandberg. Ensuring privacy with constrained additive noise by minimizing Fisher information. *Automatica*, 99:275–288, 2019.

[GBD16] Maximilian Golla, Benedict Beuscher, and Markus Dürmuth. On the security of cracking-resistant password vaults. In *ACM CCS 2016*, pages 1230–1241, 2016.

[HJW15] Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Minimax estimation of discrete distributions. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 2291–2295, 2015.

[HMB+21] Hsiang Hsu, Natalia Martinez, Martin Bertran, Guillermo Sapiro, and Flavio P. Calmon. A survey on statistical, information, and estimation—theoretic views on privacy. *IEEE BITS the Information Theory Magazine*, 1(1):45–56, 2021.

[JGAP23] Mireya Jurado, Ramon G. Gonze, Mario S. Alvim, and Catuscia Palamidessi. Analyzing the Shuffle Model Through the Lens of Quantitative Information Flow. In *2023 IEEE 36th Computer Security Foundations Symposium (CSF)*, pages 423–438, July 2023.

[JR13] Ari Juels and Ronald L. Rivest. Honeywords: Making password-cracking detectable. In *ACM CCS 2013*, page 145–160, 2013.

[JR14] Ari Juels and Thomas Ristenpart. Honey encryption: Security beyond the brute-force bound. In *EUROCRYPT 2014*, pages 293–310. Springer, 2014.

[JRT16] Joseph Jaeger, Thomas Ristenpart, and Qiang Tang. Honey encryption beyond message recovery security. In *EUROCRYPT 2016*, pages 758–788, 2016.

[KM16] M.H.R. Khouzani and Pasquale Malacaria. Relative perfect secrecy: Universally optimal strategies and channel design. In *2016 IEEE 29th Computer Security Foundations Symposium (CSF)*, pages 61–76, 2016.

[KM19] M.H.R. Khouzani and Pasquale Malacaria. Generalized entropies and metric-invariant optimal countermeasures for information leakage under symmetric constraints. *IEEE Transactions on Information Theory*, 65(2):888–901, 2019.

[LTZ21] Xinze Li, Qiang Tang, and Zhenfeng Zhang. Fooling an Unbounded Adversary with a Short Key, Repeatedly: The Honey Encryption Perspective. In *2nd Conference on Information-Theoretic Cryptography (ITC 2021)*, volume 199, pages 23:1–23:21, 2021.

[LWH22] Zhengzhi Lin, Yueyao Wang, and Yili Hong. The Poisson Multinomial distribution and its applications in voting theory, ecological inference, and machine learning. arXiv:2201.04237, 2022.

[MF13] Ali Makhdoumi and Nadia Fawaz. Privacy-utility tradeoff under statistical uncertainty. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1627–1634, 2013.

[MUS+16] William Melicher, Blase Ur, Sean M. Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Fast, lean, and accurate: Modeling password guessability using neural networks. In *USENIX Security 2016*, pages 175–191, August 2016.

[NPR13] Silvia Noschese, Lionello Pasquini, and Lothar Reichel. Tridiagonal Toeplitz matrices: properties and novel applications. *Numerical Linear Algebra with Applications*, 20(2):302–326, 2013.

[PW22] Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2022.

[RW06] A. Russell and Hong Wang. How to fool an unbounded adversary with a short key. *IEEE Transactions on Information Theory*, 52(3):1130–1140, 2006.

[SCW26a] Pengcheng Su, Haibo Cheng, and Ping Wang. Bayesian Advantage of Re-Identification Attack in the Shuffle Model. In *2026 IEEE 39th Computer Security Foundations Symposium (CSF)*, pages 49–63, July 2026.

[SCW26b] Pengcheng Su, Haibo Cheng, and Ping Wang. Decomposition-Based Optimal Bounds for Privacy Amplification via Shuffling. In *2026 IEEE Symposium on Security and Privacy (SP)*, pages 2191–2210, May 2026.

[SCW26c] Pengcheng Su, Haibo Cheng, and Ping Wang. Key-space complexity of information-theoretic target-distribution semantic security. Cryptology ePrint Archive, Paper 2026/1899, 2026.

[SCW26d] Pengcheng Su, Haibo Cheng, and Ping Wang. Which privacy blanket is optimal in the shuffle model? Cryptology ePrint Archive, Paper 2026/667, 2026.

[Sha49] C. E. Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28(4):656–715, 1949.

[SXD25] Mayuri Sridhar, Hanshen Xiao, and Srinivas Devadas. PAC-private algorithms. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 3839–3857, 2025.

[UO23] Ayşe Ünsal and Melek Önen. Information-theoretic approaches to differential privacy. *ACM Comput. Surv.*, 56(3), October 2023.

[WAMG09] Matt Weir, Sudhir Aggarwal, Breno de Medeiros, and Bill Glodek. Password cracking using probabilistic context-free grammars. In *IEEE S&P 2009*, pages 391–405, 2009.

[WCW+18] Ding Wang, Haibo Cheng, Ping Wang, Jeff Yan, and Xinyi Huang. A security analysis of honeywords. In *NDSS 2018*, 2018.

[WE18] Isabel Wagner and David Eckhoff. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.*, 51(3), June 2018.

[WZD+22] Ding Wang, Yunkai Zou, Qiying Dong, Yuanming Song, and Xinyi Huang. How to attack and generate honeywords. In *IEEE S&P 2022*, pages 966–983, 2022.

[XD23] Hanshen Xiao and Srinivas Devadas. PAC privacy: Automatic privacy measurement and control of data processing. In *Advances in Cryptology – CRYPTO 2023*, pages 611–644, Cham, 2023.

[XWSD25] Hanshen Xiao, Jun Wan, Elaine Shi, and Srinivas Devadas. One-sided bounded noise: Theory, optimization algorithms and applications. In *Proceedings of the 2025 ACM CCS*, pages 4214–4228, 2025.

A Background

A.1 Mutual-Information-Based Privacy

A fundamental problem in privacy-preserving data analysis is how to *quantify* information leakage (see surveys [WE18, HMB+21, BGY+21, UO23]). Beyond differential privacy [DMNS06],

alternative frameworks have been proposed from the perspectives of *quantitative information flow* (QIF) [ACPS12, AAC+15, ACM+20] and *mutual information* [dPCF12, MF13, CY16].

Mutual information has received renewed attention in recent years, partly because PAC privacy [XD23] provides an operational interpretation of it as a privacy measure and demonstrates its applicability to black-box mechanisms, particularly in machine learning [SXD25, XWSD25]. In essence, the classical data-processing inequality ensures that a bound on $I(Y_1; \boldsymbol{Z})$ controls the expected KL divergence between an adversary's posterior and prior distributions under any inference attack, thereby limiting the expected improvement in attack success probability. Analogous guarantees can be formulated using more general f-divergence measures [PW22].

Whereas classical PAC privacy uses Gaussian noise, the multinomial noise induced by blanket messages in the shuffle model can be viewed as a discrete analogue.

Discussion. To the best of our knowledge, the connection between the shuffle-only model and mutual-information-based privacy has not previously been explicitly characterized. We derive exact identities and asymptotic expansions for mutual information about both the target message and its shuffled position in the canonical setting. We further characterize the blanket distribution Q that minimizes the leading-order message leakage for a given target distribution P, establishing a principled approach to noise design.

A.2 Quantitative Information Flow

Quantitative Information Flow (QIF) aims to characterize the refinement of an adversary's *inferential capability* regarding a secret input upon observing the output of a mechanism. A distinguishing feature of the QIF framework is its explicit dependence on the prior distribution of the input; it defines "privacy leakage" as the gain in the adversary's *optimal inference success rate* (or its generalized utility) [ACPS12, AAC+15]. For a comprehensive treatment and discussion of the QIF framework, we refer the reader to the monograph [ACM+20].

Formal Framework: Channels, Priors, and Gains. Let the secret input be a random variable $X \in \mathbb{X}$ with a prior distribution $\boldsymbol{\pi} \in \mathcal{P}(\mathbb{X})$. A (potentially randomized) mechanism can be abstracted as a *channel* $C : \mathbb{X} \to \mathbb{Y}$, defined by the conditional probabilities $C[y \mid x] = \Pr[Y = y \mid X = x]$, where $Y \in \mathbb{Y}$ is the output observable by the adversary.

To unify various adversarial objectives, we introduce a *gain function* $g : \mathbb{X} \times \mathcal{A} \to \mathbb{R}$, where $\mathcal{A}$ is the set of possible actions available to the adversary. The value $g(x, a)$ represents the gain attained by the adversary when the true secret is x and the chosen action is a.

Prior/Posterior Vulnerability and g-leakage. Given a prior $\boldsymbol{\pi}$ and a gain function g, the adversary's optimal expected gain before observing any output is termed the *prior vulnerability*:

$$V_g(\boldsymbol{\pi}) = \max_{a \in \mathcal{A}} \sum_{x \in \mathbb{X}} \boldsymbol{\pi}(x) g(x, a).$$

Upon observing an output $Y = y$, the adversary's belief is updated to the posterior distribution $\boldsymbol{\pi}(\cdot \mid y)$, yielding an optimal expected gain of $\max_{a \in \mathcal{A}} \sum_x \boldsymbol{\pi}(x \mid y) g(x, a)$. Taking the expectation over all possible outputs $y \in \mathbb{Y}$ yields the *posterior vulnerability*:

$$V_g(\boldsymbol{\pi}, C) = \sum_{y \in \mathbb{Y}} \Pr[Y = y] \cdot \max_{a \in \mathcal{A}} \sum_{x \in \mathbb{X}} \boldsymbol{\pi}(x \mid y) g(x, a).$$

Definition A.1. $\mathcal{L}_g^+(\boldsymbol{\pi}, C) = V_g(\boldsymbol{\pi}, C) - V_g(\boldsymbol{\pi})$ *is the additive g-leakage with respect to* C *under prior distribution* π*.*

This paper focuses on two canonical instantiations of the gain function, catering to different data types:

- **One-try Secret Guessing:** For discrete random variables, we consider the maximum success probability in a single inference attempt. This is formally defined by $\mathcal{A} = \mathbb{X}$ and $g(x, a) = \mathbb{1}[x = a]$. Under this gain function, the posterior vulnerability $V_g(\boldsymbol{\pi}, C)$ corresponds to the *Bayes vulnerability*, representing the adversary's optimal probability of correctly identifying the secret in one shot.
- **Negative L_2 Loss:** For random variables supported on $\mathbb{X} \subseteq \mathbb{R}$, we characterize the adversary's precision in estimating a numerical value. This is modeled by the negative squared error loss, where $\mathcal{A} = \mathbb{R}$ and $g(x, a) = -(x - a)^2$. In this setting, the adversary's optimal action a is the posterior mean $\mathbb{E}[X \mid Y = y]$, and the posterior vulnerability reflects the negative of the expected posterior variance of the secret.

