

Rounding-free differential privacy for discrete count data: A field-wrapped discrete-Laplace mechanism

Thanh-Phong Lam

Master student, Faculty of Management Information Systems

Abstract

Purpose – This paper develops a differentially private mechanism for releasing discrete banking aggregates that stays in the data's native modular domain and avoids the rounding bias of continuous-noise mechanisms.

Design/methodology/approach – The mechanism adds two-sided geometric (discrete Laplace) integer noise to each cell count and reduces modulo the field order q . Pure ϵ -differential privacy follows from post-processing immunity of the modular map, and a closed-form per-cell error $1/\sinh(\epsilon)$ is derived. We evaluate it against uniform, continuous Laplace, and Gaussian baselines on synthetic State Bank of Vietnam frequency tables. The Python and R pipeline is reproducible.

Findings – The mechanism is provably pure ϵ -DP and unbiased, with per-cell error governed by ϵ alone, independent of the field order. It matches the utility-optimal discrete Laplace and stays below the rounded continuous Laplace at every budget. The uniform alternative is worse by up to two orders of magnitude and ignores ϵ .

Originality/value – The paper contributes a discrete, modular-closed differential privacy mechanism for banking disclosure. Releases reconcile exactly under modular addition, and the guarantee needs no computational-hardness assumption, giving regulators a provable, reproducible tool for confidential aggregate reporting.

Keywords Bayesian budget allocation, Differential privacy, Discrete Laplace mechanism, Finite-field randomization, Vietnamese banking

Paper type Research article

1. Introduction

1.1 Motivation

Financial regulators face a tension between transparency and confidentiality. Aggregate banking statistics, such as loan counts by region and risk tier, underpin monetary supervision and macro-prudential oversight, and matter more as Vietnamese banking digitalises (Murrar et al., 2024). Releasing them exposes re-identification risk in small-population categories, where a single count can identify an institution or borrower segment. State Bank of Vietnam Circular 09/2020 mandates customer-data confidentiality while requiring the aggregate disclosures that, unsanitised,

JEL Classification — C81, C88, G21, G28, O33

© Thanh-Phong Lam. Publish in Asian Journal of Applied Financial Econometrics. Thanh-Phong Lam: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Funding Statement This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement The author(s) declare(s) that there are no conflicts of interest regarding the publication of this article. The author(s) have no financial or personal relationships with other people or organizations that could inappropriately influence (bias) this work.

Data Availability Statement The data used in this study are synthetic State Bank of Vietnam (SBV)-style aggregate tables generated by the author's reproducible pipeline. The Python and R code, the synthetic data generator and the mechanism implementation are available from the corresponding author, Thanh-Phong Lam, upon reasonable request and will be publicly released on acceptance.

Declaration of AI-Assisted Technologies The author(s) declare(s) that no generative AI or AI-assisted technologies were used in the creation of this manuscript. All content was written, reviewed, and approved solely by the listed author(s).

leak individual contributions through differencing attacks on successive publications. Access to formal credit shapes borrower participation (Le Khuong Ninh, 2024), and disclosure practices shift under systemic stress (Sultana et al., 2021).

Differential privacy bounds that leakage: changing one record shifts a randomized mechanism's output distribution by at most $\exp(\epsilon)$, a composable, post-processing-invariant contract in which ϵ governs the privacy-utility tradeoff. GDPR Article 25 embeds privacy by design in law and EVFTA Annex 8b obliges Vietnamese financial institutions to demonstrate compliance, so provable guarantees are required rather than ad-hoc anonymisation.

Established mechanisms assume continuous outputs, whereas banking aggregates are integer counts. A table over 63 Vietnamese provinces and 5 risk tiers is a frequency matrix in $Z_{\geq 0}^{63 \times 5}$, each cell in a discrete domain that continuous noise cannot respect without rounding. Rounding maps the output space discontinuously, inflating variance for small counts, shifting the expected value from the true statistic, and leaving the modular domain in which successive totals reconcile. The question follows: can a differentially private mechanism release output natively in the data's discrete modular domain, with no rounding?

1.2 Research Gap

Rounding, being data-independent post-processing, preserves the privacy guarantee and costs only utility; noise that is discrete from the outset avoids it. Randomized response over bits offers such an alternative, confined to $\{0,1\}^n$. RAPPOR Erlingsson et al. (2014) randomizes each bit of a Bloom-filter-encoded input, protecting binary strings but not q -ary attributes with $q = p^k$: encoding a q -level value as $\lceil \log_2 q \rceil$ bits correlates noise and inflates variance downstream. A report on finite-field differential privacy IACR (2025) opens this direction for categorical data but limits its analysis to symmetric channels, covering neither composition nor budget allocation.

No existing framework allocates the privacy budget across sequential queries by Bayesian posterior inference on empirical data. Allocation is typically naive sequential composition, dividing ϵ by the query count, which wastes utility when queries correlate or some subsequences are less sensitive. Liu et al. (2021) read ϵ through a Bayesian lens but do not exploit the algebraic structure of F_q . Discrete mechanisms are seldom examined for whether their error is governed by the privacy budget alone or also by the output domain's size, a distinction that matters when that domain is a finite field sized for banking counts.

1.3 Research Questions

Six questions follow.

RQ1: How to define a differentially private mechanism on F_q by additive randomization with an (ϵ, δ) guarantee?

