International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 8, August-2018

A DEXTEROUS INEQUALITY QUERY AUDITING WITHOUT REFUTATION THREAD USING CASTLE TECHNOLOGY

K.Vinotha[#1], Dr. G.Kesavaraj[#2],

#1M.Phil Scholar (Full-Time), #2 Assistant Professor,
Department of Computer Science,
Vivekanandha College of Arts and Sciences for Women, (Autonomous)
Tiruchengode, Namakkal-DT, TamilNadu, INDIA

Abstract

Cloud computing primarily based technology and their impact is growing currently each day. Cloud computing is employed altogether the realm of business, Education, Social Impact and data miner. The uses enhance the danger of knowledge saddlery and forgery. Because the services square measure sharing at intervals totally different cloud users. That the security problems in cloud computing surroundings square measure a significant concern. Because the users all have confidence the cloud vendors, thus there's a requirement protection and therefore the data management from the cloud users. To beat this downside, we tend to introduce a replacement model that's CASTLE info auditing, which associate degree difference query is auditing theme that evaluates the danger of respondent supported the query history. Moreover, we tend to proposes relax CASTLE to extend the utility by returning answer with slight perturbations.

Keywords: Castle, Relax castle, privacy, auditing, query denial, optimization.

I. Introduction

Information technologies are extensively accustomed collect and share personal knowledge among varied parties via the cloud. In areas comparable to aid analysis, crime analysis, client relationship management, credit analysis. It is essential to extend a protected knowledge sharing, accessing, and commerce mechanism. Taking the assurance associated answerable Act (HIPAA) as an example, to avoid re-identification, entities should take away or perturbs a minimum of [18] alphabetic character knowledge parts after they share sensitive knowledge. Many works are projected that explore the trade-off between utility and privacy. Their final goal is to guard individual's nonpublic info from unauthorized access.

One possible answer that was addressed by Adam and Worthmann is query auditing. We tend to get back the classic query auditing within the cloud platform that serves the confuse as associate auditor.

Query auditing refers to the method of auditing whether or not respondent a new predictable query can foundation privacy compromise once given a sensitive dataset similarly as a sequence of antecedently answered queries on one attribute and also the corresponding answers.

We initial study difference query auditing, where the query is of the form f $(\tilde{X})\leq$?a.$\tilde{X}$ Could be aset of the sensitive dataset, and also the answer is either 'yes' or 'no'. The perform f will be any polynomial time calculable perform (e.g., logistic, linear regression), that is additional general than the combination query.

The dataset of n individuals defines a n-dimensional house, wherever every individual corresponds to at least one dimension. Observe that associate variance query over this dataset, as well as its answer ('yes' or 'no') defines a n-dimensional valid region. Given a collection of previous requests and their answers, associate offender will slim down the doable values for the dataset by computing the intersection of all valid regions similar to all answered queries (called the answer house to any or all such queries). Then a naïve auditing mechanism is to deny this query if there's a singular answer for all historical inequalities with this one (which reveals that the dataset is re-identified).

In this paper, we tend to propose CASTLE that with success audits the difference queries while not denial threats. To audit a received query, we tend to estimate the amount of the posterior answer house. To enhance the utility, that is outlined because the variety of answered queries, we tend to additional contribution a relaxed CASTLE by adding errors to the answers. To keep up the service quality (introducing too several errors seriously degrades the quality), the error budget is outlined because the add of the additional errors. Increasing the quantity of answered queries underneath the constraint of the error budget, involves associate other complex improvement task–online max with an unknown query distribution. Moreover, we tend to still fix the privacy problems with differential privacy, with the aim of utterly protective the knowledge that's provided by the information house owners. Finally, we tend to show that our methodology will audit the mixed queries (min, max, and sum) and isn't restricted to the difference query via another extension.