The QIF framework, under the two specific gain functions discussed above, provides the formal definitions and operational interpretations for the privacy metrics adopted in this paper: *message recovery advantage* and *expected posterior variance*:

$$\begin{aligned} \mathrm{Adv}_{\mathrm{MR}}(Y_1; \boldsymbol{Z}) &:= \mathbb{E}_{\boldsymbol{Z}}\left[\max_y \Pr(Y_1 = y \mid \boldsymbol{Z})\right] - \max_y \Pr(Y_1 = y), \\ \mathrm{EPV}(Y_1; \boldsymbol{Z}) &:= \mathbb{E}_{\boldsymbol{Z}}[\mathrm{Var}(Y_1 \mid \boldsymbol{Z})]\,. \end{aligned}$$

Discussion. The study of message recovery advantage (MRA) in the shuffle model was initiated by [JGAP23], which considered the special case of i.i.d. uniformly distributed messages in the shuffle-only model and k-RR-plus-shuffle mechanisms. We extend this analysis to the canonical setting, deriving exact expressions for MRA under arbitrary target and blanket distributions.

To the best of our knowledge, this is the first work to study expected posterior variance in the shuffle model.

A.3 Proof of Property 2.1

Proof. The target position K is uniform on $[n]$. Conditional on $K = k$, the entry Z_k follows P, while the remaining entries are independent samples from Q. Thus, for $\boldsymbol{z} \in \mathrm{supp}(Q)^n$,

$$\Pr[K = k, \boldsymbol{Z} = \boldsymbol{z}] = \frac{1}{n} P(z_k) \prod_{j \neq k} Q(z_j) = \frac{1}{n} r(z_k) Q^{\otimes n}(\boldsymbol{z}).$$

Summing over k gives

$$M_n(\boldsymbol{z}) = Q^{\otimes n}(\boldsymbol{z}) \frac{1}{n} \sum_{k=1}^{n} r(z_k),$$

which proves the likelihood-ratio representation. For a positive-probability output, Bayes' rule then yields

$$\Pr[K = k \mid \boldsymbol{Z} = \boldsymbol{z}] = \frac{r(z_k)}{\sum_{i=1}^{n} r(z_i)}.$$

Finally, $Y_1 = Z_K$, so conditioning on K gives the message-posterior formula. □

A.4 Blanket Decomposition and Reduction

In contrast to the explicit formulas derived in the canonical shuffle-only model, the posteriors of K and Y_1 do not admit the same simple likelihood-ratio formulas when user distributions P_i are heterogeneous. Nevertheless, for any privacy metric satisfying the data processing inequality (including mutual information, TV-information, and QIF measures), the analysis of the heterogeneous shuffle-only setting can be reduced to a tractable homogeneous case via blanket decomposition.

In the *shuffle-only* model, this reduction requires the inputs X_i to be *mutually independent* to ensure the existence of a well-defined post-processing function. In contrast, the *shuffle-DP* model circumvents this requirement by providing $\boldsymbol{X}_{-1}$ as side information to the post-processing step.

In the following, we introduce the blanket decomposition and reduction for the shuffle-only model.

Definition A.2 (Blanket Decomposition [BBGN19])**.** *Let* $\gamma := \sum_{y\in\mathbb{Y}} \inf_{i\in[2,n]} P_i(y)$. *The* blanket distribution Q^{B} *associated with* $(P_2, \ldots, P_n)$ *is*

$$Q^{\mathrm{B}}(y) \;=\; \frac{1}{\gamma} \inf_{i\in[n]-\{1\}} P_i(y).$$

Each P_i *then admits the decomposition*

$$P_i \;=\; \gamma\, Q^{\mathrm{B}} \;+\; (1-\gamma)\,\mathrm{LO}_i,$$

where the left-over distribution LO_i *is given by*

$$\mathrm{LO}_i(y) \;=\; \frac{P_i(y) - \gamma Q^{\mathrm{B}}(y)}{1-\gamma}, \qquad \forall\, y \in \mathbb{Y}.$$

Operationally, sampling from P_i *is equivalent to drawing from* Q^{B} *with probability* γ*, and from* LO_i *otherwise.*

Remark A.1. *When* $\gamma = 0$*,* Q^B *may be chosen arbitrarily, since its mixture weight is zero. Excluding some background users may yield a positive blanket weight; see Example A.1. When* $\gamma = 1$*, all background distributions coincide with* Q^B*, and the left-over distributions are irrelevant, so the reduction is immediate.*

For analytical tractability, we utilize the *generalized blanket distribution* $\bar{Q}^{\mathrm{B}}$ [SCW26b], defined on the augmented space $\bar{\mathbb{Y}} = \mathbb{Y} \cup \{\perp\}$:

$$\bar{Q}^{\mathrm{B}}(y) \;=\; \begin{cases} \inf_{i\in[n]-\{1\}} P_i(y), & y \in \mathbb{Y}, \\ 1-\gamma, & y = \perp\,. \end{cases}$$

Theorem A.1 (Blanket Reduction [SCW26b])**.** *Let* $\boldsymbol{Z}^r = \mathcal{S}\big(\{Y_1\} \cup \{Y_i' : i = 2, \ldots, n\}\big)$*, where* $Y_i' \overset{i.i.d.}{\sim} \bar{Q}^{\mathrm{B}}$ *and* $\mathcal{S}$ *applies a uniformly random permutation. Then there exists a post-processing function* f^{B} *such that*

$$\big(Y_1,\; f^{\mathrm{B}}\big(\boldsymbol{Z}^r, \{\mathrm{LO}_i\}_{i=2}^n\big)\big) \;\overset{d}{=}\; (Y_1,\; \boldsymbol{Z})\,,$$

i.e., the distribution of $\boldsymbol{Z}$ *is obtained by post-processing* $\boldsymbol{Z}^r$ *using* $\{\mathrm{LO}_i\}_{i=2}^n$.

Operationally, given $\boldsymbol{Z}^r$ containing c occurrences of $\perp$, the post-processing f^{B} selects c background users uniformly at random without replacement and replaces those $\perp$ symbols with independent samples from the corresponding left-over distributions LO_i. Full proofs and procedural details appear in [SCW26b].

Property A.1. *The mutual information satisfies* $I(Y_1; \boldsymbol{Z}) \le I(Y_1; \boldsymbol{Z}^r)$.

Proof. By Theorem A.1, given the set $\{\mathrm{LO}_i\}_{i=2}^n$, we have the Markov chain $Y_1 \to \boldsymbol{Z}^r \to \boldsymbol{Z}$. The result follows immediately from the data-processing inequality. □

This reduction is particularly effective when the P_i are similar. If a particular P_i deviates substantially from the rest, one may exclude that user from the analysis: the output $\boldsymbol{Z}$ without that user's contribution can be post-processed to reconstruct the full $\boldsymbol{Z}$ by re-inserting that user's message (at a uniformly random position).

Example A.1. *Suppose* $P_1 = (0.4, 0.6)$, $P_2 = (0.3, 0.7)$, $P_3 = (0.5, 0.5)$, *and* $P_4 = (1.0, 0)$. *User 4's distribution differs markedly from the others, so we may exclude it. The generalized blanket distribution of* P_2 *and* P_3 *is* $\bar{Q}^{\mathrm{B}} = (0.3, 0.5, 0.2)$, *and*

$$I(Y_1; \boldsymbol{Z}) \;\le\; I(Y_1; \boldsymbol{Z}^r),$$

where $\boldsymbol{Z}^r = \mathcal{S}\big(\{Y_1\} \cup \{Y_i' : i = 2, 3\}\big)$ *and* $Y_i' \overset{i.i.d.}{\sim} \bar{Q}^{\mathrm{B}}$.

B Missing Proofs in Section 3

B.1 Proof of Theorem 3.1

Proof. Identify $\mathrm{supp}(P)$ with $[d]$. We may restrict the decoy alphabet to $[d]$ without loss: mapping every symbol outside $[d]$ to a fixed symbol in $[d]$ preserves the decoy-generation constraints and cannot increase leakage. Let T denote the output histogram. Conditional on T, the uniformly shuffled ordering is independent of M, so

$$I(M; \boldsymbol{Z}) = I(M; T).$$

We first prove (4). Fix $E \subseteq [d]$ with $q := P(E) \in (0, 1)$, and set $B = \mathbf{1}_{\{M \in E\}}$. Let $\mu_b = \mathrm{Law}(T \mid B = b)$, and let $N_E(T)$ count the output messages in E. Independence of the decoys from M gives

$$\mathbb{E}_{\mu_1} N_E - \mathbb{E}_{\mu_0} N_E = 1.$$

Since $0 \le N_E \le n$, $D_{\mathrm{TV}}(\mu_1, \mu_0) \ge 1/n$. Writing $\mu = q\mu_1 + (1-q)\mu_0$, data processing and Pinsker's inequality yield

$$\begin{aligned} I(M; T) \ge I(B; T) &= qD_{\mathrm{KL}}(\mu_1 \| \mu) + (1-q)D_{\mathrm{KL}}(\mu_0 \| \mu) \\ &\ge 2q(1-q)D_{\mathrm{TV}}(\mu_1, \mu_0)^2 \\ &\ge \frac{2q(1-q)}{n^2}. \end{aligned}$$

Maximizing over E proves the claimed lower bound.

If $n \ge d$, choose a fixed multiset of n messages containing every symbol in $[d]$. Given $M = i$, generate the decoys by removing one occurrence of i from this multiset. The output histogram is then independent of M, giving $L_{\mathrm{dep}}(P, n) = 0$. Strict separation follows from (4).

Now suppose $2 \le n < d$. The Khouzani–Malacaria characterization [KM16, KM19] recalled in Section 4.2 gives $L_{\mathrm{dep}}(P, n) = H(P) - H(\boldsymbol{\pi}^*)$. Its equality conditions require every output of positive probability to induce a posterior supported on exactly n messages. In particular, an optimal output must contain n distinct messages.

This condition cannot hold for a message-oblivious mechanism. Indeed, let D_1 be any one of its decoys. Since D_1 is independent of M,

$$\Pr(M = D_1) = \sum_{i=1}^{d} p_i \Pr(D_1 = i) \geq \min_{i \in [d]} p_i > 0.$$

On this collision event, the output contains at most $n-1$ distinct messages, and its posterior therefore has support size at most $n-1$. Thus no message-oblivious mechanism attains the dependent optimum.