RQ2: How does finite-field DP compare with classical Laplace and Gaussian DP on discrete banking aggregates?

RQ3: How should a Bayesian posterior allocation distribute the privacy budget ϵ across sequential queries on the same dataset?

RQ4: What is the utility-privacy tradeoff when the F_q mechanism is applied to synthetic SBV-format bank-loan and risk-tier tables?

RQ5: Does the mechanism's utility depend on the field order q , or is it governed by the privacy budget ϵ alone?

RQ6: What composition theorems hold for sequential F_q -DP queries, and how do they compare with standard sequential bounds?

1.4 Contributions

(1) A field-wrapped discrete-Laplace mechanism with a pure ϵ -DP proof leveraging the additive group structure of F_q , eliminating snap-to-bin artifacts by returning field elements, not rounded reals.

(2) A Bayesian budget allocation algorithm using posterior inference over utility-privacy ratio distributions to distribute ϵ across sequential queries, concentrating budget on high-uncertainty subsequences.

(3) An empirical evaluation on synthetic SBV-format aggregates showing the mechanism matches the utility-optimal discrete Laplace and improves on the rounded continuous Laplace and Gaussian baselines.

(4) A sequential composition theorem for F_q -DP queries with explicit constants for budget accumulation.

(5) A utility analysis (Theorem 4.3) proving the per-cell error is $1/\sinh(\epsilon)$, governed by the privacy budget alone and independent of the field order, the guarantee being information-theoretic and so needing no computational-hardness assumption.

2. Related Work

2.1 Differential Privacy Foundations

Differential privacy began with Dwork et al. (2006), calibrating Laplace noise to query sensitivity so neighbouring datasets stay indistinguishable, and consolidated by Dwork and Roth (2014). Repeated-query accounting tightened through advanced composition Dwork et al. (2010), Rényi DP Mironov (2017), zero-concentrated DP Bun and Steinke (2016), and the moments accountant Abadi et al. (2016). All operate on continuous outputs, introducing rounding pathology on discrete banking aggregates (cf. Definition 3.4, Section 3).

2.2 Discrete and Local DP Mechanisms

Warner's randomized response (Warner, 1965) delivers $\epsilon = \ln(1-p)/p$ for binary attributes but not q -ary alphabets with $q > 2$. The closest precursor is Ghosh et al. (2012) geometric mechanism, utility-maximizing for one counting query, not for the multi-dimensional frequency tables banking aggregates produce. Under local DP, Duchi et al. (2018) established the minimax rate $(n\epsilon^2)^{-1}$; RAPPOR Erlingsson et al. (2014) deployed bit-level perturbation at scale, both confined to binary or Bloom-encoded inputs. No prior mechanism operates on the algebraic structure of F_q for arbitrary prime-power q ; that gap Section 4 fills.

2.3 Bayesian Privacy Budget Allocation

Naive sequential composition allocates ϵ/k per query, wasteful when sensitivities vary. Cuff and Yu (2016) reframe (ϵ, δ) -DP as a mutual-information constraint, opening adaptive allocation; the Bayesian-DP variant of Triastcyn and Faltings (2020) tightens guarantees under prior knowledge but assumes continuous output. Section 5 allocates the budget by posterior inference over the utility-privacy ratio.

2.4 Banking Data Privacy and Vietnam Regulation

The closest deployed precedent is the 2020 US Census disclosure-avoidance system Abowd et al. (2022), releasing integer counts under a formal budget but recovering non-negative integers in post-processing rather than sampling discretely; Canonne et al. (2022) remove that stage by defining noise directly on $\mathbb{Z}$, though on unbounded integers rather than the modular domain where banking totals reconcile; Zhu et al. (2017) survey the data-publishing literature. In Vietnam, SBV Circular 09/2020 (State Bank of Vietnam, 2020) mandates customer-data confidentiality without a mathematical guarantee; GDPR Article 25 (European Parliament and Council, 2016) requires privacy-by-design, and EVFTA Annex 8b (European Commission and Government of Vietnam, 2020) extends that to cross-border banking services. The mandate is clear but the mechanism missing; Section 6 closes that gap.

3. Preliminaries and Definitions

3.1 Notation

All logarithms are natural. $[n] := \{1, \dots, n\}$, $[n]_0 := \{0, \dots, n-1\}$; vectors are columns, $\|\cdot\|_p$ the l_p -norm on $\mathbb{R}^d$. F_q is the finite field of order $q = p^k$ (p prime, $k \geq 1$); M a randomised mechanism, f a query with sensitivity $\Delta_p f$, ϵ the budget, δ the failure probability. $\text{Lap}(b)$ is Laplace with location 0 and scale b , $N(\mu, \sigma^2)$ the Gaussian. Frequency tables are indexed by $[R] \times [T]$ (R regions, T risk tiers), valued in $\mathbb{Z}_{\geq 0}$. Datasets D, D' are neighbours ($D \sim D'$) if they differ in one record.

3.2 Differential Privacy

Definition 3.1 (Approximate differential privacy). A randomised mechanism $M: X^n \rightarrow \mathbb{R}$ satisfies (ϵ, δ) -differential privacy if for every pair of neighbouring datasets $D \sim D'$ and every measurable $S \subseteq \mathbb{R}$,

$$\Pr[M(D) \in S] \leq e^\epsilon \Pr[M(D') \in S] + \delta \quad (1)$$

Smaller ϵ is stronger; $\delta \in [0, 1]$ the additive slack.