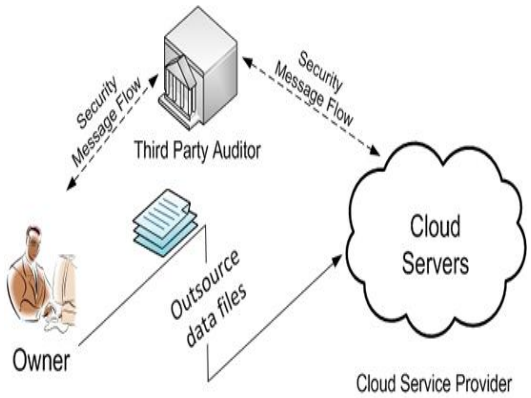

Fig 1.1: Query auditing system

II. Literature Survey

Within the age of huge data, increasing quantities of uncommunicative facts could also be finished, inclusive of that in [1]. A SDB generally refers to a information used for applied math analysis functions. SDB contains knowledge at the individual record level; users area unit generally solely allowed to raise queries over aggregates. To guard the privacy of info that will be sensitive at the individual record level [2].They embody auditing queries area unit query restrictions, cell suppression [3], [4]. Providing approximate answers and anonymous knowledge assortment [5],[6].Absolute security by checking the query history, auditing permits USA to answer a query only if it's secure to try and do therefore [3]. Simulatable audit examine a replacement query only supported past query answers while not consulting the information, this theme effectively prevents the query denial threat, [7]A relaxed theme at every auditing time, an huge range of possible information solution, that area component trustworthy to the past query answers.

This theme has 2 limitations they're computationally expense, no guarantee of security [7].Sum-only queries that we have a tendency to decision changed typical auditing. At every auditing time, it first of all computes the bounds of the solution to the new query by inspecting past answered queries. Then it computes the bounds of every information variable by inspecting past answered queries and also the derived bounds of the solution to the new query [8]. once responsive every query accurately, Reiss [13] proved the higher and lower bounds on least amount range of answered queries to compromise the dataset, and acknowledged that no dataset are often compromised in (2k-(1+1)/r), whatever the query size is outlined as k, every query overlaps by r parts at the most, and also the values of l parts area unit best-known prior to. In our work, rather than quantifying a lowest amount range of answered queries, we have a tendency to apply perturbation [14]–[15] to more improve the higher and lower bounds on the amount of answered queries. Not like in differential privacy [16], if we have a tendency to perturb our answer, the perturbation can solely enlarge the doable resolution house of a variable; it will not generate the answer incorrect. During this paper, we have a tendency to study CASTLE auditing that with success audits the difference queries even as not denial threats. To audit a received query, we have a tendency to estimate the amount of the posterior assessment house. We have a tendency to stringently go with the first premise of auditing: unceasingly monitor the user's data that's consequential from responses to past queries and use it to work out the method to answer to a replacement query.

However we have a tendency to area unit aware that responses embody each query answers and denials.

III. CASTLE: Inequality Query Auditing

3.1 CASTLE Overview

Makes "cloud first" doable for federal agencies. it develop fashionable application higher and quicker. And cloud computing transforms the legal relationship between people and their individual proceedings. Federal-having (or relevancy a system of states inside that many countries of a unity but keep free-lance in internal associations.

With reference to (or) denoting the central government as distinguished from the separate units constituting a federation. To solve the query auditing drawback, we have a tendency to propose the CASTLE algorithmic rule. The essential strategy is to envision.

Whether or not the posterior answer house is finite by the safe zone (i.e., the quantitative relation of $S_s^t \cup S_z$is large) via sampling. During this approach, our theme is applied to any polynomial computation query. Our algorithmic rule is freed from denial threats since the answer residence won't be narrowed right down to some extent and can be finite from below by the safe zone.

3.2 CASTLE Description

Specifically, for any private dataset $X = \{x_{,\ldots\ldots,}x_n\}$, to come to a assessment whether or not the recently received query q_t is answered, we want to envision whether or not the likelihood P r, which is defined in Eq. 1, is less than $1 - \delta$. If $P r \geq 1 - \delta$, q_t will be correctly answered; otherwise, it will be denied.

One intuitive strategy for estimating the likelihood is to calculate the degree ratio$\frac{|Sz \cap S_s^t|}{|Sz|}$, that is, to whether or not the answer house will embody the whole safe zone. But, existing volume estimation algorithms rectangular quantity uniquely applicable to H-polytopes and P-polytopes, and also the dimensions of intersection space $|Sz \cap S_s^t|$ cannotbe there directly sculptured as AN H-polytope (defined as AX = b) or a P-polytopes (defined as $AX \leq b$) if we have a tendency to enable any polynomial-time assessable operate. Shoddier, precisely estimating the degree of a broken-backed polytope has been experienced to be #P-hard by Expert worker and Frieze. Hence, we have a tendency to calculate the degree of the intersection house and also the safe zone, and approximate the degree with the technique of sampling.

