Privacy Preserved Gross Domestic Product (GDP) Calculation

**Introduction:**

The Gross Domestic Product (GDP), a pivotal economic metric, encapsulates the monetary worth of ultimate goods and services—those procured by end consumers—fabricated within a nation during a designated interval, such as a quarter or a year [1]. It constitutes the comprehensive summation of all augmented values produced within an economy. The nomenclature "Gross" within GDP signifies the inclusive tallying of products, irrespective of their ensuing utilization. A product's purpose could encompass consumption, investment, or the substitution of an asset.

Diverse methodologies are at one's disposal for GDP computation: 1. The Production Approach 2. The Expenditure Approach 3. The Income Approach. Within the ambit of this discourse, we gravitate towards the Income Approach as the foundational framework for GDP assessment. In the Income Approach paradigm, GDP materializes as the amalgamation of Total National Income (TNI), Sales Tax (SL), Depreciation (D), and Net Foreign Factor Income (F). TNI designates the cumulative assemblage of all earnings that a nation's residents and businesses accrue over a specified duration. While orchestrating GDP (and by extension, TNI) calculations, the imperative of preserving the privacy of each participant in the computational schema stands paramount.

This paper delineates a gamut of methodologies to safeguard the confidentiality of subjects embroiled in this evaluative process. Two distinct methodologies are expounded herein. The inaugural methodology hinges on the aegis of Paillier Encryption, thereby effectuating privacy preservation. The subsequent methodology revolves around the tenets of differential privacy, thereby engendering data confidentiality. To our knowledge, this marks the inaugural endeavor towards the computation of GDP and TNI whilst assiduously upholding the privacy prerogatives of the contributors enmeshed in this intricate calculus.

**Implementation:**

The potency of both global and local economies exerts its influence upon every individual. At the crux of this economic understanding lies the Gross Domestic Product (GDP), a metrical gauge of an economy's expanse, efficacy, and overarching well-being [2]. This quantification transpires both annually and quarterly in the United States. In the context of India, it is subjected to quarterly and annual assessment, with the revelation of each quarter's statistics transpiring with a deferment of two months from the denouement of the respective quarter's operational timeline. The annual GDP dataset is unshrouded on the 31st of May, embellished with a similar two-month delay.

In the United Kingdom, the choreography is slightly divergent. Here, novel GDP data is perpetually being assembled on a monthly basis, although it's the quarterly compilations—a triad of months captured in tandem—that command the broadest spectrum of attention.

A picture containing cartoon, design

Description automatically generated

Fig 1.

The computation of the Gross Domestic Product (GDP) is underpinned by diverse methodological frameworks:

1. The Expenditure Approach
2. The Income Approach
3. The Production Approach

Among these, the Expenditure Approach emerges as the most ubiquitously employed GDP derivation technique, hinging on the pecuniary outlays undertaken by ultimate consumers. This encompasses a multitude of instances, such as consumer disbursements on sustenance, the acquisition of services, corporate investments in industrial equipment, and the procurement of commodities and amenities by both governmental bodies and foreign entities.

The formulation that encapsulates the Expenditure Approach is as follows:

GDP = C + G + I + NX (1)

In this context, the variables bear the following connotations:

* C designates consumption, denoting the collective magnitude of private consumer disbursement within a nation's economic fabric.
* G corresponds to aggregate government expenditure, encompassing disbursements that span remuneration for government personnel, infrastructural undertakings like road construction and repair, allocations to public educational institutions, and military outlay.
* I amalgamate the entirety of a country's investment expenditures.
* NX, symbolizing net exports, signifies the differential between a nation's exports and imports.

The income approach encapsulates the summation of earnings engendered by the provision of goods and services.

GDP (Income Approach): Total National Income + Sales Tax + Depreciation + Net Foreign Factor Income (2)

The Total National Income amalgamates the entirety of wages, rentals, interests, and profits, among other economic inflows.

Sales Tax denotes the levy imposed by the government on consumer expenditures allocated towards goods and services.

Depreciation pertains to the allocation of cost to a tangible asset throughout its operational lifespan.