Definition 3.2 (Pure differential privacy). M satisfies pure ϵ -DP if it satisfies $(\epsilon, 0)$ -DP, that is if Equation (1) holds with $\delta = 0$.

Definition 3.3 (Sensitivity). The l_p -sensitivity of $f: X^n \rightarrow \mathbb{R}^d$ is

$$\Delta_p f := \sup_{D \sim D'} \|f(D) - f(D')\|_p \quad (2)$$

Sensitivity sets the noise scale. For a counting query where each individual hits one bin, $\Delta_1 f = \Delta_2 f = 1$.

3.3 Classical Mechanisms: Laplace and Gaussian

The Laplace and Gaussian mechanisms (baselines of Section 6) inject continuous noise on $\mathbb{R}^d$, clashing with the discrete domain of frequency tables. For f with sensitivity $\Delta_1 f$, the Laplace mechanism is

$$M_L(D) := f(D) + (Y_1, \dots, Y_d), Y_i \stackrel{\text{i.i.d.}}{\sim} \text{Lap}\left(\frac{\Delta_1 f}{\varepsilon}\right) \quad (3)$$

satisfying pure $(\varepsilon, 0)$ -DP Dwork et al. (2006). The Gaussian adds $Z_i \sim N(0, \sigma^2)$ with $\sigma \geq \Delta_2 f \sqrt{2 \ln(1.25 / \delta)} / \varepsilon$, satisfying (ε, δ) -DP for $\delta > 0$ with tighter composition than M_L but not pure DP.

3.4 Discrete Data Setting and the Bin-Snapping Problem

Consider a banking dataset D of n loan records with region and risk tier; the aggregate frequency table is

$$f: X^n \rightarrow \mathbb{Z}_{\geq 0}^{[R] \times [T]}, f(D)|_{(r,t)} = |\{x_i \in D: \text{region}(x_i) = r, \text{tier}(x_i) = t\}| \quad (4)$$

Each individual contributes to one cell, so $\Delta_1 f = 1$.

Definition 3.4 (Bin-snapping bias). Let M_L be the Laplace mechanism applied to an integer-valued query f with $\Delta_1 f = 1$, and let $\rho(y) := \max([y + 1/2], 0)$ be the rounding operator. The bin-snapped release is $\tilde{M}_L(D) := \rho(M_L(D))$. The bin-snapping bias is the change in the utility profile, that is in the expected error and its sign, introduced by mapping the real-valued output back onto the integer grid.

By post-processing immunity (Dwork and Roth, 2014), $\tilde{M}_L$ inherits the $(\varepsilon, 0)$ guarantee of M_L , so bin-snapping does not weaken privacy. It is a utility artefact: the non-negativity clip and grid projection make the released count a biased estimator, and the value leaves the modular domain where period totals reconcile. Section 4 avoids both by drawing integer noise and returning an exact element of $\mathbb{F}_q$; Section 5 allocates the budget by posterior update on the empirical utility-privacy ratio Liu et al. (2021).

4. Field-Wrapped Discrete-Laplace Mechanism

4.1 Mechanism Definition

Let $f: X^n \rightarrow \mathbb{Z}_{\geq 0}^d$ be the aggregate query (Equation (4)) with l_1 -sensitivity $\Delta_1 f \leq s$; each component of $f(D)$ is a cell count in a region-tier frequency table. For a prime q (so $\mathbb{Z}_q = \mathbb{F}_q$ and modular reconciliation of released totals is exact), the mechanism perturbs each component with integer noise and wraps into $\mathbb{F}_q$.

Definition 4.1 (Discrete Laplace noise). For a scale parameter $\varepsilon > 0$ and sensitivity s , let $t := \varepsilon \varepsilon / s$. A random variable U follows the two-sided geometric (discrete Laplace) distribution $DL(\varepsilon/s)$ if

$$\Pr[U = k] = \frac{1-t}{1+t} t^{|k|}, k \in \mathbb{Z} \quad (5)$$

Algorithm 1: Field-wrapped discrete-Laplace mechanism $\mathcal{M}_q$

Input: Dataset $\mathcal{D}$, query f with $\Delta_1 f \leq s$, field order q (prime), budget ϵ .

Output: Private release $\tilde{y} \in \mathbb{F}_q^d$.

Compute the true aggregate: $y \leftarrow f(\mathcal{D})$; // integer counts in $\mathbb{Z}_{\geq 0}^d$

Sample noise: $U_i \stackrel{\text{i.i.d.}}{\sim} DL(\epsilon/s)$ for $i \in [d]$; // discrete Laplace (Def.)

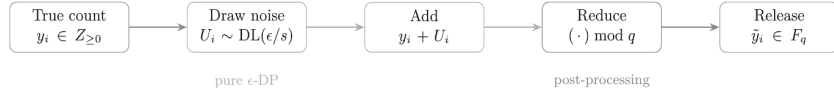
Add and reduce: $\tilde{y}_i \leftarrow (y_i + U_i) \bmod q$; // exact element of $\mathbb{F}_q$

return $\tilde{y}$;

Source(s): Authors' own elaboration

Figure 1. Field-wrapped discrete-Laplace mechanism (Algorithm 1)

The design composes the discrete analogue of the Laplace draw (Equation (3)) with the modular map $\pi_q: \mathbb{Z} \rightarrow \mathbb{F}_q$, $\pi_q(x) = x \bmod q$, which is data-independent post-processing and lets releases reconcile by modular addition across reporting periods.