Finally, joint decoy distributions on the finite alphabet $[d]^{n-1}$ form a compact probability simplex, and mutual information is continuous in the joint decoy distribution. Hence the message-oblivious minimum is attained. Since equality with the dependent optimum is impossible, the two optimal values are strictly separated. □

B.2 Proof of Lemma 3.1

Proof. Write $\mu = \mathbb{E}S$ and set $a = \max\{1, 2\sigma\}$. By Chebyshev's inequality,

$$\Pr\big(|S - \mu| \leq a\big) \geq 1 - \frac{\sigma^2}{a^2} \geq \frac{3}{4}.$$

Define the nondecreasing function $f : \mathbb{R} \to [0, 1]$ by

$$f(t) = \min\left\{1,\, \max\left\{0,\, \frac{t - (\mu - a - 1)}{2a + 2}\right\}\right\}.$$

On the event $\{|S - \mu| \leq a\}$, both S and $S + 1$ lie in the linear segment of f, and hence

$$f(S + 1) - f(S) = \frac{1}{2a + 2}.$$

Outside this event, the difference remains nonnegative because f is nondecreasing. The variational characterization of total variation therefore gives

$$D_{\mathrm{TV}}\big(\mathrm{Law}(S), \mathrm{Law}(S + 1)\big) \geq \mathbb{E}\big[f(S + 1) - f(S)\big] \geq \frac{3}{4(2a + 2)} = \frac{3}{8(a + 1)} \geq \frac{3}{16(\sigma + 1)},$$

where the last inequality follows from $a \leq 2\sigma + 1$. □

B.3 Proof of Theorem 3.2

Proof. Let $B = \mathbf{1}_{\{M=y\}} \sim \mathrm{Bern}(p)$. The y-th coordinate of the output satisfies

$$T_y = B + S_y.$$

Since B is a deterministic function of M and T_y is a deterministic function of T, the data-processing inequality yields

$$I(M; T) \geq I(B; T_y).$$

Because S_y is independent of B, the conditional output laws are

$$W_0 = \mathrm{Law}(S_y), \qquad W_1 = \mathrm{Law}(S_y + 1),$$

and their marginal mixture is

$$W = (1-p)W_0 + pW_1.$$

By Pinsker's inequality,

$$\begin{aligned} I(B;T_y) &= (1-p)D_{\mathrm{KL}}(W_0\|W) + pD_{\mathrm{KL}}(W_1\|W) \\ &\geq 2(1-p)D_{\mathrm{TV}}(W_0,W)^2 + 2pD_{\mathrm{TV}}(W_1,W)^2 \\ &= 2p(1-p)D_{\mathrm{TV}}(W_0,W_1)^2, \end{aligned}$$

where the last equality uses $D_{\mathrm{TV}}(W_0,W) = pD_{\mathrm{TV}}(W_0,W_1), \quad D_{\mathrm{TV}}(W_1,W) = (1-p)D_{\mathrm{TV}}(W_0,W_1)$. Applying Lemma 3.1 gives

$$I(M;T) \geq 2p(1-p)\left(\frac{3}{16(\sigma_y+1)}\right)^2 = \frac{9p(1-p)}{128(\sigma_y+1)^2}. \qquad \square$$

B.4 Proof of Theorem 3.3

Proof. Fix y with $p_y \in (0,1)$. Let $B = \mathbf{1}_{\{M=y\}}$ and $B_i = \mathbf{1}_{\{H_i=y\}}$ for $i = 2,\ldots,n$. The decoy count in coordinate y is

$$S_y = \sum_{i=2}^{n} B_i, \qquad B_i \sim \mathrm{Bern}(Q_i(y)),$$

where the B_i are mutually independent and jointly independent of M. Therefore,

$$\sigma_y^2 := \mathrm{Var}(S_y) = \sum_{i=2}^{n} Q_i(y)(1-Q_i(y)) \leq \frac{n-1}{4}.$$

Applying Theorem 3.2 gives

$$I(M;T) \geq \frac{9p_y(1-p_y)}{128(\sigma_y+1)^2} \geq \frac{9p_y(1-p_y)}{128(1+\sqrt{n-1}/2)^2} = \frac{9p_y(1-p_y)}{32(\sqrt{n-1}+2)^2}.$$

Since this bound is uniform over all decoy distributions and P is fixed and non-degenerate, it follows that

$$L_{\mathrm{het}}(P,n) = \Omega_P(n^{-1}), \qquad L_{\mathrm{can}}(P,n) = \Omega_P(n^{-1}).$$

For the matching upper bound, choose $Q_i = P$ for every $i = 2,\ldots,n$. This yields a canonical construction with i.i.d. decoys distributed according to the message prior. The canonical mutual-information expansion (Theorem 5.2) gives

$$I(M;T) = \frac{d-1}{2n} + O_P(n^{-3/2}),$$

where $d = |\mathrm{supp}(P)|$. Because this construction belongs to the heterogeneous class,

$$L_{\mathrm{het}}(P,n) = O_P(n^{-1}), \qquad L_{\mathrm{can}}(P,n) = O_P(n^{-1}).$$

Combining the upper and lower bounds proves (6). $\square$

B.5 Proof of Lemma 3.2

Proof. Since the mixture W dominates every W_i, the inequality $D_{\mathrm{KL}}(U\|V) \le \chi^2(U\|V)$ gives

$$\begin{aligned} I(M;T) &= \sum_{i=1}^{d} p_i D_{\mathrm{KL}}(W_i\|W) \\ &\le \sum_{i=1}^{d} p_i \sum_{t:W(t)>0} \frac{(W_i(t)-W(t))^2}{W(t)} \\ &= \sum_{1\le i<j\le d} p_i p_j \sum_{t:W(t)>0} \frac{(W_i(t)-W_j(t))^2}{W(t)}. \end{aligned}$$

The last equality is the pairwise-difference identity for weighted variance.

Fix $i<j$ and an output t, and write $a=W_i(t)$ and $b=W_j(t)$. When $a+b>0$, we have

$$\frac{(a-b)^2}{W(t)} \le \frac{(\sqrt{a}-\sqrt{b})^2(\sqrt{a}+\sqrt{b})^2}{p_i a + p_j b} \le \frac{2}{\min\{p_i,p_j\}}(\sqrt{a}-\sqrt{b})^2,$$

where we used

$$(\sqrt{a}+\sqrt{b})^2 \le 2(a+b), \qquad p_i a + p_j b \ge \min\{p_i,p_j\}(a+b).$$

If $a=b=0$, the corresponding contribution is zero. Multiplying by $p_i p_j$ and summing over all outputs and pairs proves (7). □

B.6 Proof of Lemma 3.3

Proof. Set

$$v_j = \sqrt{\frac{2}{L+1}}\sin((j+1)\theta), \qquad j=0,\dots,L-1,$$

and extend v_j by zero outside this interval. The identity

$$\sum_{j=0}^{L-1} \sin^2((j+1)\theta) = \frac{L+1}{2}$$

gives $\sum_j v_j^2 = 1$. Moreover, the trigonometric recurrence yields

$$v_{j-1} + v_{j+1} = 2\cos\theta\, v_j, \qquad j=0,\dots,L-1,$$

with boundary values $v_{-1}=v_L=0$. Multiplying by v_j and summing gives

$$2\sum_{j\in\mathbb{Z}} v_j v_{j-1} = 2\cos\theta \sum_{j\in\mathbb{Z}} v_j^2 = 2\cos\theta.$$

Since $\sqrt{r(j)} = v_j$ and $\sqrt{r^+(j)} = v_{j-1}$, this proves the affinity identity. Both square-root vectors have unit norm, so

$$\left\|\sqrt{r} - \sqrt{r^+}\right\|_2^2 = 2 - 2\sum_{j\in\mathbb{Z}} \sqrt{r(j)r^+(j)} = 2 - 2\cos\theta.$$ □

B.7 Proof of Theorem 3.4

Proof. Consider the first $d-1$ coordinates of T. Conditional on $M=d$, their joint distribution is $r^{\otimes(d-1)}$. Conditional on $M=i<d$, only coordinate i is shifted by one unit, replacing its distribution r with r^{+}. The final coordinate is uniquely determined by the first $d-1$ coordinates because the total count is n. Consequently, retaining or omitting that coordinate does not change the pairwise square-root distances.

Write $W_i = \mathrm{Law}(T \mid M=i)$ and $c=\cos\theta$. By Lemma 3.3 and the multiplicativity of square-root affinity for product distributions,

$$\sum_t \sqrt{W_i(t)W_d(t)} = c, \qquad i<d,$$

and

$$\sum_t \sqrt{W_i(t)W_j(t)} = c^2, \qquad i,j<d, \quad i\neq j.$$

It follows that

$$\left\|\sqrt{W_i}-\sqrt{W_d}\right\|_2^2 = 2(1-c), \qquad i<d,$$

whereas

$$\left\|\sqrt{W_i}-\sqrt{W_j}\right\|_2^2 = 2(1-c^2) \le 4(1-c), \qquad i,j<d, \quad i\neq j.$$

Thus every pair satisfies

$$\left\|\sqrt{W_i}-\sqrt{W_j}\right\|_2^2 \le 4(1-\cos\theta) \le 2\theta^2.$$

Applying Lemma 3.2 gives

$$\begin{aligned} I(M;T) &\le 2\sum_{i<j}\max\{p_i,p_j\}\left\|\sqrt{W_i}-\sqrt{W_j}\right\|_2^2 \\ &\le 4\theta^2\sum_{i<j}\max\{p_i,p_j\} \le 4\theta^2\binom{d}{2} = \frac{2d(d-1)\pi^2}{(L+1)^2}. \end{aligned}$$

If some prior probabilities vanish, apply the lemma after restricting to the positive-probability inputs; the same upper bound remains valid. For fixed d, we have $L=\Theta_d(n)$, proving (10). □

B.8 Proof of Theorem 3.5

Proof. Let $M\sim\mathrm{Bern}(p)$. For two independent decoys $B_1\sim\mathrm{Bern}(q_1)$ and $B_2\sim\mathrm{Bern}(q_2)$, the shuffled histogram is determined by $T=M+S$, where $S=B_1+B_2$. Write $s_j=\Pr(S=j)$ and $L=\log(1/p)$.

We first establish a useful upper bound. For $0<p\le 1/2$ and $s_1,s_2>0$,

$$I(M;T) \le p\left(2+\log\frac{1}{s_1}+s_1\log\frac{1}{s_2}+s_2L\right). \tag{32}$$

Indeed, the conditional output distributions are

$$W_0=(s_0,s_1,s_2,0), \qquad W_1=(0,s_0,s_1,s_2).$$

Choose the reference distribution

$$R=((1-p)s_0,(1-p)s_1,(1-p)s_2,p),$$

which is normalized because $s_0 + s_1 + s_2 = 1$.