Net Foreign Factor Income underscores the contrast between the income generated by a country's citizens and enterprises in foreign territories, juxtaposed against the income procured by foreign citizens and enterprises within the domestic sphere. This dichotomy serves as the focal point of this paper's emphasis on GDP computation via the Income approach.

On the other hand, the production approach revolves around the consolidation of value addition across each stage of the production continuum. Here, value addition signifies the aggregate revenue deduced from total sales minus the valuation of intermediary inputs utilized in the production continuum. For instance, flooring is an intermediary input, whereas bread is the ultimate end product.

The computation of Gross Domestic Product (GDP) holds profound significance in appraising a nation's performance. Presently, GDP computation relies on an array of estimations, a practice that may yield outcomes not commensurate with a desired echelon of precision. This discrepancy is intrinsically linked to the notion that individual engagement might not reach the sought-after level of involvement. The reluctance of participating individuals to disclose their private fiscal reserves or earnings to the governmental apparatus constitutes a pivotal factor. The looming specter of punitive measures imposed by authorities for tax evasion amplifies these apprehensions, constituting a principal concern.

A salient solution emerges if the GDP calculation mechanism could furnish a certain degree of confidentiality for individual users. This would entail refraining from ascertaining the entirety of an individual's income while still permitting the submission of income data to the system in segmented parcels, amalgamated with analogous inputs from other system participants. Such a safeguarded confidentiality schema augments the allure of GDP computation, engendering greater enthusiasm among users to divulge their fiscal particulars. To the best of our knowledge, this paper represents the inaugural endeavor in the pursuit of calculating a privacy-preserved GDP.

The subsequent sections of this paper are meticulously dedicated to the exposition of three distinct methodologies for the calculation of GDP, all with an underpinning of privacy preservation. Firstly, the GDP calculation is executed through the utilization of Paillier Encryption. The succeeding segment embarks on GDP computation anchored in the tenets of Differential Privacy (DP). Finally, within the third subsection, our focus sharpens upon the derivation of GDP via the prism of Self-Healed Differential Privacy, an advancement building upon the second phase previously delineated.

**Preserving Privacy in GDP Computation through Paillier Encryption**

Safeguarding privacy assumes paramount significance in shielding users' information from unauthorized exposure. Within this context, the computation of a nation's GDP can be ingeniously executed while upholding privacy, employing the efficacy of Paillier Encryption (PE). PE constitutes a form of partially homomorphic encryption that endows the capacity for additive operations on data encrypted homomorphically. The generation of the encryption key within the PE framework entails the selection of two random numbers, p and q, whereby their product yields n, where n signifies the public key, and p and q correspond to the private key components.

The process of encryption and subsequent decryption transpires through the following delineated steps: For an unencrypted message m, bearing the constraint m<n, a random number r is selected, satisfying the condition r<n, and an element g which is g ∈ Zn2 (integers ranging from 1 to n^2), functioning as the generator.

Cipher Text c is determined by the equation c = gmrn mod n2 (3)

The decryption procedure is conducted as follows: for plaintext m,

m =L(cλ mod n2). μ mode n (4)

Where μ =(L(gλ mode n2))-1 mod n, and λ is the least common multiple of (p-1) and (q-1). Function L is defined as

L(x)=

The additive homomorphic characteristic of Paillier encryption is harnessed, such that: Decrypt (add (Encrypt(m1), Encrypt(m2))) =m1+m2 (5)

This signifies that the decryption of the combined encrypted values equates to the summation of the original unencrypted values they represent.

The researchers adeptly compute a privacy-conserving Gross Domestic Product (GDP) utilizing the additive homomorphic property inherent in Paillier Encryption, as elucidated in (5). The authors introduce the innovative concept of "candidates" as the agents to facilitate this endeavor. Within the system's framework, a variable quantity of candidates is generated, denoted as "n," a numerical parameter constrained to a value beneath a designated threshold, symbolized by "Ө."