We first introduce AN oracle that may tell whether or not some extent is within a polytopes or not. Then, we have a tendency to sample sufficiently several points inside a bounding space with better-known volume, i.e., safe zone, to estimate the degree quantitative relation by asking the oracle whether or not some extent is additionally inside the intersection house. During this approach, our mechanism is appropriate to several polynomial-time calculable operate of the shape q_t: f $(\tilde{X}) \leq ?$ a_t. To sample sufficiently several points inside the safe zone, we have a tendency to implement the method Dimensions, that is represent in [17]. It samples suitably Sizable amount of point's in an additional ordinarily broken-backed body to estimate the quantitative relation. For more rapidly convergence, we have a predisposition to conjointly be appropriate the operate spherical on any broken-backed body that's not spherical enough. later on, rather than honestly sampling points from the same distribution, which needs AN exponential variety of samples, the same as [17] we have a tendency to construct a cooling schedule, that is denoted the same the same as Get Annealing Schedule, trustworthy with that we have a tendency to sample from a distinct distribution f_i in every part, and its variance slowly will enhance to 1, at that opinion it's fundamentally the uniform distribution. Furthermore, we encompass a use the Hit And Run algorithmic rule to get a hold samples in each ingredient [18].

Algorithm 1: *Estimation of* ***Pr*** *(Probability)*

Input: $S_z = \{l_1 \leq x_1 \leq u_1, \ldots, l_n \leq x_n \leq u_n\}$, and S_s^{t-1}, $<q_t, a_t>$; n, ϵ, a_O, ratio, r_{steps}, W;

Output: $Pr \leftarrow \frac{|N^t|}{|N|}$

Step 1*: Let* $N_t = \emptyset$ *and* $N = \emptyset$

Step 2: *Update* S_s^{t-1} *to* $S_s^{t'}$ *with* $<q_t, a_t>$

Step 3: T = *Round* (S_z, r_{steps})

Step 4: *Set* $S_z' = T \cdot S_z$ *and* $S_s^{t'} = T \cdot S_s^t$

Step 5: $<f_o, \cdots, f_m>$ = *GetAnnealingSchedule* $(S_z', a_O, ratio)$

Step 6: *for* $i = 1$; $i < m$; $i++$ *do*

Step 7: *Set converged = false*

Step 8: *while converged = false do*

Step 9: *Sample P based on HitAndRun* (K', f_{i-1})

Step 10: $N \leftarrow N \cup P$

Step 11: *if P is within* $S_s t'$ *then*

Step 12: $N_t \leftarrow N_t \cup P$

Step 13: *end if*

Step 14: *converged = Checkconverged (€/m, W)*

Step 15: *end while*

Step 16: *end for*

Step 17: *Return Pr = |Nt|/|N|*

Operate Check converged is named to envision whether or not the points within the broken-backed body commencement a distribution square measure more or less proportional to intermediate operate f_i Additional information concerning these algorithms is found in [17] when sampling sufficiently several points inside the safe zone, we have a tendency to check whether or not sampled purpose is inside the intersection of this answer residence (modeled by S_s^{t-1}) and also the topological space (formed by the freshly received query q_t with its correct answer at). The algorithmic rule that's wont to estimate Pr is specifically delineate in rule one, wherever S_z is that the pre-defined safe zone, and therefore there solution area is shapely by S_s^{t-1}. ϵ is that the target relative error fraction between the calculable volume and therefore the actual volume, and each one a_O and magnitude relation area unit parameters that area unit relating to the tempering schedule.

The establishment execute could be a distribution with variance $1/2a_O$, and magnitude relation is that the cooling magnitude relation. r_{steps} is that them is calculation steps, and W is that the size of the window that's accustomed live the convergence.

3.3 Performance Analysis

3.3.1 Parameter Selection

According to [9], it's been tried that $Pr(D / \in Sz) < \epsilon$ when we set $aO = (n+\sqrt{8n \ln(1/\epsilon)})/2$ for Associate in Nursing calculable convexo-convex body Sz and a willy-nilly sampled purpose D. Moreover, let D be a random purpose in Sz with a likelihood density that's proportional to $e^{-ai\ ||x||2}$, $ai+1 = ai \times ratio$, wherever quantitative relation $= 1 - 1/n$, and $n \geq 4$. we will estimate every quantitative relation in every part error $\epsilon/2\sqrt{m}$, where m is that the total range of phases.

3.3.2 Collusion Resistance

When δ approaches zero, the intersected answer house for any somebody is physically delimited from below by the safe zone, even once there's collusion. Once δ is accrued, we have a tendency to analyze the worst-case state of affairs once there's collusion. That is, for any fastened δ, we have a tendency to analyze the minimum range of adversaries that square measure needed to reveal the personal dataset.