Writing $W = (1-p)W_0 + pW_1$, the relative-entropy chain rule gives

$$I(M;T) = (1-p)D_{\mathrm{KL}}(W_0\|R) + pD_{\mathrm{KL}}(W_1\|R) - D_{\mathrm{KL}}(W\|R) \le (1-p)D_{\mathrm{KL}}(W_0\|R) + pD_{\mathrm{KL}}(W_1\|R).$$

The two divergences satisfy

$$D_{\mathrm{KL}}(W_0\|R) = \log\frac{1}{1-p}$$

and

$$\begin{aligned} D_{\mathrm{KL}}(W_1\|R) &= s_0 \log\frac{s_0}{(1-p)s_1} + s_1\log\frac{s_1}{(1-p)s_2} + s_2\log\frac{s_2}{p} \\ &= (1-s_2)\log\frac{1}{1-p} + \sum_{j=0}^{2} s_j \log s_j + s_0\log\frac{1}{s_1} + s_1\log\frac{1}{s_2} + s_2 L, \end{aligned}$$

where $L = \log(1/p)$ and $0\log 0 := 0$. Consequently,

$$\begin{aligned} I(M;T) &\le (1-ps_2)\log\frac{1}{1-p} + p\left(\sum_{j=0}^{2} s_j\log s_j + s_0\log\frac{1}{s_1} + s_1\log\frac{1}{s_2} + s_2 L\right) \\ &\le \log\frac{1}{1-p} + p\left(\log\frac{1}{s_1} + s_1\log\frac{1}{s_2} + s_2 L\right) \\ &\le p\left(2 + \log\frac{1}{s_1} + s_1\log\frac{1}{s_2} + s_2 L\right). \end{aligned}$$

Here the second inequality uses $\sum_j s_j \log s_j \le 0$, $s_0 \le 1$, and $s_1 \le 1$; the last uses $-\log(1-p) \le 2p$ for $0 < p \le 1/2$.

Heterogeneous upper bound. Choose $q_1 = \frac{1}{L}$, $\quad q_2 = \frac{1}{\log L}$. Then

$$s_1 = q_1 + q_2 - 2q_1q_2 = \Theta((\log L)^{-1}), \qquad s_2 = q_1q_2 = \frac{1}{L\log L}.$$

By (32),

$$I(M;T) \le p\left(2 + \log\frac{1}{s_1} + s_1\log\frac{1}{s_2} + s_2 L\right) = O(p\log\log L).$$

Therefore,

$$L_{\mathrm{het}}(P_p, 3) = O(p\log\log L).$$

Canonical lower bound. Now suppose $q_1 = q_2 = q$, so that $S \sim \mathrm{Bin}(2,q)$. Fix $\delta \in (0, 1/2)$ and set $a = L^{-1/2+\delta}$.

If $q \ge a$, the event $T = 3$ occurs with probability pq^2 and uniquely identifies $M = 1$. Using the posterior representation of mutual information,

$$I(M;T) = \mathbb{E}_T D_{\mathrm{KL}}(P_{M|T}\|P_M) \ge pq^2 L \ge pL^{2\delta}.$$

Otherwise, $q < a$. Let

$$u = \Pr(T = 1 \mid M = 1) = (1-q)^2$$

and

$$v = \Pr(T = 1) = (1-p)2q(1-q) + p(1-q)^2.$$

Then

$$u \geq (1-a)^2, \qquad v \leq 2a + p \leq 3a$$

for sufficiently small p. By data processing for KL divergence,

$$I(M;T) \geq pD_{\mathrm{KL}}(P_{T|M=1} \| P_T) \geq p\, d_{\mathrm{KL}}(u\|v) \geq p\left(u \log \frac{1}{v} - h(u)\right) \geq \left(\frac{1}{2} - \delta - o(1)\right) p \log L,$$

where h denotes binary entropy. Since $L^{2\delta} \gg \log L$, both cases give the latter lower bound. As $\delta > 0$ is arbitrary,

$$L_{\mathrm{can}}(P_p, 3) \geq \left(\frac{1}{2} - o(1)\right) p \log L.$$

Canonical upper bound. Conversely, choosing $q = L^{-1/2}$ gives

$$s_1 = 2q(1-q), \qquad s_2 = L^{-1}.$$

Applying (32) again,

$$L_{\mathrm{can}}(P_p, 3) \leq I(M;T) \leq p\left(\frac{1}{2}\log L + O(1)\right).$$

Consequently,

$$L_{\mathrm{can}}(P_p, 3) = \left(\frac{1}{2} + o(1)\right) p \log L.$$

Combining the canonical lower bound with the heterogeneous upper bound yields

$$\frac{L_{\mathrm{het}}(P_p, 3)}{L_{\mathrm{can}}(P_p, 3)} = O\left(\frac{\log\log L}{\log L}\right) \longrightarrow 0.$$

This completes the proof. □

B.9 Prior-Independent Analysis of the Decoy Hierarchy

Theorem B.1 (Prior-independent optimal leakage rates)**.** *For every fixed alphabet size $d \geq 2$, as $n \to \infty$,*

$$\begin{aligned} \mathfrak{L}_{\mathrm{can}}(d,n) &= \Theta_d(n^{-1}), \\ \mathfrak{L}_{\mathrm{het}}(d,n) &= \Theta_d(n^{-1}), \\ \mathfrak{L}_{\mathrm{obl}}(d,n) &= \Theta_d(n^{-2}). \end{aligned} \tag{33}$$

For integers $d \geq 2$ and $n \geq 2$,

$$\mathfrak{L}_{\mathrm{dep}}(d,n) = [\log(d/n)]_+, \qquad [x]_+ := \max\{x, 0\}.$$

Proof.

A uniform lower bound for independent decoys. Fix two distinct messages $a, b \in [d]$, and choose the prior

$$P(a) = P(b) = \frac{1}{2}, \qquad P(m) = 0 \quad \text{for } m \notin \{a, b\}.$$

Consider any heterogeneous independent mechanism. Let S_a be the number of decoys equal to a. Since S_a is a sum of $n-1$ independent Bernoulli variables,

$$\mathrm{Var}(S_a) \le \frac{n-1}{4}.$$

Writing $B = \mathbf{1}_{\{M=a\}}$, the corresponding output count is $T_a = B + S_a$, with $B \perp S_a$. By (5),

$$I_P(M; Z) = I_P(M; T) \ge I(B; T_a) \ge \frac{9}{128(\sqrt{n-1}+2)^2} = \Omega(n^{-1}).$$

The same prior and lower bound apply to every heterogeneous independent mechanism. Thus, taking the supremum over priors and then the infimum over mechanisms preserves the bound:

$$\mathfrak{L}_{\mathrm{het}}(d, n) = \Omega(n^{-1}).$$

Since canonical mechanisms form a subclass of heterogeneous mechanisms,

$$\mathfrak{L}_{\mathrm{can}}(d, n) \ge \mathfrak{L}_{\mathrm{het}}(d, n),$$

so the same lower bound holds for canonical shuffling.

A uniform upper bound for canonical shuffling. Choose the uniform blanket $Q = U_d$, independently of the prior P. Let $R = Q^{\otimes n}$ be the reference distribution, and write $W_m = \mathrm{Law}(Z \mid M = m)$. The likelihood-ratio representation gives

$$\frac{dW_m}{dR}(z) = \frac{N_m(z)}{nq_m}, \qquad q_m = \frac{1}{d},$$

where $N_m(z)$ counts the occurrences of m in z. Under R, we have $N_m \sim \mathrm{Bin}(n, q_m)$, and hence

$$\chi^2(W_m \| R) = \mathbb{E}_R\left[\left(\frac{N_m}{nq_m} - 1\right)^2\right] = \frac{1 - q_m}{nq_m} = \frac{d-1}{n}.$$

For any prior $P = (p_1, \ldots, p_d) \in \Delta_d$, let $W_P = \sum_m p_m W_m$ denote the marginal output distribution. The relative-entropy chain rule and the inequality $D_{\mathrm{KL}} \le \chi^2$ yield

$$\begin{aligned} I_P(M; Z) &= \sum_{m=1}^{d} p_m D_{\mathrm{KL}}(W_m \| R) - D_{\mathrm{KL}}(W_P \| R) \\ &\le \sum_{m=1}^{d} p_m D_{\mathrm{KL}}(W_m \| R) \\ &\le \sum_{m=1}^{d} p_m \chi^2(W_m \| R) = \frac{d-1}{n}. \end{aligned}$$

Thus the same canonical mechanism satisfies

$$\sup_{P \in \Delta_d} I_P(M; Z) \le \frac{d-1}{n},$$

which proves $\mathfrak{L}_{\mathrm{can}}(d,n) = O_d(n^{-1})$. Since canonical mechanisms form a subclass of heterogeneous independent mechanisms,

$$\mathfrak{L}_{\mathrm{het}}(d,n) \le \mathfrak{L}_{\mathrm{can}}(d,n),$$

so the same upper bound holds for heterogeneous shuffling. Together with the preceding lower bounds, this establishes

$$\mathfrak{L}_{\mathrm{can}}(d,n) = \Theta_d(n^{-1}), \qquad \mathfrak{L}_{\mathrm{het}}(d,n) = \Theta_d(n^{-1}).$$

A uniform upper bound using correlated noise. Use the construction in (8). The proof of Theorem 3.4 yields the uniform bound

$$\sup_{P \in \Delta_d} I_P(M;Z) \le \frac{2d(d-1)\pi^2}{(L+1)^2} = O_d(n^{-2}).$$

The constant is uniform over P. It follows that

$$\mathfrak{L}_{\mathrm{obl}}(d,n) = O_d(n^{-2}).$$

A uniform lower bound for message-oblivious mechanisms. Choose again the prior uniform on the fixed pair $\{a,b\}$. The general message-oblivious lower bound from Theorem 3.1 gives, for every message-oblivious mechanism,

$$I_P(M;Z) \ge \frac{2(1/2)(1/2)}{n^2} = \frac{1}{2n^2}.$$

Taking the supremum over priors and then the infimum over mechanisms therefore yields

$$\mathfrak{L}_{\mathrm{obl}}(d,n) \ge \frac{1}{2n^2}.$$

This proves the third rate characterization.

Exact characterization of message-dependent decoys. For the lower bound, take the uniform prior $P = U_d$. For any mechanism, the genuine message must occur among the n output entries, so every posterior has support size at most $\min\{d,n\}$. Consequently,

$$I_{U_d}(M;Z) = \log d - H(M \mid Z) \ge \log d - \log \min\{d,n\} = [\log(d/n)]_+.$$

This holds for every message-dependent mechanism and therefore lower-bounds the minimax leakage.