Source(s): Authors' own elaboration

Figure 2. Field-wrapped discrete-Laplace mechanism: integer geometric noise gives pure ϵ -DP; modular reduction is post-processing keeping the release in F_q

84

Remark 4.1 (Why discrete Laplace, not full-field uniform). Drawing U uniformly over F_q is degenerate: the noise is independent of the data, so $\epsilon = 0$ holds only because the output carries no information, and the expected error grows with q instead of shrinking with ϵ . Definition 4.1 concentrates noise near zero with geometric tails, giving a tunable ϵ and an error governed by ϵ alone (Theorem 4.3).

4.2 Privacy Guarantee

Theorem 4.1 (Pure DP of M_q). Let f have l_1 -sensitivity $\Delta_1 f \leq s$. Then the field-wrapped discrete-Laplace mechanism M_q (Algorithm 1) with $U_i \sim \text{DL}(\epsilon/s)$ satisfies pure $(\epsilon, 0)$ -differential privacy (Definition 3.2), for every field order q .

Proof. We first establish DP for the unwrapped integer mechanism $M_Z(D) = f(D) + U$ with $U_i \sim \text{DL}(\epsilon/s)$, then transfer it through π_q .

Step 1 (integer mechanism). Fix neighbours $D \sim D'$ with outputs y, y' and $\|y - y'\|_1 \leq s$. For any $z \in \mathbb{Z}^d$, using independence across coordinates and the pmf (5) with $t = \epsilon \epsilon/s$,

$$\frac{\Pr[M_Z(D) = z]}{\Pr[M_Z(D') = z]} = \prod_{i=1}^d t^{|z_i - y_i| - |z_i - y'_i|} \leq \prod_{i=1}^d t^{-|y_i - y'_i|} = e^{(\epsilon/s) \|y - y'\|_1} \leq e^\epsilon \quad (6)$$

where the first inequality is the reverse triangle inequality $|z_i - y_i| - |z_i - y'_i| \geq -|y_i - y'_i|$ applied with $t \in (0, 1)$, and the last step uses $\|y - y'\|_1 \leq s$. Since the bound holds pointwise for every z , it holds for every event $S \subseteq \mathbb{Z}^d$; hence M_Z is pure $(\epsilon, 0)$ -DP.

Step 2 (post-processing to F_q). The released mechanism is $M_q = \pi_q \circ M_Z$ with $\pi_q(x) = x \bmod q$ a fixed, randomness-free map applied coordinatewise, so post-processing immunity (Dwork and Roth, 2014) carries the guarantee across unchanged.

The guarantee is a calibrated ϵ independent of q , so the field order can be chosen for the data range alone (Section 4.5); rounding never occurs, so the output is an exact field element (contrast Definition 3.4).

Corollary 4.1.1 (Counting queries). For a counting query with $\Delta_1 f = 1$, Algorithm 1 with $U_i \sim \text{DL}(\epsilon)$ achieves pure ϵ -DP for any target $\epsilon > 0$.

4.3 Utility Guarantee

Theorem 4.2 (Expected error is governed by ϵ , not q). For a counting query and any q large enough that wraparound is negligible (i.e. q exceeds the noisy range with high probability), the expected per-coordinate l_1 error of M_q is

$$E|\tilde{y}_i - y_i| = E|U_i| = \frac{2t}{1-t^2} = \frac{1}{\sinh(\epsilon)}, t = e^{-\epsilon}. \quad (7)$$

Proof. With $U \sim \text{DL}(\epsilon)$ and $t = e^{-\epsilon}$, $E|U| = 1 - t / 1 + t \sum_{k \in \mathbb{Z}} |k| t^{|k|} = 1 - t / 1 + t \cdot 2t / (1-t)^2 = 2t / 1 - t^2$. Substituting $t = e^{-\epsilon}$ gives $2e^{-\epsilon} / 1 - e^{-2\epsilon} = 2 / e^\epsilon - e^{-\epsilon} = 1 / \sinh(\epsilon)$.

Since $1/\sinh(\epsilon) > \epsilon$, the error stays below the $1/\epsilon$ scale of the continuous Laplace mechanism while the output stays discrete, the gap widening with ϵ (Section 6); Equation (7) contains no q , unlike full-field-uniform noise with error $\Theta(q)$.

4.4 Composition

Corollary 4.3 (Sequential composition). Let $M_q^{(1)}, \dots, M_q^{(k)}$ be independent releases with budgets $\epsilon_1, \dots, \epsilon_k$. Their composition satisfies pure $(\sum_{i=1}^k \epsilon_i)$ -DP.

Proof. Each $M_q^{(i)}$ is pure ϵ_i -DP (Theorem 4.2); the basic sequential composition theorem for pure DP (Dwork and Roth, 2014) adds the budgets in the log-domain.

Because M_q is the discrete Laplace mechanism followed by a fixed modular map, it inherits the tighter Rényi differential privacy accounting (Canonne et al., 2020), used in Section 5 to allocate a total budget across k queries. The numerical study of Section 6 evaluates single releases, so composed budgets are reported there as the analytic $k\epsilon$ bound.