For the purpose of clarity, let us denote three illustrative candidates: Alice, Bob, and Charlie. These entities assume the role of receiving fractional segments of participants' income. The disbursements from each participant are randomly apportioned among these candidates. The candidates then collectively deliberate and resolve the frequency at which they would accept these income portions from each participant, subsequently disseminating this information across all participants.

Consequently, participants are empowered to fractionate their fiscal data, with the candidates sharing their public encryption keys with all participants. These segmented values undergo encryption via the Paillier scheme at the candidate side. Among the participants, a singular leader is elected within the cluster. During each time division, this leader undertakes the responsibility of receiving encrypted values from all participants within the group. The leader then accumulates these encrypted values in a cumulative manner. Once the leader receive these values , it perform the additive homomorphic encryption process on their end and subsequently dispatches the amalgamated result to the designated "candidates.".

**Participant 1**

Encryption

Encryption

**Participant 2**

**Participant /**

**Leader**

Encryption

**GDP**

**Candidate 1**

Decryption

Key Generation

Key Publish

**Candidate N**

Fig 2

It's noteworthy that each candidate remains unaware of the specific income values attributed to individual participants. The leader, operating within this framework, remains uninformed about the specific values held by other participants in the group, given that these values are maintained in an encrypted format. After accumulating the contributions from participants over "Ө" rounds, the candidates engage in the additive homomorphic encryption of all encrypted ciphertexts. The resulting composite ciphertext is subsequently decrypted using the private key of the candidate on the client side. The summation of the deciphered values originating from all participants ultimately represents the cumulative aggregate transmitted by the entire cohort of participants.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Alice | Bob | Charlie |
| Participant1 (300) | 80 | 120 | 100 |
| Participant2 (600) | 160 | 200 | 240 |
| Participant3 (450) | 100 | 150 | 200 |
| Participant4 (180) | 100 | 20 | 60 |
| Sum (1530) | 440 | 490 | 600 |

Table 1

The essence of Figure 2 can be encapsulated as follows:

1. Random selection of candidates transpires within the system.
2. Candidates undertake the generation of both public and private keys through the utilization of Paillier Encryption.
3. The public keys devised by candidates are disseminated to all entities within the capsule.
4. Participants congregate into groups, designating a singular leader for each respective cluster.
5. The participants proceed to fractionate their data into "Ө" distinct divisions, each divergence tailored to avoid uniformity.
6. The encrypted data is orchestrated by participants and subsequently conveyed to their designated leaders.
7. Leaders enact homomorphic addition on the accumulated data and transmit the outcome to candidates.
8. Candidates receive and execute a homomorphic addition on this data, ultimately decrypting the GDP component.
9. The summation of values across all candidates constitutes the GDP.

**Preserving Privacy in GDP Computation Through Differential Privacy**

The utilization of Differential Privacy (DP) engenders an indistinguishability property, offering an ingenious safeguard. This property manifests through the perturbation of query outcomes in such a manner that an inquirer is rendered incapable of discerning the presence or absence of individual data, let alone the specific individual data itself.

Within the purview of DP, two datasets, denoted as "D" and "D̅," are considered neighbors if they solely diverge in a solitary row. A query "Y" exhibits ε-differential privacy across all conceivable outcomes "y" and all neighboring datasets "D" and "D̅" in compliance with the following expression:

Pr[Y(D)=y] ≤ eε \* Pr[Y(D̅)=y] (3)

Here, the symbol ε embodies the quantum of privacy fortification. Equation (3) delineates that the probability of an algorithm generating a specific output "y" on dataset "D" remains at most eε times the probability of producing the identical output on dataset "D̅." Should the system involve a participation cohort of "n" individuals, then "D" and "D̅" ∈ Rn \*Rt, wherein "t" represents the number of partitions.

Introducing the concept of sensitivity "S" for a function f: Rn \* Rt -> Rt, where

S(f): = max || f(D) - f(D̅) || (4)

The notion of sensitivity, within the ambit of differential privacy, represents the magnitude by which output data can maximally alter upon the inclusion or exclusion of a solitary individual's data. It quantifies the influence of a singular participant within the dataset. A lower sensitivity value corresponds to a diminished potential for deriving information pertaining to individual data from the ultimate output. Here, "||. ||" signifies the distance matrix employed to quantitatively assess the variance between function outputs for neighboring datasets. The sensitivity metric stands as an amalgamation of individual partitions across temporal index "t."