For the matching upper bound when $n < d$, given $M = m$, choose an n-element set $A \subseteq [d]$ uniformly among those containing m, and release a uniformly random ordering Z of A. This mechanism is independent of the prior. Since A is recoverable from Z and the ordering is independent of M conditional on A, for every prior P,

$$\begin{aligned} I_P(M;Z) &= I_P(M;A) \\ &= H_P(A) - H_P(A \mid M) \\ &\le \log \binom{d}{n} - \log \binom{d-1}{n-1} = \log(d/n). \end{aligned}$$

When $n \ge d$, fix a multiset of size n containing every symbol in $[d]$. On input m, use the multiset with one occurrence of m removed as the decoys. The released sequence is a uniformly random ordering of the same fixed multiset, independently of m, and hence has zero leakage for every prior. □

C Missing Proofs in Section 5

C.1 Proof of Lemma C.1

Lemma C.1 (Global quadratic remainder)**.** *For every $u \geq -1$, $f(x) = x \log x$ (with convention $f(0) = 0$), we have*

$$\left| f(1+u) - u - \frac{u^2}{2} \right| \leq \frac{|u|^3}{2}. \tag{34}$$

Proof. For $u \geq 0$, we have $f(1) = 0$, $f'(1) = f''(1) = 1$, and $f'''(t) = -t^{-2}$ for $t > 0$. Taylor's formula with integral remainder gives

$$f(1+u) - u - \frac{u^2}{2} = -\frac{1}{2} \int_0^u \frac{(u-t)^2}{(1+t)^2}\, dt.$$

Its absolute value is at most $\frac{1}{2}\int_0^u (u-t)^2\, dt = u^3/6$.

For $u = -v$, where $0 \leq v < 1$, the logarithm series yields

$$f(1-v) = -v + \sum_{k=2}^{\infty} \frac{v^k}{k(k-1)}.$$

Consequently,

$$0 \leq f(1-v) + v - \frac{v^2}{2} = \sum_{k=3}^{\infty} \frac{v^k}{k(k-1)} \leq v^3 \sum_{k=3}^{\infty} \frac{1}{k(k-1)} = \frac{v^3}{2}.$$

Continuity at $v = 1$ completes the proof. □

C.2 Proof of Theorem 5.1

Proof. Let $\boldsymbol{Z}^* \sim Q^{\otimes n}$ and define the likelihood ratio $r(y) = \frac{P(y)}{Q(y)}, y \in \operatorname{supp}(Q)$.

Since K is uniform over $[n]$, conditioned on $K = i$, the i-th coordinate of $\boldsymbol{Z}$ follows P, while all remaining coordinates are i.i.d. from Q. Consequently,

$$D_{\mathrm{KL}}\big(\mathsf{Law}(K, \boldsymbol{Z}) \,\|\, U_{[n]} \otimes Q^{\otimes n}\big) = D_{\mathrm{KL}}(P \| Q),$$

where $U_{[n]}$ denotes the uniform distribution on $[n]$.

On the other hand, the chain rule for relative entropy gives

$$D_{\mathrm{KL}}\big(\mathsf{Law}(K, \boldsymbol{Z}) \,\|\, U_{[n]} \otimes Q^{\otimes n}\big) = I(K; \boldsymbol{Z}) + D_{\mathrm{KL}}(M_n \| Q^{\otimes n}).$$

Combining these identities proves the first identity.

It remains to estimate the second term. By Property 2.1, the density of M_n relative to $Q^{\otimes n}$ is

$$L(\boldsymbol{z}) = \frac{1}{n} \sum_{i=1}^{n} r(z_i).$$

Writing $f(t) = t \log t$ with $f(0) = 0$, we obtain

$$D_{\mathrm{KL}}(M_n \| Q^{\otimes n}) = \mathbb{E}_{\boldsymbol{Z}^* \sim Q^{\otimes n}}[f(L(\boldsymbol{Z}^*))].$$

Let

$$U_i = r(Z_i^*) - 1, \qquad \overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i.$$

Then the U_i are i.i.d., with $\mathbb{E}[U_i] = 0,\ \mathbb{E}[U_i^2] = \chi^2(P\|Q)$. Hence,

$$\mathbb{E}[\overline{U}] = 0, \qquad \mathbb{E}[\overline{U}^2] = \frac{\chi^2(P\|Q)}{n}.$$

Lemma C.1, proved in Appendix C.1, gives

$$\left| f(1+u) - u - \frac{u^2}{2} \right| \le \frac{|u|^3}{2} \qquad \text{for all } u \ge -1.$$

Applying this bound with $u = \overline{U}$ yields

$$\left| D_{\mathrm{KL}}(M_n\|Q^{\otimes n}) - \frac{\chi^2(P\|Q)}{2n} \right| \le \frac{1}{2}\mathbb{E}[|\overline{U}|^3].$$

Since the alphabet is finite and $P \ll Q$, the likelihood ratio r is bounded. In particular, letting $C = \max_{y\in\mathrm{supp}(Q)} |r(y) - 1|$, we have $|U_i| \le C$. By independence and centering,

$$\mathbb{E}[\overline{U}^4] = \frac{n\mathbb{E}[U_1^4] + 3n(n-1)(\mathbb{E}[U_1^2])^2}{n^4} \le \frac{C^4(3n-2)}{n^3}.$$

Therefore, by Cauchy–Schwarz,

$$\mathbb{E}[|\overline{U}|^3] \le \sqrt{\mathbb{E}[\overline{U}^2]\mathbb{E}[\overline{U}^4]} \le \frac{C^3\sqrt{3n-2}}{n^2} = O(n^{-3/2}).$$

The proof is done. □

C.3 Proof of Theorem 5.2

Proof. Let $\boldsymbol{Z}^* \sim Q^{\otimes n}$ and define

$$r_y(z) = \frac{\mathbb{1}[z = y]}{Q(z)}, \qquad r_P(z) = \frac{P(z)}{Q(z)}.$$

For each $y \in \mathrm{supp}(P)$, let

$$L_y(\boldsymbol{z}) = \frac{1}{n}\sum_{j=1}^{n} r_y(z_j), \qquad L_P(\boldsymbol{z}) = \frac{1}{n}\sum_{j=1}^{n} r_P(z_j).$$

These are the likelihood ratios of $\mathsf{Law}(\boldsymbol{Z} \mid Y_1 = y)$ and $\mathsf{Law}(\boldsymbol{Z})$, respectively, relative to $Q^{\otimes n}$.

By the chain rule for relative entropy,

$$\begin{aligned} I(Y_1; \boldsymbol{Z}) &= \mathbb{E}_{y\sim P} D_{\mathrm{KL}}\big(\mathsf{Law}(\boldsymbol{Z} \mid Y_1 = y)\|Q^{\otimes n}\big) - D_{\mathrm{KL}}(M_n\|Q^{\otimes n}) \\ &= \mathbb{E}_{Q^{\otimes n}}\left[\sum_y p_y f(L_y) - f(L_P)\right], \end{aligned} \tag{35}$$

where $M_n = \mathsf{Law}(\boldsymbol{Z})$ and $f(t) = t\log t$, with $f(0) = 0$.

We first estimate the two terms on the right-hand side. For any distribution $A \ll Q$, define

$$L_A(\boldsymbol{Z}^*) = \frac{1}{n}\sum_{j=1}^{n} \frac{A(Z_j^*)}{Q(Z_j^*)}.$$

Writing

$$U_j = \frac{A(Z_j^*)}{Q(Z_j^*)} - 1, \qquad \overline{U} = \frac{1}{n}\sum_{j=1}^{n} U_j,$$

we have $\mathbb{E}[U_j] = 0$ and $\mathbb{E}[\overline{U}^2] = \frac{\chi^2(A\|Q)}{n}$.

Let

$$C_A = \max_{z \in \mathrm{supp}(Q)} \left| \frac{A(z)}{Q(z)} - 1 \right|.$$

Since the alphabet is finite, $C_A < \infty$. Lemma C.1 together with the second- and fourth-moment identities for $\overline{U}$, yields

$$\left| \mathbb{E}_{Q^{\otimes n}}[f(L_A)] - \frac{\chi^2(A\|Q)}{2n} \right| \le \frac{C_A^3\sqrt{3n-2}}{2n^2}. \tag{36}$$

Applying (36) to $A = \delta_y$ for every $y \in \mathrm{supp}(P)$, and then to $A = P$, gives

$$I(Y_1; \boldsymbol{Z}) = \frac{1}{2n}\left(\sum_y p_y \chi^2(\delta_y\|Q) - \chi^2(P\|Q) \right) + O(n^{-3/2}).$$

Since

$$\chi^2(\delta_y\|Q) = \frac{1}{q_y} - 1, \qquad \chi^2(P\|Q) = \sum_y \frac{p_y^2}{q_y} - 1,$$

we obtain

$$\sum_y p_y \chi^2(\delta_y\|Q) - \chi^2(P\|Q) = \sum_y p_y \left(\frac{1}{q_y} - 1 \right) - \left(\sum_y \frac{p_y^2}{q_y} - 1 \right) = \sum_y \frac{p_y(1-p_y)}{q_y}.$$

Therefore,

$$I(Y_1; \boldsymbol{Z}) = \frac{1}{2n}\sum_y \frac{p_y(1-p_y)}{q_y} + O(n^{-3/2}).$$

More precisely, letting $C_y = \max\left\{1, \frac{1}{q_y} - 1\right\}$, $C_P = \max_{z \in \mathrm{supp}(Q)} \left| \frac{p_z}{q_z} - 1 \right|$, the same argument used in the proof of Theorem 5.1 yields the explicit bound

$$\left| I(Y_1; \boldsymbol{Z}) - \frac{1}{2n}\sum_y \frac{p_y(1-p_y)}{q_y} \right| \le \frac{\sqrt{3n-2}}{2n^2}\left(\sum_y p_y C_y^3 + C_P^3 \right).$$

This holds for every $n \ge 1$, completing the proof. □

D Missing Proofs in Section 6

D.1 Proof of Theorem 6.1

Proof. Let $r(y) = P(y)/Q(y)$ for $y \in \mathrm{supp}(Q)$, and let $\boldsymbol{Z}^* \sim Q^{\otimes n}$. Define

$$L(\boldsymbol{z}) = \frac{1}{n}\sum_{j=1}^{n} r(z_j).$$

Recall that K is uniform over $[n]$ and that the density of $\boldsymbol{Z}$ relative to $Q^{\otimes n}$ is L.