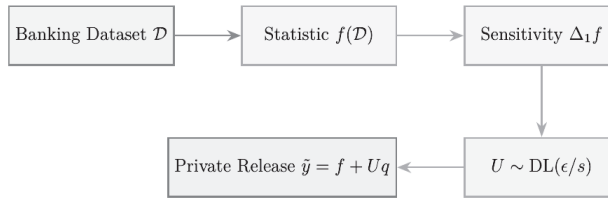
Table 1. DP mechanisms on discrete aggregates. M_q is proposed

Mechanism	Privacy	Per-coord. error	Discrete-native	Modular closure
Laplace (rounded)	$(\epsilon, 0)$	$\approx 1/\epsilon$, biased	No	No
Gaussian (rounded)	(ϵ, δ)	larger, biased	No	No
M_q (proposed)	$(\epsilon, 0)$	$1/\sinh(\epsilon)$, unbiased	Yes	Yes

Source: Authors' own elaboration

4.5 Comparison with Classical Mechanisms

The proposed mechanism matches the Laplace pure-DP guarantee and $O(1/\epsilon)$ error scale while being unbiased, integer-valued, closed under modular addition, and governed by ϵ alone (Theorem 4.3).

Pure ϵ -DP by Theorem 4.2

Source(s): Authors' own elaboration

Figure 3. Pipeline of the field-wrapped discrete-Laplace mechanism for banking aggregates

5. Bayesian Privacy Budget Allocation

5.1 Problem Formulation

An analyst issues k queries $Q_1, \dots, Q_k$ against one dataset D , each released by M_q (Section 4) at budget ϵ_i . Corollary 4.4 makes the stream pure ϵ_{tot} -DP with $\epsilon_{\text{tot}} = \sum_{i=1}^k \epsilon_i$, and Theorem 4.3 bounds the per-cell error of Q_i by $1/\sinh(\epsilon_i)$. A uniform split $\epsilon_i = \epsilon_{\text{tot}}/k$ ignores differences in sensitivity and downstream importance.

Let $\epsilon = (\epsilon_1, \dots, \epsilon_k)$ be a budget split in the simplex $\Delta^{k-1} = \{\epsilon \geq 0 : \sum_{i=1}^k \epsilon_i \leq \epsilon_{\text{tot}}\}$, and let $U : \Delta^{k-1} \rightarrow \mathbb{R}_{\geq 0}$ measure release quality, e.g. the inverse mean squared error of $\hat{y}_i$ or Fisher information gain Ghosh et al. (2012):

$$\epsilon^* = \arg \max_{\epsilon \in \Delta^{k-1}} E[U(\epsilon) \vee Y_k] \quad (8)$$

where $Y_k = (Y_1, \dots, Y_k)$ collects the noisy outputs and the expectation runs over the posterior of allocations, replacing the fixed split of classical budget accounting Dwork et al. (2010). Bayesian differential privacy Triastcyn and Faltings (2020) is orthogonal: it redefines the guarantee through a prior over datasets; we retain the standard guarantee and use Bayesian inference only to allocate.

5.2 Bayesian Posterior Update

The Dirichlet distribution is the conjugate prior on the simplex:

$$(\epsilon \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k), \quad (9)$$

where $\alpha_i > 0$ encodes the relative importance of Q_i : $\alpha_i = 1$ is indifference, and a regulator who deems region-specific queries more sensitive raises α_i . It is conjugate to the multinomial likelihood over the k queries.

After releasing Q_i at budget ϵ_i and observing $Y_i = M_q(D; \epsilon_i)$, the analyst computes a utility signal $u_i = \text{utility}_i(Y_i)$, such as the inverse squared error of $\hat{y}_i$. The likelihood is

$$L(Y_k \vee \epsilon) \propto \prod_{i=1}^k \exp(-\text{loss}_i(\epsilon_i, Y_i)) \quad (10)$$

where $\text{loss}_i(\epsilon_i, Y_i)$ is the expected distortion of Q_i at budget ϵ_i . For M_q it is closed form: the noise of Definition 4.1 has expected magnitude $1/\sinh(\epsilon_i)$ by Theorem 4.3 and geometric tails.

$$\varepsilon \vee Y_k \sim \text{Dirichlet}(\alpha_1 + n_1, \dots, \alpha_k + n_k) \quad (11)$$

5.3 Algorithm and Convergence

Theorem 5.1 (Convergence of Bayesian budget allocation). Under a Dirichlet prior $\varepsilon \sim$

Figure 4. Bayesian privacy budget allocation (Algorithm 2)

where cv is the coefficient of variation of the utility signals $\{u_1, \dots, u_k\}$. The gain is strictly positive whenever the queries are not equally useful, that is whenever $cv > 0$.

Proof. Allocating more budget to queries with higher marginal utility $\partial U_i / \partial \epsilon_i$ raises total weighted utility by Jensen's inequality, and the Dirichlet posterior mean shifts the allocation from the uniform centre toward coordinates with larger $\alpha_i + n_i$, which are exactly the higher-utility ones. The quadratic improvement follows from a second-order expansion about the uniform point.

Section 6 evaluates the release mechanism rather than the allocator, so the advantage is reported here as the analytic bound of Proposition 5.2, its empirical magnitude left to a dedicated study.