3.3.3 Denial Threat Removal

Our technique may also handle denial threats. Denial threats occur once the denied query, together with antecedently answered queries, is often accustomed re-identify a particular worth. However, in our technique, while not the right answers, the 'denied' query can not be accustomed subdivide the answer house. Thus, denial isn't any longer a helpful answer. to Illustrate, the new received query qt : P(x1, x2, x4, x7) ≤? at are going to be denied below our mechanism once the likelihood P r is a smaller amount than 1−δ. during this approach, the answer house are often divided by neither $\sum(x1, x2, x4, x7) \leq$ at $\sum(x1, x2, x4, x7) \geq$ at since the info shopper doesn't understand whether or not the right answer is 'yes' or 'no'.

3.3.4 Query Type

Since we have a tendency to don't have to be compelled to estimate the amount of the answer house, our mechanism are often effectively applied to any polynomial-time calculable operate, that's of the form f ($\widetilde{X}$)≤?at.

3.3.5 Extension to More Attributes

Historically applied math question considers one specific attribute. To multiple attributes, we are going to build a secure zone that corresponds to every specific attribute (e.g., age, salary). For a question that's computed with one specific attribute (e.g., age), we are going to attempt to confirm whether or not respondent will lead to a privacy breach, i.e., the age-related answer house can now not embody the safe zone (the likelihood is a smaller amount than 1-δ).

3.3.6 Dynamic Dataset Consideration

Our theme will still reaches good performance once the worth of some components within the dataset square measure modified if the pre-defined safe zone still includes the new dataset (which is mapped as a high dimensional point). However, if ever-changing the dimensions of the dataset, another safe zone has to be engineered, and our theme are going to be supported the new safe zone.

Algorithm two is predicated on Algorithm 1, wherever the answer area is barely updated with qt and its correct answer once P r ≥ one − δ; otherwise, it remains an equivalent. In Algorithm two, it takes the subsequent as input: the query and its answer <q_t, a_t>. If the proper answer to a query, e.g., f ($\tilde{X}$)≤? a, is e.g., 'no', then it is stored in the form e.g., <−f(X), −a> or <f(X), a> with correct answer 'yes'. The safe zone is denoted as S_z; the answer area is shapely by $S_s^{t'}$; The dataset is denoted as X and its size is n; and the privacy parameter δ is defined in Definition one.

Let $a_O = (n + \sqrt{8n \ln(1/\in)})/2$, ratio = 1 − 1/n, $r_{steps} = 8n^3$ and $W = 4n^2 + 500$. The output is the correct answer 'yes' or 'no' or 'denial'.

Algorithm 2: CASTLE
Input: *$X = \{x_1; \ldots, x_n\}$;*
S_z *= $\{ l_1 \leq x_1 \leq u_1; \ldots, l_n \leq x_n \leq u_n\}$; and $S_s^{t'}$; f_t; n; δ ; ∈; a_O; ratio; r_{steps}; W;*
Output: *'yes/no' or 'denial'*
Step1: *Get <q_t; a_t> from f_t;*
Step2:
Step3: *Estimate P_r with following inputs via*
Algorithm 1: *S_z, S_s^{t-1},q_t; a_t; n;∈; a_O; ratio; r_{steps}; W;*
Step4: *if Pr ≥1-δ then*
Step5: *Update S_s^{t-1} to S_s^t with <q_t; a_t>*
Step 6: *Return 'a_t'*
Step7: *end if*
Step8: *Return `denial'*

IV. RELAXED CASTLE: UTILITY MAXIMIZATION

4.1 Relaxed CASTLE Overview

Briefly, any fresh received query are 1st audited by CASTLE. If P r (D $\in S_s^t$|D $\in$ S_z) $\geq 1 - \delta$, then the auditing takings usually. Otherwise, q_t is denoted as Associate in nursing insecure query, and that we can appreciate the minimum perturbation that the likelihood is larger than $1 - \delta$. After that, we are able to improve the utility by respondent this insecure query with a small perturbation that consumes the error budget. Since the error budget is proscribed to E, ideally, forever we must always always opt to answer the insecure queries that need the littlest perturbation for the simplest utility. If the sequence of all queries were well-known ahead, it'd be simple to settle on the optimum set of insecure queries for maximizing the effectiveness ψ although agreeable the error budget constraint (by avariciously and iteratively selecting the insecure query with the littlest error demand). However, the sequence of all queries isn't well-known ahead. Hence, once given the minimum perturbation of Associate in Nursing insecure query, we have a tendency to raise Expert to make a decision whether or not to perform perturbation or deny the queries to maximize the utility (approximates the optimum set) at intervals the error budget limitation.