Since, conditioned on $K = i$, the i-th coordinate follows P and all other coordinates are i.i.d. from Q, we have

$$\Pr[K = i, \boldsymbol{Z} = \boldsymbol{z}] = \frac{1}{n} Q^{\otimes n}(\boldsymbol{z})\, r(z_i).$$

On the other hand,

$$\Pr[K = i]\Pr[\boldsymbol{Z} = \boldsymbol{z}] = \frac{1}{n} Q^{\otimes n}(\boldsymbol{z})\, L(\boldsymbol{z}).$$

By the definition of TV-information,

$$\begin{aligned} I_{\mathrm{TV}}(K;\boldsymbol{Z}) &= D_{\mathrm{TV}}\big(\mathsf{Law}(K,\boldsymbol{Z}), \mathsf{Law}(K)\otimes\mathsf{Law}(\boldsymbol{Z})\big) \\ &= \frac{1}{2n}\mathbb{E}_{\boldsymbol{Z}^*\sim Q^{\otimes n}}\left[\sum_{i=1}^{n} |r(Z_i^*) - L(\boldsymbol{Z}^*)|\right] \\ &= \frac{1}{2}\mathbb{E}_{\boldsymbol{Z}^*\sim Q^{\otimes n}}\left[|r(Z_1^*) - L(\boldsymbol{Z}^*)|\right], \end{aligned}$$

where the last equality follows by exchangeability.

Since $Z_1^* \sim Q$,

$$D_{\mathrm{TV}}(P,Q) = \frac{1}{2}\mathbb{E}_{\boldsymbol{Z}^*\sim Q^{\otimes n}}[|r(Z_1^*) - 1|].$$

Using the reverse triangle inequality, we obtain

$$\begin{aligned} |I_{\mathrm{TV}}(K;\boldsymbol{Z}) - D_{\mathrm{TV}}(P,Q)| &\le \frac{1}{2}\mathbb{E}_{\boldsymbol{Z}^*\sim Q^{\otimes n}}\left[|L(\boldsymbol{Z}^*) - 1|\right] \\ &\le \frac{1}{2}\sqrt{\mathbb{E}_{\boldsymbol{Z}^*\sim Q^{\otimes n}}[(L(\boldsymbol{Z}^*) - 1)^2]}. \end{aligned}$$

By independence and $\mathbb{E}_{Y\sim Q}[r(Y)] = 1$,

$$\mathbb{E}[(L(\boldsymbol{Z}^*) - 1)^2] = \frac{1}{n}\,\mathrm{Var}_{Y\sim Q}[r(Y)] = \frac{\chi^2(P\|Q)}{n}.$$

Consequently, $|I_{\mathrm{TV}}(K;\boldsymbol{Z}) - D_{\mathrm{TV}}(P,Q)| \le \frac{1}{2}\sqrt{\frac{\chi^2(P\|Q)}{n}}$. Since $\chi^2(P\|Q) < \infty$ on a finite alphabet whenever $P \ll Q$, the right-hand side converges to zero as $n \to \infty$, proving the theorem. □

D.2 Proof of Theorem 6.2

Proof. Let $\boldsymbol{Z}^* \sim Q^{\otimes n}$ and, for each $i \in [d]$, define

$$L_i(\boldsymbol{z}) = \frac{1}{n}\sum_{k=1}^{n} \frac{\mathbb{1}[z_k = i]}{q_i}, \qquad \overline{L}(\boldsymbol{z}) = \sum_{j=1}^{d} p_j L_j(\boldsymbol{z}).$$

Since $P \ll Q$, these likelihood ratios are well-defined on the support of P.

By the canonical shuffle likelihood-ratio identity, the distributions of $\boldsymbol{Z} \mid Y_1 = i$ and $\boldsymbol{Z}$ have densities L_i and $\overline{L}$, respectively, with respect to $Q^{\otimes n}$.

Using the joint-distribution representation of TV-information, we obtain

$$\begin{aligned} I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) &= D_{\mathrm{TV}}\big(\mathsf{Law}(Y_1, \boldsymbol{Z}), P \otimes \mathsf{Law}(\boldsymbol{Z})\big) \\ &= \frac{1}{2} \sum_{i=1}^{d} p_i \mathbb{E}_{\boldsymbol{Z}^* \sim Q^{\otimes n}} \left[\left| L_i(\boldsymbol{Z}^*) - \overline{L}(\boldsymbol{Z}^*) \right| \right]. \end{aligned}$$

Let N_i denote the number of occurrences of symbol i in $\boldsymbol{Z}^*$. Then

$$(N_1, \ldots, N_d) \sim \mathrm{Mult}(n, Q), \qquad L_i(\boldsymbol{Z}^*) = \frac{N_i}{nq_i}.$$

Substituting these identities yields the claim. □

D.3 Full Proof of Theorem 6.3

Proof. Without loss of generality, let the message alphabet be $\mathcal{M} = \{0, 1\}$, with target distribution $P = (1-p, p)$ and blanket distribution $Q = (1-q, q)$, where $p, q \in (0, 1)$. Let $\boldsymbol{Z}$ denote the shuffled output consisting of one sample from P and $n-1$ independent samples from Q. For any realization $\boldsymbol{z}$, let n_1 and $n_0 = n - n_1$ denote the numbers of occurrences of 1 and 0, respectively.

Since the message alphabet is binary, the TV-information can be written as

$$I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) = \mathbb{E}_{\boldsymbol{Z}} \left[\left| p - \Pr(Y_1 = 1 \mid \boldsymbol{Z}) \right| \right]. \tag{37}$$

By Bayes' rule and the likelihood-ratio representation of the canonical shuffle model,

$$\Pr(Y_1 = 1 \mid \boldsymbol{Z} = \boldsymbol{z}) = \frac{p\, n_1/(nq)}{p\, n_1/(nq) + (1-p)\, n_0/(n(1-q))}. \tag{38}$$

Let $\boldsymbol{Z}^* \sim Q^{\otimes n}$ be the reference experiment consisting of n independent blanket samples. The likelihood ratio of the shuffled output distribution relative to $Q^{\otimes n}$ is

$$L(\boldsymbol{z}) = p \frac{n_1}{nq} + (1-p) \frac{n_0}{n(1-q)}.$$

Using $\Pr(\boldsymbol{Z} = \boldsymbol{z}) = L(\boldsymbol{z}) Q^{\otimes n}(\boldsymbol{z})$, we may rewrite (37) as

$$\begin{aligned} I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) &= \mathbb{E}_{\boldsymbol{Z}^*} \left[L(\boldsymbol{Z}^*) \left| p - \frac{p\, n_1/(nq)}{L(\boldsymbol{Z}^*)} \right| \right] \\ &= p(1-p)\, \mathbb{E}_{\boldsymbol{Z}^*} \left[\left| \frac{n_1}{nq} - \frac{n_0}{n(1-q)} \right| \right] \\ &= \frac{p(1-p)}{nq(1-q)} \mathbb{E}_{N \sim \mathrm{Bin}(n,q)} \left[|N - nq| \right]. \end{aligned}$$

Therefore, $I_{\mathrm{TV}}(Y_1; \boldsymbol{Z}) = \frac{p(1-p)}{n}\, \phi_n(q)$, where $\phi_n(q) = \frac{1}{q(1-q)} \mathbb{E}_{N \sim \mathrm{Bin}(n,q)}[|N - nq|]$. Since $p(1-p)/n$ is independent of q, minimizing the TV-information is equivalent to minimizing $\phi_n(q)$. The conclusion follows immediately from Lemma 6.1. □

D.4 Proof of Lemma 6.1

Proof. The normalized Mean Absolute Deviation of a binomial random variable $X \sim \mathrm{Bin}(n, q)$ is given by the following identity (originally attributed to De Moivre [DZ91, BK13]):

$$\phi_n(q) = \frac{2n\binom{n-1}{k}q^{k+1}(1-q)^{n-k}}{q(1-q)} = 2n\binom{n-1}{k}q^k(1-q)^{n-k-1}, \text{ where } k = \lfloor nq \rfloor.$$

Since $k = \lfloor nq \rfloor$ is constant on the intervals $I_k = [\frac{k}{n}, \frac{k+1}{n})$ for $k \in \{0, 1, \ldots, n-1\}$, we analyze $\phi_n(q)$ as a sequence of functions $f_k(q) = 2n\binom{n-1}{k}q^k(1-q)^{n-k-1}$ defined on each I_k.

To find the extremum of $f_k(q)$ within I_k, we take the derivative with respect to q:

$$f_k'(q) = 2n\binom{n-1}{k}q^{k-1}(1-q)^{n-k-2}\left[k - (n-1)q\right].$$

Setting $f_k'(q) = 0$ yields the critical point $q_k^* = \frac{k}{n-1}$.

Note that $f_0(q)$ is strictly decreasing on $(0, 1/n)$, whereas $f_{n-1}(q)$ is strictly increasing on $[1 - 1/n, 1)$.

For $k \in \{1, \ldots, n-2\}$, the point q_k^* always lies within the interval I_k since:

$$\frac{k}{n} \le \frac{k}{n-1} \le \frac{k+1}{n}$$

holds for all $n \ge 2$. Since $f_k(q)$ is a unimodal function that increases for $q < q_k^*$ and decreases for $q > q_k^*$, it reaches a *local maximum* at q_k^*.

Because $\phi_n(q)$ is continuous on $(0, 1)$ and unimodal within each interval I_k, the **local minima** of $\phi_n(q)$ must occur at the boundary points of these intervals, i.e., at $q \in \{\frac{k}{n}\}_{k=1}^{n-1}$.

Let $a_k = \phi_n(\frac{k}{n})$. We wish to find k that minimizes:

$$a_k = 2n\binom{n-1}{k}\left(\frac{k}{n}\right)^k\left(\frac{n-k}{n}\right)^{n-k-1}.$$

Consider the ratio of consecutive terms:

$$\frac{a_{k+1}}{a_k} = \left(\frac{k+1}{k}\right)^k \cdot \left(\frac{n-k-1}{n-k}\right)^{n-k-1}.$$

Let $g(x) = (1 + 1/x)^x$. It is well known that $g(x)$ is strictly increasing and converges to e as $x \to \infty$. The ratio can be written as:

$$\frac{a_{k+1}}{a_k} = \frac{g(k)}{g(n-k-1)}.$$

- If $k < n - k - 1$ (i.e., $k < \frac{n-1}{2}$), then $g(k) < g(n-k-1)$, which implies $\frac{a_{k+1}}{a_k} < 1$. Thus, the sequence a_k is strictly decreasing.
- If $k > n - k - 1$ (i.e., $k > \frac{n-1}{2}$), then $g(k) > g(n-k-1)$, which implies $\frac{a_{k+1}}{a_k} > 1$. Thus, the sequence a_k is strictly increasing.
- If n is odd and $k = \frac{n-1}{2}$, then $k = n - k - 1$, implying $\frac{a_{k+1}}{a_k} = 1$, so $a_{(n-1)/2} = a_{(n+1)/2}$.

Conclusion. The sequence a_k reaches its minimum when k is as close to $n/2$ as possible. Specifically:

- If n is even, the minimum is at $k = n/2$, corresponding to $q = 1/2$.
- If n is odd, the minimum is at $k = (n-1)/2$ and $k = (n+1)/2$, corresponding to $q = \frac{n-1}{2n}$ and $q = \frac{n+1}{2n}$.