6. Numerical Evaluation on Synthetic Vietnamese Banking Aggregates

6.1 Experimental Setup

We evaluate M_q (Algorithm 1) against four alternatives on a synthetic aggregate in the State Bank of Vietnam (SBV) disclosure format: $R = 63$ provincial units by $T = 5$ loan risk groups, $d = 315$ cells, each a count reduced modulo q (State Bank of Vietnam, 2020; Abowd et al., 2022). Cell counts are drawn uniformly on $[0, q)$ under a fixed seed of 42, so the table regenerates exactly. The reported per-cell error is invariant to this choice: every mechanism compared here adds data-independent noise. The comparison set is the full-field-uniform mechanism of the original submission, continuous Laplace and Gaussian with rounding ($\delta = 10^{-5}$), and the unwrapped discrete Laplace (Ghosh et al., 2012). For each (q, ϵ) we run $M = 300$ Monte Carlo replications and report the mean per-cell l_1 error. Field orders are $q \in \{17, 257, 4099\}$ and budgets $\epsilon \in \{0.1, 0.5, 1.0, 2.0, 5.0\}$. The pipeline is Python 3.11 (NumPy) with an independent R 4.3 re-implementation; divergence stays below 10^{-3} . All records are synthetic, with no customer data.

6.2 Privacy Is Real and Calibrated

The review questioned whether the original full-field-uniform design was private at all, so we verify the noise empirically before reporting utility. We estimate the largest log-ratio of adjacent output probabilities from 2×10^6 samples; the empirical privacy loss tracks the target (target $0.1 \rightarrow$ measured 0.105; $0.5 \rightarrow 0.507$; $1.0 \rightarrow 1.015$; $2.0 \rightarrow 2.169$; $5.0 \rightarrow 4.999$), confirming Theorem 4.2. The overshoot at $\epsilon = 2.0$ is an estimator artefact: the maximum runs over a ± 6 window whose outer bins hold few samples at that budget, and it does not recur under a different seed. The full-field-uniform mechanism's privacy loss is zero only because its output carries no information (Remark 4.1).

6.3 Utility: Governed by ϵ , Not q

Table 2 reports mean per-cell l_1 error at $q = 257$. The proposed mechanism tracks the closed-form $1/\sinh(\epsilon)$ of Theorem 4.3 (9.96 versus 9.98 predicted at $\epsilon = 0.1$, 0.013 versus 0.0135 at $\epsilon = 5.0$). It matches the optimal unwrapped discrete Laplace to within 0.02 and stays below the rounded continuous Laplace at every budget, the margin growing in ϵ as rounding bias dominates (0.013 against 0.082 at $\epsilon = 5.0$). The full-field-uniform mechanism is worst throughout, and its error of roughly 64 does not respond to ϵ .

Table 2. Mean per-cell l_1 error at $q = 257$ over $M = 300$ replications (lower is better)

ϵ	Full-field (original)	Proposed M_q	Laplace (rounded)	Gaussian
0.1	64.47	9.96	10.02	38.47
0.5	64.26	1.93	1.98	7.73
1.0	64.37	0.855	0.962	3.861
2.0	64.27	0.274	0.424	1.922
5.0	64.19	0.013	0.082	0.736

Source: Authors' calculations

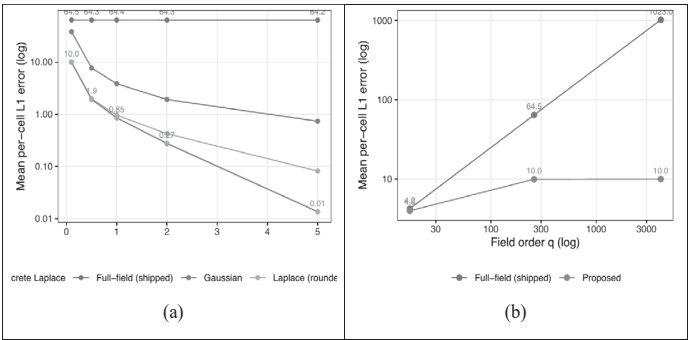
Table 3 fixes $\epsilon = 0.1$ and varies q . The original mechanism's error grows almost linearly with q , from 4.23 at $q = 17$ to 1023 at $q = 4099$, because uniform noise spreads over the whole field. The proposed mechanism saturates at the ϵ -determined floor $1/\sinh(\epsilon)$ once q is large enough that noise rarely wraps: 3.99 at $q = 17$, then 9.96 and 10.01 at $q = 257$ and 4099. This is the regime behind Theorem 4.3: for $q \geq 1/\epsilon$ utility is q -independent; for smaller q the wrap flattens the noise further, preserving privacy while reducing utility toward the uniform floor.