Algorithm 4: RELAXED CASTLE
Input: *$X = \{x_1, \cdots, x_n\}$;*
$S_z = \{l_1 \leq x_1 \leq u_1, \cdots, l_n \leq x_n \leq u_n\}$;
and S_s^{t-1},ft; <q_t, a_t>; n, δ, ϵ_s, a_O, ratio, r_{steps}, W;
Output: *'answer' or new answer with a'$_t$ or 'denial'*
Step 1: *Obtain <q_t, a_t> from f_t;*
Step2: *Evaluate Algorithm 1 with inputs X, S_Z , S_S^{t-1},f_t, n, ϵ_s, a_O, ratio, r_{steps}, W;*
Step3: *if answer is 'answer' then*
Step4: *Update S_S^{t-1} to S_s^t with <q_t, a_t>*
Step5:*else*
Step 6: *Resort to Expert with input e_t, t, τ, C_c, E, LB, UB*
Step7:*where e_t = min {e|a'$_t$ ← a_t + e; Pr˜ ≥ 1 – δ }*
Step8: *if Expert outputs 'answer' then*
Step 9: *Update S_s^{t-1} to S_s^t with <qt, a'$_t$>*
Step 10: *Return a'$_t$, where a'$_t$ = a_t + e_t,*
Step 11: *end if*
Step 12: *end if*
Step 13: *Return 'denial'*

If the sequence of all queries were well-known ahead, it'd be simple to settle on the optimum set of insecure queries for maximizing the effectiveness ψ although agreeable the error budget constraint (by avariciously and iteratively selecting the insecure query with the littlest error demand). However, the sequence of all queries isn't well-known ahead. Hence, once given the minimum perturbation of Associate in Nursing insecure query, we have a tendency to raise Expert to make a decision whether or not to perform perturbation or deny the queries to maximize the utility (approximates the optimum set) at intervals the error budget limitation. Next, we have a tendency to gift the implementation of Expert.

Expert 1st learns the query distribution by perpetually acceptive unconfident queries within the terribly starting. When an exact amount, it decides whether or not to Associate in Nursingswer or deny an insecure query by scrutiny the rewards of respondent or denying supported the query distribution.

4.2 Performance Analysis

Our mechanism is still free of denial threats since a data consumer cannot infer whether the privacy compromise is caused by answer 'yes' (i.e., $f(\tilde{X}) \leq a$) or 'no' (i.e., $f(\tilde{X}) \geq a$). Specifically, even if the answer is a slightly perturbed, e.g., $f(\tilde{X}) \leq a+e$ or $f(\tilde{X}) \geq a-e$, the exact answer ('yes' or 'no') to the original query cannot be determined, e.g., f(~X) ≤?a since the dataset is mapped as a point, which is randomly within the safe zone (not in the center). Therein, the perturbed answer is not determined by the exact answer but by the distance between the safe zone and the received query $f(\tilde{X}) \leq a$.

4.2.1 Differential Privacy

As is well known, differential privacy guarantees that the presence of one individual data instance will not influence the probability of the distribution of the results based on the rest of the data.

4.2.2 Sensitivity and Privacy

Since differential privacy is a notion for protecting a sensitive dataset, we should first define the sensitive dataset in this scenario.

The sensitive dataset is defined as $D = \{s_1, \cdots, s_n\}$, where one record (whose value is $s_i = r_i - l_i$) reveals the sensitivity of one data instance that is provided by a different owner. We use a mechanism with differential privacy to re-define the safe zone where the length of each dimension is si + ni , and ni is a noise that is randomly sampled from a distribution where one of the most important scale parameters is Δf. Here, Δf is the sensitivity and is defined as:

$$\Delta f = \max \| f(D_1) - f(D_2) \|_2 ,$$

where D_1 and D_2 are two arbitrary neighboring datasets that differing on one data instance, and $k \cdot k_2$ is the L_2-norm of a vector. In this paper, D_1 and D_2 are two arbitrary subsets of D, and are different in only one specific record (assuming that there are d data in D_1). Moreover, $f(D_1)$ (or $f(D_2)$) is the sum of the true d values. Hence, the sensitivity is defined as the maximum value of the whole dataset D.