This completes the proof. □

E Missing Proofs in Section 7

E.1 Proof of Theorem 7.1

Proof. Let $\boldsymbol{Z}^* \sim Q^{\otimes n}$ denote the reference experiment consisting of n independent blanket samples. For any realization $\boldsymbol{z}$, let $N_i(\boldsymbol{z})$ be the number of occurrences of symbol i.

By the likelihood-ratio representation of the canonical shuffle model,

$$\frac{\Pr(\boldsymbol{Z} = \boldsymbol{z} \mid Y_1 = i)}{Q^{\otimes n}(\boldsymbol{z})} = \frac{N_i(\boldsymbol{z})}{nq_i}.$$

Consequently, the marginal likelihood ratio is

$$L(\boldsymbol{z}) := \frac{\Pr(\boldsymbol{Z} = \boldsymbol{z})}{Q^{\otimes n}(\boldsymbol{z})} = \sum_{j=1}^{d} p_j \frac{N_j(\boldsymbol{z})}{nq_j}.$$

By Bayes' rule, whenever $L(\boldsymbol{z}) > 0$,

$$\Pr(Y_1 = i \mid \boldsymbol{Z} = \boldsymbol{z}) = \frac{p_i N_i(\boldsymbol{z})/(nq_i)}{L(\boldsymbol{z})}.$$

Applying a change of measure from $\mathsf{Law}(\boldsymbol{Z})$ to $Q^{\otimes n}$, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}}\left[\max_{i\in[d]} \Pr(Y_1 = i \mid \boldsymbol{Z})\right] &= \mathbb{E}_{\boldsymbol{Z}^*}\left[L(\boldsymbol{Z}^*) \max_{i\in[d]} \frac{p_i N_i(\boldsymbol{Z}^*)/(nq_i)}{L(\boldsymbol{Z}^*)}\right] \\
&= \mathbb{E}_{\boldsymbol{Z}^*}\left[\max_{i\in[d]} \frac{p_i N_i(\boldsymbol{Z}^*)}{nq_i}\right] \\
&= \mathbb{E}_{\boldsymbol{N}\sim\mathrm{Mult}(n,Q)}\left[\max_{i\in[d]} \frac{p_i N_i}{nq_i}\right].
\end{aligned}$$

Here, the integrand is defined to be zero when $L(\boldsymbol{Z}^*) = 0$.

Subtracting the optimal prior guessing probability $\max_{i\in[d]} p_i$ yields (29). □

E.2 Expression for the Expected Posterior Variance

For notational clarity, we state the theorem for discrete distributions P and Q on $\mathbb{R}$. The result extends naturally to distributions with probability densities by defining the likelihood ratio as the ratio of these densities.

Theorem E.1 (Expected posterior variance)**.** *Consider the canonical shuffle-only model with real-valued messages, where $Y_1 \sim P$ is shuffled with $n-1$ independent samples from Q. P and Q are discrete distributions on $\mathbb{R}$. Assume $P \ll Q$ and $\mathbb{E}_P[Y_1^2] < \infty$.*

Define the likelihood-ratio $r(y) = \frac{P(y)}{Q(y)}$ on $\mathrm{supp}(Q)$. Let $A_1, \ldots, A_n$ be i.i.d. samples from Q, and define $S_k = \sum_{i=1}^{k} r(A_i), S_0 = 0$.

Then, for $n \geq 2$, the expected posterior variance satisfies

$$\mathrm{EPV}(Y_1; \boldsymbol{Z}) = \mathbb{E}_{Y\sim P}[Y^2] - \mathop{\mathbb{E}_{Y\sim P}}_{S_{n-1}}\left[\frac{r(Y)Y^2}{r(Y)+S_{n-1}}\right] - (n-1)\mathop{\mathbb{E}_{Y,Y'\overset{\text{i.i.d.}}{\sim} P}}_{S_{n-2}}\left[\frac{YY'}{r(Y)+r(Y')+S_{n-2}}\right],$$

where the variables in each expectation are independent, and the distributions of S_{n-1} and S_{n-2} are as defined above.

Proof. Let $\boldsymbol{Z}^* = (Z_1^*, \ldots, Z_n^*) \sim Q^{\otimes n}$, and write

$$r_i = r(Z_i^*), \qquad \beta(\boldsymbol{Z}^*) = \sum_{i=1}^{n} r_i.$$

The likelihood ratio of the shuffled output distribution M_n relative to $Q^{\otimes n}$ is

$$\frac{M_n(\boldsymbol{z})}{Q^{\otimes n}(\boldsymbol{z})} = \frac{1}{n}\sum_{i=1}^{n} r(z_i).$$

Conditioned on an observed output $\boldsymbol{Z} = \boldsymbol{z}$ with $\beta(\boldsymbol{z}) > 0$, the posterior probability that the target message occupies position i is

$$\pi_i(\boldsymbol{z}) := \Pr(K = i \mid \boldsymbol{Z} = \boldsymbol{z}) = \frac{r(z_i)}{\beta(\boldsymbol{z})}.$$

Since $Y_1 = Z_K$, the posterior moments are

$$\mathbb{E}[Y_1 \mid \boldsymbol{Z} = \boldsymbol{z}] = \sum_{i=1}^{n} \pi_i(\boldsymbol{z}) z_i, \qquad \mathbb{E}[Y_1^2 \mid \boldsymbol{Z} = \boldsymbol{z}] = \sum_{i=1}^{n} \pi_i(\boldsymbol{z}) z_i^2.$$

These identities remain valid when multiple positions contain the same message.

By the law of total expectation,

$$\mathrm{EPV}(Y_1; \boldsymbol{Z}) = \mathbb{E}[Y_1^2] - \mathbb{E}_{\boldsymbol{Z}}\left[\left(\mathbb{E}[Y_1 \mid \boldsymbol{Z}]\right)^2\right]. \tag{39}$$

We now evaluate the second term. By changing measure to $Q^{\otimes n}$,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}}\left[\left(\mathbb{E}[Y_1 \mid \boldsymbol{Z}]\right)^2\right] &= \mathbb{E}_{Q^{\otimes n}}\left[\frac{\beta}{n}\left(\frac{\sum_{i=1}^n r_i Z_i^*}{\beta}\right)^2\right] \\
&= \frac{1}{n}\mathbb{E}_{Q^{\otimes n}}\left[\frac{\left(\sum_{i=1}^n r_i Z_i^*\right)^2}{\beta}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{Q^{\otimes n}}\left[\frac{r_i^2 (Z_i^*)^2}{\beta}\right] + \frac{1}{n}\sum_{i\neq j}\mathbb{E}_{Q^{\otimes n}}\left[\frac{r_i r_j Z_i^* Z_j^*}{\beta}\right].
\end{aligned} \tag{40}$$

Terms with $\beta = 0$ are defined to be zero; such outputs have zero probability under the actual shuffled-output distribution.

By exchangeability, all diagonal terms in (40) have the same expectation, as do all off-diagonal terms. Therefore,

$$\mathbb{E}_{\boldsymbol{Z}}\left[\left(\mathbb{E}[Y_1 \mid \boldsymbol{Z}]\right)^2\right] = \mathbb{E}_{Q^{\otimes n}}\left[\frac{r_1^2(Z_1^*)^2}{r_1 + S_{n-1}}\right] + (n-1)\mathbb{E}_{Q^{\otimes n}}\left[\frac{r_1 r_2 Z_1^* Z_2^*}{r_1 + r_2 + S_{n-2}}\right], \tag{41}$$

where S_{n-1} and S_{n-2} denote sums of independent likelihood-ratio weights associated with the remaining $n-1$ and $n-2$ samples, respectively.

Finally, using $Q(y)r(y) = P(y)$, the diagonal term becomes

$$\mathbb{E}_{Q^{\otimes n}}\left[\frac{r_1^2(Z_1^*)^2}{r_1 + S_{n-1}}\right] = \mathbb{E}_{\substack{Y\sim P \\ S_{n-1}}}\left[\frac{r(Y)Y^2}{r(Y) + S_{n-1}}\right],$$

while the off-diagonal term becomes

$$\mathbb{E}_{Q^{\otimes n}}\left[\frac{r_1 r_2 Z_1^* Z_2^*}{r_1 + r_2 + S_{n-2}}\right] = \mathbb{E}_{\substack{Y,Y'\overset{\text{i.i.d.}}{\sim} P \\ S_{n-2}}}\left[\frac{YY'}{r(Y) + r(Y') + S_{n-2}}\right].$$

Substituting these expressions into (41) and then into (39) establishes the claim. □

E.3 Proof of Theorem 7.2

Proof. Let $\mu = \mathbb{E}_P[Y_1]$ and $\sigma_P^2 = \mathrm{Var}_P(Y_1)$. By the law of total variance,

$$\mathrm{EPV}(Y_1; \boldsymbol{Z}) = \sigma_P^2 - \mathrm{Var}\left(\mathbb{E}[Y_1 \mid \boldsymbol{Z}]\right). \tag{42}$$

Let $\boldsymbol{Z}^* = (Z_1^*, \ldots, Z_n^*) \sim Q^{\otimes n}$, and define

$$r(y) = \frac{p_y}{q_y}, \qquad L = \frac{1}{n}\sum_{i=1}^n r(Z_i^*), \qquad T = \frac{1}{n}\sum_{i=1}^n r(Z_i^*)(Z_i^* - \mu).$$

Under the canonical shuffle model, the likelihood ratio of the shuffled-output distribution relative to $Q^{\otimes n}$ is L. Moreover, the posterior mean satisfies

$$\mathbb{E}[Y_1 \mid \boldsymbol{Z} = \boldsymbol{z}] - \mu = \frac{\sum_{i=1}^n r(z_i)(z_i - \mu)}{\sum_{i=1}^n r(z_i)}.$$

Consequently, changing measure to $Q^{\otimes n}$ gives

$$\mathrm{Var}\left(\mathbb{E}[Y_1 \mid \boldsymbol{Z}]\right) = \mathbb{E}_{Q^{\otimes n}}\left[L\left(\frac{T}{L}\right)^2\right] = \mathbb{E}_{Q^{\otimes n}}\left[\frac{T^2}{L}\right], \tag{43}$$

where T^2/L is defined to be zero when $L = 0$.