S(f) = (5)

The Laplace mechanism constitutes a technique harnessed for introducing perturbations into the output within any given time interval "t," ensuring the preservation of differential privacy. This mechanism operates through the introduction of noise via the Laplace distribution.

The Laplace distribution, an essential component, is a probabilistic distribution characterized by a density function represented as follows:

Lapλ(x)=exp (- ) (6)

Within this context, λ assumes the role of the introduced noise magnitude. The amplification of λ equates to an augmentation in the amplitude of the incorporated noise. This scholarly exposition undertakes the procedure of obfuscating actual participant data by cloaking it with Laplace noise, thereby shrouding authentic GDP values from external observers. Here, individual participants undertake the autonomous task of imbuing the noise, ensuring that external observers remain oblivious to this manipulation, a process entirely devoid of interdependence on fellow participants within the system.

This pursuit is actualized by each participant, at any given juncture "t," infusing noise extracted from the Gamma distribution. Notably, the Laplace distribution can be disassembled into a sequence of distinct individual distributions. This deconstruction is represented as:

Lapλ(x)= =: Gλ  (7)

Here, G1 and G2 denote two mutually independent and identically distributed Gamma distributions, both characterized by shape parameter and scale parameter λ.

During each discrete time instance denoted as "t," each participant takes the initiative to incorporate gamma noise, as outlined in (7), to their respective GDP values, subsequently transmitting this augmented data to the aggregator. The aggregator then undertakes the task of summing all the data points received at time "t." Consequently, the resultant value for participant "i" at time "t" materializes as follows:

Y I,t =Xi,t = Xi,t Gλ (8)

Thus, for any given temporal point "t," the summative value across all participants crystallizes as:

Yt= Yi,t (9)

Leveraging the expression in equation (8), we can refine equation (9) to read as follows:

Yt= xi.t + (10)

By invoking the formulation delineated in equation (7), equation (10) assumes a modified structure as depicted below:

Yt= xi.t + Lapλ (11)

Hence, the central idea is to incorporate Gamma noise during each temporal interval, ensuring that the authentic partitioned GDP values remain concealed from potential intruders.