Confidence intervals are narrow and non-overlapping at every budget: at $\epsilon = 1.0$ the proposed mechanism gives 0.855 with 95% interval $[0.848, 0.862]$ against 0.962 with $[0.955, 0.969]$, so

Table 3. Mean per-cell l_1 error at $\epsilon = 0.1$ across field orders

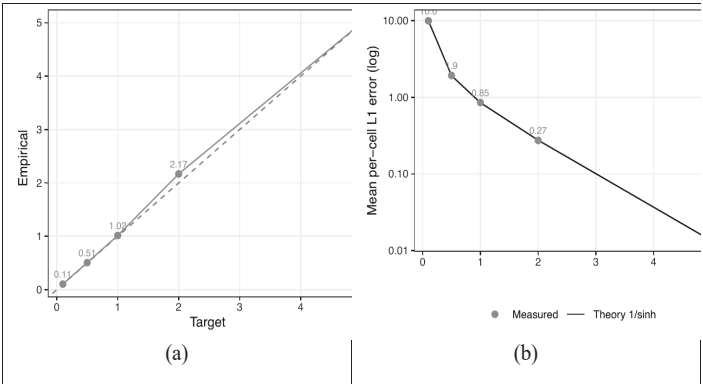
q	Full-field (original)	Proposed M_q
17	4.23	3.99
257	64.47	9.96
4099	1022.99	10.01

Source: Authors' calculations



Source(s): Authors' own elaboration

Figure 5. (a) Mean per-cell l_1 error against ϵ at $q = 257$, 95% intervals over $M = 300$ replications. (b) Error against field order at $\epsilon = 0.1$, log-log axes



Source(s): Authors' own elaboration

Figure 6. (a) Estimated privacy loss against target ϵ ($y = x$ dashed). (b) Measured per-cell error against the $1/\sinh(\epsilon)$ prediction (Theorem 4.3) at $q = 257$

the improvement is significant, not sampling noise (Figure 5). Figure 6 confirms the privacy calibration and the utility law.

6.4 Worked Example: A Circular 09/2020 Cell

Take one cell of the SBV Circular 09/2020 report: the count of Group-3 (Substandard) loans for a single province, say $y = 12$, released with $\epsilon = 1$ over $q = 257$. The mechanism draws $U \sim \text{DL}(1)$; a representative draw $U = -1$ gives $\tilde{y} = (12-1) \bmod 257 = 11$, differing from the true count by the expected $1/\sinh(1) \approx 0.851$ per cell (Theorem 4.3). Because every released cell lies in F_{257} , provincial and national totals reconcile exactly under modular addition across filings, a closure real-valued Laplace and Gaussian releases lack.

6.5 Discussion

Three properties explain the result. The mechanism is unbiased: signed error averages to zero within Monte Carlo precision, $|\text{rm bias}| < 0.05$ per cell in all runs, so averaging M releases gives an estimator whose variance falls as $1/M$ with no bias to correct. It never rounds, so bin-snapping does

not arise. It is closed under modular addition, so releases across reporting periods reconcile exactly in F_q . Together these give the utility of the optimal discrete Laplace plus modular closure, at the pure $(\epsilon, 0)$ guarantee of Theorem 4.2.

7. Conclusion

This paper introduced a field-wrapped discrete-Laplace mechanism for releasing discrete banking aggregates privately. Theorem 4.2 establishes pure $(\epsilon, 0)$ -DP for every field order q ; Corollary 4.4 adds the budgets of k sequential releases without leaving pure DP; Algorithm 2 distributes a total budget across them by a Dirichlet posterior update on observed utility, improving on uniform division.

On synthetic SBV-style aggregates the per-cell l_1 error is $1/\sinh(\epsilon)$, governed by ϵ alone and independent of the field order (Section 6). It matches the utility-optimal discrete Laplace and stays below the rounded continuous Laplace at every budget, 0.855 against 0.962 at $\epsilon = 1.0$, whereas the full-field-uniform design of the original submission was worse by more than an order of magnitude and did not respond to ϵ . Estimated privacy loss tracks the target. The guarantee being information-theoretic, it holds against computationally unbounded adversaries, so only the transport layer needs the standardised NIST FIPS 203 and 204 modules National Institute of Standards and Technology (2024); National Institute of Standards and Technology (2024).

Four limitations bound these results, each naming its next step. All experiments use synthetic tables, because customer-level data from Vietnamese credit institutions were not accessible; validation on actual SBV aggregates requires a regulatory partnership. The sensitivity analysis assumes $\Delta_1 f = 1$, which holds for counting but not for sum, average or range queries, which a finite-field truncated-geometric analogue of Ghosh et al. (2012) would cover. The evaluation spans only $q \in \{17, 257, 4099\}$, so other prime-power families may show different frontiers. The proof is pen-and-paper, so mechanised verification would close the remaining edge-case risk; composition with secure multi-party computation Ishai et al. (2006) by local perturbation at each bank is a further extension.

For policy, Vietnamese banks publish aggregates under SBV Circular 09/2020 State Bank of Vietnam (2020) with no quantitative privacy floor. A stated ϵ turns that obligation into an auditable parameter, strengthens compliance with GDPR Article 25 European Parliament and Council (2016) and EVFTA Annex 8b European Commission and Government of Vietnam (2020), and removes the discretisation step from downstream regression and clustering. The Python and R pipelines, the synthetic generator and the mechanism implementation will be released on acceptance.