We first compute the second moment of T. Let $X_i = r(Z_i^*)(Z_i^* - \mu)$. The X_i are independent and identically distributed, with

$$\mathbb{E}[X_i] = \sum_y q_y r(y)(y - \mu) = \mathbb{E}_P[Y_1 - \mu] = 0.$$

Therefore,

$$\mathbb{E}[T^2] = \frac{1}{n}\mathbb{E}[X_1^2] = \frac{1}{n}\sum_y q_y r(y)^2 (y-\mu)^2 = \frac{1}{n}\sum_y \frac{p_y^2 (y-\mu)^2}{q_y}. \tag{44}$$

It remains to show that

$$\mathbb{E}\left[\frac{T^2}{L}\right] = \mathbb{E}[T^2] + O(n^{-3/2}). \tag{45}$$

Since the alphabet is finite and $P \ll Q$, the likelihood ratio r is bounded. Define

$$W = \max_y r(y), \qquad D = \max_y |r(y)(y-\mu)|,$$

and

$$B = \max_{y \in \mathrm{supp}(P)} |y - \mu|.$$

All three constants are finite.

On the event $L \geq 1/2$, we have $\left|\frac{T^2}{L} - T^2\right| \leq 2T^2|L-1|$. By Cauchy–Schwarz,

$$\mathbb{E}\left[\left|\frac{T^2}{L} - T^2\right| \mathbf{1}_{\{L \geq 1/2\}}\right] \leq 2\sqrt{\mathbb{E}[T^4]\,\mathbb{E}[(L-1)^2]}. \tag{46}$$

By independence and centering,

$$\mathbb{E}[T^4] = \frac{n\mathbb{E}[X_1^4] + 3n(n-1)(\mathbb{E}[X_1^2])^2}{n^4} \leq \frac{D^4(3n-2)}{n^3}.$$

Similarly, $\mathbb{E}[(L-1)^2] = \frac{\chi^2(P\|Q)}{n}$. Hence, the right-hand side of (46) is $O(n^{-3/2})$.

On the complementary event $L < 1/2$, we use $|T| \leq BL$ to obtain

$$\left|\frac{T^2}{L} - T^2\right| \leq B^2 L(1-L) \leq B^2.$$

Since L is the average of n independent random variables in $[0, W]$ with mean one, Hoeffding's inequality yields

$$\Pr(L < 1/2) \leq \exp\left(-\frac{n}{2W^2}\right).$$

Thus,

$$\mathbb{E}\left[\left|\frac{T^2}{L} - T^2\right| \mathbf{1}_{\{L < 1/2\}}\right] = O(e^{-cn}) \tag{47}$$

for some constant $c > 0$ depending only on P and Q.

Combining (46) and (47) proves (45). Substituting (44) into (43), and then into (42), establishes

$$\mathrm{EPV}(Y_1; \boldsymbol{Z}) = \sigma_P^2 - \frac{1}{n}\sum_y \frac{p_y^2 (y-\mu)^2}{q_y} + O(n^{-3/2}).$$

We now optimize the leading-order term. By the Cauchy–Schwarz inequality, the optimal Q is supported on $\mathrm{supp}(P)$ and

$$q_y \propto p_y |y - \mu|.$$

Under the assumption that $y \neq \mu$ for all $y \in \mathrm{supp}(P)$, every numerator is positive, so this uniquely determines an admissible blanket distribution:

$$q_y^* = \frac{p_y|y-\mu|}{\mathbb{E}_P|Y_1-\mu|}.$$

Finally, if $\mu \in \mathrm{supp}(P)$, the formal optimizer assigns zero mass to μ. Although this violates $P \ll Q$, the infimum can be approached by setting $Q_\varepsilon = (1-\varepsilon)Q^* + \varepsilon P,\ \varepsilon > 0$, and letting $\varepsilon \to 0$. This completes the proof. □

F Analysis of the Canonical Shuffle-Only Model when $P \not\ll Q$

We extend the analysis to distributions P and Q on a common finite alphabet, allowing P to assign positive mass outside $\mathrm{supp}(Q)$.

Let $\mathbb{Y} = \mathbb{Y}_1 \cup \mathbb{Y}_2$, where

$$\mathbb{Y}_2 = \{y \mid P(y) > 0 \text{ and } Q(y) = 0\}, \qquad \mathbb{Y}_1 = \mathbb{Y} \setminus \mathbb{Y}_2.$$

Define the indicator random variable $T = \mathbb{1}[Y_1 \in \mathbb{Y}_2]$. When $T = 1$, the event $Y_1 \in \mathbb{Y}_2$ implies that Y_1 takes a value impossible under Q, which makes the shuffled output $\boldsymbol{Z}$ reveal Y_1 completely. Conditioning on $T = 0$, the distribution reduces to a restricted case $P' = P\,|_{\mathbb{Y}_1}$ such that $P' \ll Q$. Let $\beta = \Pr[T = 1] = P(\mathbb{Y}_2)$, so that $P = (1-\beta)P\,|_{\mathbb{Y}_1} + \beta P\,|_{\mathbb{Y}_2}$.

When $\beta = 0$, we have $P \ll Q$. When $\beta = 1$, the supports of P and Q are disjoint, so $I(K;\boldsymbol{Z}) = \log n$ and $I(Y_1;\boldsymbol{Z}) = H(P)$. We therefore focus on the case $0 < \beta < 1$ below.

The following results generalize the asymptotic expansions obtained in the absolutely continuous case.

Theorem F.1. *In the canonical shuffle-only model, when $P \not\ll Q$, the mutual information between the shuffle position K and the shuffled output $\boldsymbol{Z}$ satisfies*

$$I(K;\boldsymbol{Z}) \;=\; \beta \log n + (1-\beta)\, D_{\mathrm{KL}}(P'\|Q) - (1-\beta)\,\frac{\chi^2(P'\|Q)}{2n} + O(n^{-3/2}), \qquad n \to \infty,$$

where $\beta = P(\mathbb{Y}_2)$, $P' = P|_{\mathbb{Y}\setminus\mathbb{Y}_2}$, and $\mathbb{Y}_2 = \{y \mid P(y) > 0$ and $Q(y) = 0\}$.

Proof. By the chain rule, $I(K;\boldsymbol{Z},T) = I(K;T) + I(K;\boldsymbol{Z} \mid T) = I(K;\boldsymbol{Z}) + I(K;T \mid \boldsymbol{Z})$, we obtain

$$I(K;\boldsymbol{Z}) = I(K;\boldsymbol{Z} \mid T) + I(K;T) - I(K;T \mid \boldsymbol{Z}).$$

Observe that T is deterministically recoverable from $\boldsymbol{Z}$ (one simply checks whether $\boldsymbol{Z}$ contains any element from $\mathbb{Y}_2$). Hence $I(K;T \mid \boldsymbol{Z}) = 0$. Furthermore, K and T are independent: T is an intermediate random variable generated in the process $Y_1 \sim P$, whereas K is drawn independently by the shuffler. Thus $I(K;T) = 0$.

Therefore,

$$I(K;\boldsymbol{Z}) = I(K;\boldsymbol{Z} \mid T) = I(K;\boldsymbol{Z} \mid T=0)\Pr[T=0] + I(K;\boldsymbol{Z} \mid T=1)\Pr[T=1].$$

When $T = 1$, i.e., $Y_1 \in \mathbb{Y}_2$, the shuffle output deterministically reveals the position of Y_1 in $\boldsymbol{Z}$. Hence $I(K;\boldsymbol{Z} \mid T=1) = \log n$.

When $T = 0$, the random variable Y_1 follows the restricted distribution $P' \ll Q$. In this case, all previously derived expansions for $I(K;\boldsymbol{Z})$ under $P \ll Q$ (see Theorem 5.1) directly apply to P':

$$I(K;\boldsymbol{Z} \mid T=0) = D_{\mathrm{KL}}(P'\|Q) - \frac{\chi^2(P'\|Q)}{2n} + O(n^{-3/2}).$$

Combining the two parts yields

$$I(K;\boldsymbol{Z}) = \beta \log n + (1-\beta)\left(D_{\mathrm{KL}}(P'\|Q) - \frac{\chi^2(P'\|Q)}{2n}\right) + O(n^{-3/2}),$$

as claimed. □

Theorem F.2. *In the canonical shuffle-only model, when $P \not\ll Q$, the mutual information between the target message and the shuffled output satisfies*

$$I(Y_1;\boldsymbol{Z}) \;=\; \sum_{y\in\mathbb{Y}_2} P(y)\log\frac{1}{P(y)} + (1-\beta)\log\frac{1}{1-\beta} + (1-\beta)\frac{\sum_i \frac{p'_i-(p'_i)^2}{q_i}}{2n} + O(n^{-3/2}), \qquad n\to\infty,$$

where $\beta = P(\mathbb{Y}_2)$, $P' = P|_{\mathbb{Y}\setminus\mathbb{Y}_2}$, and $\mathbb{Y}_2 = \{y \mid P(y) > 0 \text{ and } Q(y) = 0\}$.

Proof. Analogously to the proof of Theorem F.1, we obtain

$$I(Y_1;\boldsymbol{Z}) = I(Y_1;\boldsymbol{Z}\mid T)+I(Y_1;T) = I(Y_1;\boldsymbol{Z}\mid T=0)\Pr[T=0] + I(Y_1;\boldsymbol{Z}\mid T=1)\Pr[T=1]+I(Y_1;T).$$

By definition,

$$I(Y_1;T) = H(Y_1) - H(Y_1\mid T) = H(Y_1) - \beta H(Y_1\mid T=1) - (1-\beta)H(Y_1\mid T=0).$$

When $T=1$, the shuffle output directly reveals the value of Y_1, thus $I(Y_1;\boldsymbol{Z}\mid T=1) = H(Y_1\mid T=1)$. When $T=0$, we again invoke Theorem 5.2 (the case $P'\ll Q$), giving

$$I(Y_1;\boldsymbol{Z}\mid T=0) = \frac{\sum_i \frac{p'_i-(p'_i)^2}{q_i}}{2n} + O(n^{-3/2}).$$

By definition,

$$H(Y_1\mid T=0) = \sum_{y\in\mathbb{Y}_1}\frac{P(y)}{1-\beta}\log\frac{1-\beta}{P(y)}.$$

Combining these parts yields

$$\begin{aligned} I(Y_1;\boldsymbol{Z}) &= H(Y_1) - (1-\beta)H(Y_1\mid T=0) + I(Y_1;\boldsymbol{Z}\mid T=0)\Pr[T=0] \\ &= \sum_{y\in\mathbb{Y}} P(y)\log\frac{1}{P(y)} - \sum_{y\in\mathbb{Y}_1} P(y)\log\frac{1-\beta}{P(y)} + (1-\beta)\frac{\sum_i \frac{p'_i-(p'_i)^2}{q_i}}{2n} + O(n^{-3/2}) \\ &= \sum_{y\in\mathbb{Y}_2} P(y)\log\frac{1}{P(y)} + (1-\beta)\log\frac{1}{1-\beta} + (1-\beta)\frac{\sum_i \frac{p'_i-(p'_i)^2}{q_i}}{2n} + O(n^{-3/2}). \end{aligned}$$

□