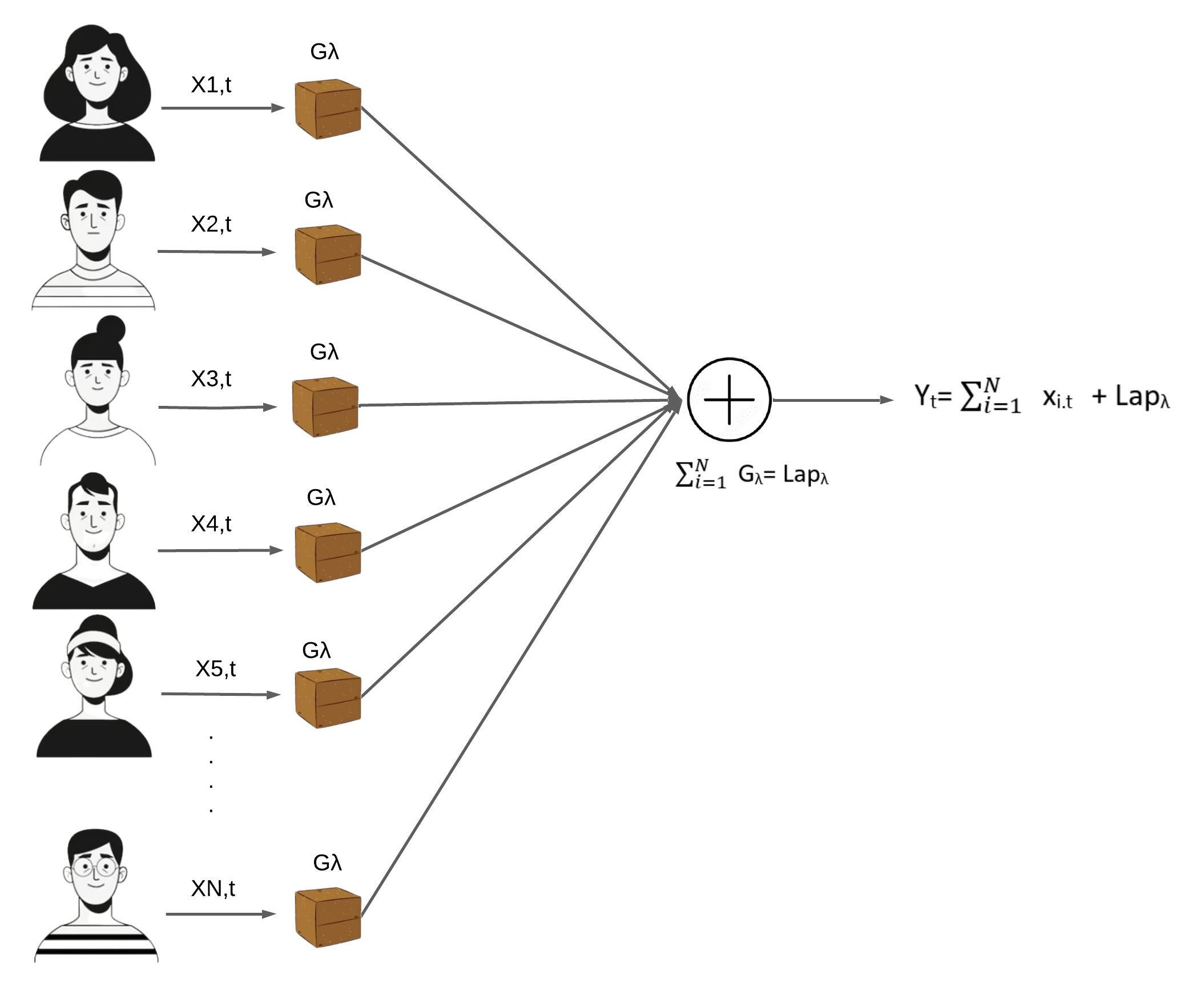


Fig 3

The computation of GDP is expressed through the following formula:

Calculated GDP = Yi (12)

However, the computed GDP exhibits divergence from the authentic counterpart by a factor of t multiplied by Laplace noise, symbolized as Lapλ. Acceptance of this computed value is contingent upon its divergence being within the error threshold of 5%. If it adheres to this error margin, the value is deemed suitable for utilization within calculations. The discrepancy between actual GDP and the perturbed counterpart is graphically elucidated in Fig. 4. The ensuing depiction encapsulates the resultant graph while adopting a scaling parameter of 0.5. This parameter selection ensued from a series of iterative trials involving the available data, culminating in the determination of an apt scaling value that effectively obfuscates the genuine GDP data at the temporal juncture "t."

A graph with orange and blue lines

Description automatically generated

Fig 4

Presented below are the juxtaposed GDP values in contrast with the calculated perturbed GDP values:

A close-up of numbers

Description automatically generated

An astute observation reveals an insignificant 0.1% discrepancy, which is comfortably contained within the prescribed permissible margins.

In the context delineated in Figure 3, we ascribe the nomenclature "capsule" to the assembly of individuals involved in the computation. Should "P" denote the populace of the nation, and "N" represent the size of the capsule, the entirety of the population can be partitioned into P/N capsules. Once the differential GDP of each capsule is ascertained, these capsules congregate to constitute the terminal nodes within a tree structure. Employing the principles expounded in the Differential Privacy calculations as referenced in [11], these values undergo recursive processing within the hierarchical tree, persisting until they ascend to the root node. The eventual outcome residing within the root node attains the coveted status of privacy-preserved GDP for the nation, an entity meticulously visualized in Figure 5.