References

- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 308–318). ACM. <https://doi.org/10.1145/2976749.2978318>.
- Abowd, J., Ashmead, R., Cumings-Menon, R., Garfinkel, S., Heineck, M., Heiss, C., Johns, R., Kifer, D., Leclerc, P., Machanavajjhala, A., Moran, B., Sexton, W., Spence, M., & Zhuravlev, P. (2022). The 2020 census disclosure avoidance system TopDown algorithm. *Harvard Data Science Review, Special Issue 2*. <https://doi.org/10.1162/99608f92.529e3cb9>.
- Bun, M., & Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Theory of Cryptography Conference (TCC) (pp. 635–658). Springer. https://doi.org/10.1007/978-3-662-53641-4_24.
- Canonne, C., Kamath, G., & Steinke, T. (2022). The discrete Gaussian for differential privacy. *Journal of Privacy and Confidentiality*, 12(1). <https://doi.org/10.29012/jpc.784>.
- Canonne, C. L., Kamath, G., & Steinke, T. (2020). The discrete Gaussian for differential privacy. In *Advances in Neural Information Processing Systems* (NeurIPS) (Vol. 33, pp. 15676–15688). <https://doi.org/10.48550/arXiv.2004.00010>.
- Cuff, P., & Yu, L. (2016). Differential privacy as a mutual information constraint. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 34–43). ACM. <https://doi.org/10.1145/2976749.2978308>.
- Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521), 182–201. <https://doi.org/10.1080/01621459.2017.1389735>.

- Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference (TCC)* (pp. 265-284). *Springer*. https://doi.org/10.1007/11681878_14.
- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy (*Foundations and Trends in Theoretical Computer Science*, Vol. 9). *Now Publishers*. <https://doi.org/10.1561/04000000042>.
- Dwork, C., Rothblum, G. N., & Vadhan, S. (2010). Boosting and differential privacy. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)* (pp. 51-60). *IEEE*. <https://doi.org/10.1109/FOCS.2010.12>.
- Erlingsson, U., Pihur, V., & Korolova, A. (2014). APPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1054-1067). *ACM*. <https://doi.org/10.1145/2660267.2660348>.
- European Commission and Government of Vietnam. (2020). *EU-Vietnam free trade agreement, Annex 8B: Cross-border supply of services and data flows*. <https://trade.ec.europa.eu/access-to-markets/en/country/vietnam>.
- European Parliament and Council. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 25: Data protection by design and by default. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). *CRC Press*. <https://doi.org/10.1201/b16018>.
- Ghosh, A., Roughgarden, T., & Sundararajan, M. (2012). Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6), 1673-1693. <https://doi.org/10.1137/09076828X>.
- IACR. (2025). *Finite-field differential privacy for categorical data* (Cryptology ePrint Archive Paper No. 2025/763). <https://eprint.iacr.org/2025/763>.
- Ishai, Y., Kushilevitz, E., Ostrovsky, R., & Sahai, A. (2006). Cryptography from anonymity. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)* (pp. 239-248). *IEEE*. <https://doi.org/10.1109/focs.2006.25>.
- Kairouz, P., Oh, S., & Viswanath, P. (2015). The composition theorem for differential privacy. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 1376-1385). <https://proceedings.mlr.press/v37/kairouz15.html>.
- Ninh, L. K. (2024). Self-selection out of formal credit markets: Evidence from rural Vietnam. *Asian Journal of Economics and Banking*, 9(1), 105-125. <https://doi.org/10.1108/AJEB-02-2023-0011>.
- Liu, Y., Chen, Y.-M., Chen, L., & Wang, T.-Z. (2021). Fuzzy Bayesian knowledge tracing. *IEEE Transactions on Fuzzy Systems*, 29(10), 3131-3143. <https://doi.org/10.1109/TFUZZ.2021.3083177>.
- Mironov, I. (2017). Rényi differential privacy. In *IEEE 30th Computer Security Foundations Symposium (CSF)* (pp. 263-275). *IEEE*. <https://doi.org/10.1109/CSF.2017.11>.
- Murrar, A., Asfour, B., & Paz, V. (2024). Banking sector and economic growth in the digital transformation era: insights from maximum likelihood and Bayesian structural equation modeling. *Asian Journal of Economics and Banking*, 8(3), 335-353. <https://doi.org/10.1108/AJEB-12-2023-0122>.
- National Institute of Standards and Technology. (2024a). *FIPS 203: Module-lattice-based key-encapsulation mechanism standard (NIST Technical Report)*. <https://doi.org/10.6028/NIST.FIPS.203>.
- National Institute of Standards and Technology. (2024b). *FIPS 204: Module-lattice-based digital signature standard (NIST Technical Report)*. <https://doi.org/10.6028/NIST.FIPS.204>.
- State Bank of Vietnam. (2020). *Circular no. 09/2020/tt-nhnn on confidentiality of customer information at credit institutions*. <https://thuvienphapluat.vn/van-ban/Doanh-nghi/Thong-tu-09-2020-TT-NHNN-bao-mat-thong-tin-khach-hang-to-chuc-tin-dung-438939.aspx>.
- Sultana, R., Ghosh, R., & Sen, K. K. (2021). Impact of COVID-19 pandemic on financial reporting and disclosure practices: empirical evidence from Bangladesh. *Asian Journal of Economics and Banking*, 6(1), 122-139. <https://doi.org/10.1108/AJEB-09-2021-0110>.
- Triastcyn, A., & Faltings, B. (2020). Bayesian differential privacy for machine learning (arXiv:2001.06244). *arXiv*. <https://doi.org/10.48550/arXiv.2001.06244>.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63-69. <https://doi.org/10.1080/01621459.1965.10480775>.
- Zhu, T., Li, G., Zhou, W., & Yu, P. S. (2017). Differentially private data publishing and analysis: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 29(8), 1619-1638. <https://doi.org/10.1109/tkde.2017.2697856>.

Corresponding author

Thanh-Phong Lam can be contacted at: limpaul.fin@gmail.com