A diagram of a diagram

Description automatically generated

Fig 5

**Privacy-Preserved GDP Computation through Self-Healed Differential Privacy**

An inherent limitation of the previously expounded differential privacy approach lies in the cumulative addition of Lapλ noise with each temporal partition. Consequently, for "t" time divisions, a cumulative total of "t \* Lapλ " noise is superimposed upon the computed GDP value. While its acceptance remains permissible when adhering to the 5% threshold, this method warrants exploration for its inherent shortcomings. In response, this paper introduces, substantiates, and validates a novel paradigm aimed at obviating noise entirely. Simultaneously, this advanced framework endeavors to yield a pristine GDP calculation that is devoid of errors, while unwaveringly safeguarding the confidentiality of individual participants ensconced within the system.

The underlying notion revolves around a gradual self-healing process aimed at mitigating the errors previously incurred, thereby culminating in the precise transmission of user values upon reaching the culmination of the temporal division denoted as "t." The solution suggests detecting the noise introduced at a particular moment and subsequently refining and rectifying the noise in successive steps. When the noise, represented as "Ө," is introduced at a given stage "l," the forthcoming steps accommodate meticulous adjustments within the applied noise, effectuating a methodical correction process.

For elucidation, if the error "Ө" is determined to undergo self-healing across "s" sequential steps, then the self-healing mechanism computes the cumulative mean of errors from the preceding "s" steps, subsequently harmonizing it with the ongoing step. The algorithm encapsulating this procedure is articulated as follows:

**Input**: Noise "Ө" induced at step "l," self-healing across "s" steps.

1. Compute the mean error "μ" over the last "s" steps utilizing the formula:

μ = (Өl-s + Өl-s+1 + Өl-s+2 + …. Өl-1 + Өl )/s , excluding the “t”’th time division in GDP calculation.

1. Evaluate the value of "μ" at step "t," representing the final time-divided step in GDP calculation, employing the formula:

μ= -1 \* (Өt-1 + Өt-2 \* (s-1)/s+ Өt-3 \* (s-2)/s + Өt-4\* (s-3)/s+ ….+ Өl \*2/s+ Өl /s)

Through this algorithmic framework, errors are systematically rectified and integrated across time divisions, ultimately leading to an enhanced accuracy in preserving user values while maintaining the desired privacy constraints.

The "μ" value is harmonized with the "Ө" noise introduced at step "l," resulting in a noise manifestation of "Ө - μ" at each step. Consequently, the cumulative effect of added noise nullifies itself over "s" consecutive steps, thus yielding a state of noise cancellation.

Presented herewith are the depictions showcasing the authentic GDP values alongside the GDP values that have undergone the process of self-noise healing:

A number of numbers and letters

Description automatically generated

Evidently, the distortions incurred through the introduction of differential privacy are rendered null, signifying the successful self-healing of errors, culminating in the pristine calculation of GDP, free from discrepancies.

Furthermore, the graphical representation in the subsequent chart offers a visual contrast between the genuine user values and the values that have undergone the self-healing process:

A graph with blue and orange lines

Description automatically generated

Fig 6.

The discernible divergence between the two sets of values in Figure 6 affirms the effective masking of user data, underscoring the pivotal role of this approach in upholding users' privacy.

**Threat To Validity:**

The researchers presented several methodologies for computing privacy-preserved GDP, with the computations conducted using test data. However, the practical application of these techniques to actual country-level GDP assessment remains unrealized due to inherent limitations in accessing individual-level data.

**Conclusion:**

In essence, the Gross Domestic Product (GDP) serves as a fundamental economic indicator, encapsulating the value of goods and services produced within a nation over specific timeframes. This paper delves into the intricacies of GDP calculation, particularly focusing on the Income Approach. It explores innovative methods to ensure the privacy of participants in this computation, presenting techniques involving encryption and differential privacy. This endeavor is groundbreaking, marking the first attempt to calculate GDP and related values while safeguarding contributors' privacy.

**Source Code Available At:**

<https://github.com/sanjaikanth/PrivacyPreservedGDP>

**References:**

[1] <https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/gross-domestic-product-GDP>

[2] <https://online.hbs.edu/blog/post/why-is-gdp-important>