4.2.3 Truncated Gaussian Mechanism

To ensure that the pre-defined safe zone (defined by the data owners) is always protected, the new safe zone that is generated by the auditor must be larger than the original. That is, we must add noises that are greater than zero to the safe zone. Since noises that are sampled from a Gaussian distribution (from negative infinity to positive infinity) cannot satisfy this requirement, the gaussian mechanism cannot be applied in this case. We propose another mechanism, called the truncated Gaussian mechanism, in which the noises are sampled from a truncated distribution. The truncated Gaussian distribution, where all noises are drawn from a group that obeys the Gaussian distribution and has a value that in the range a < u < b, obeys the formula

$$\frac{g(v;\mu,\sigma,a,b) = 1\ \sigma\ \varphi(\ v-\mu\ \sigma\ )}{\Phi(\ b-\mu\ \sigma\ ) - \Phi(\ a-\mu\ \sigma\ )},$$

Where $\varphi(v) = \sqrt{}\ 1\ 2\pi \exp - 1\ 2\ v\ 2$.

Here, we set a = 0, b = ∞, and μ = 0.

Since b =∞, the cumulative distribution

$$\sigma\ (\frac{b-\mu}{\sigma}\ )= 1$$

Proving that the truncated Gaussian mechanism also satisfies ($\in_d$, δ_d)−differential privacy is equivalent to proving that the probability of privacy loss (greater than $\in_{d)}$ is less than δd. The privacy loss here is:

$$\frac{g(v;\mu = 0,\sigma,a,b)}{g(v + \Delta f;\mu = 0,\sigma,a,b)} = |\ \ln \frac{(\ e\ (-1/2\sigma\ 2\ )v\ 2}{e\ (-1/2\sigma 2)(v+\Delta f)\ 2\ )|},$$

which is the same as in the Gaussian mechanism . Hence, we can easily obtain that the truncated Gaussian mechanism.

4.2.4 Compos ability

Under our scenario, our mechanism suffers from repeated applications: the auditor repeatedly calls the mechanism with (ϵ_d, δ_d)−differential privacy to generate a safe zone (one dimension is generated per call). Then, after several applications, from the results, the adversary might have enough information, such that the mechanism can no longer satisfy (ϵ_d, δ_d) −differential privacy. According to the composition theorem , the sequential applications of the mechanism can satisfy (ε, D)−differential privacy when an auditor sets up a sequence of privacy budget for each mechanism $\epsilon_i = \epsilon/n$, $\delta i = D/n$ (where n is the number of dimensions of each safe zone).

V. Evaluation

In this section, all of our schemes area unit evaluated in 2 aspects: potency and Utility. One in all the classic works on question auditing is [19]. Once this work, few works concentrate in implementing a interrogation auditing mechanism; instead, they study the theoretical results with relevance classical compromise (privacy breach once one worth is pinpointed). Thus, during this work we have a tendency to solely compare our work with simulatable auditing as (hereina after noted SIMULATABLE). Our experiment relies on a Chicago worker earnings dataset, which has the names of staff and their salaries. The progressive techniques on volume estimation area polytope. Unit solely appropriate for estimating high-dimensional bodies that area unit portrayed as H-polytope or P- Therefore, during this work, we have a tendency to solely value some sorts of difference queries, admire add and liquid ecstasy.

CASTLE VS. SIMULATABLE

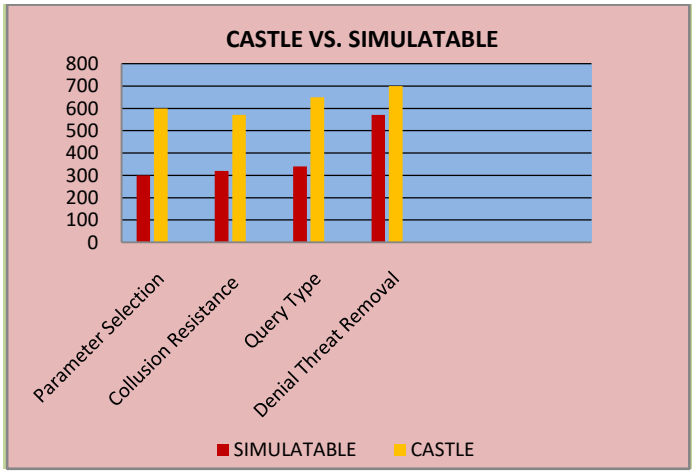

Fig 5.1: Castle vs. Simulatable

CASTLE VS. Relax CASTLE

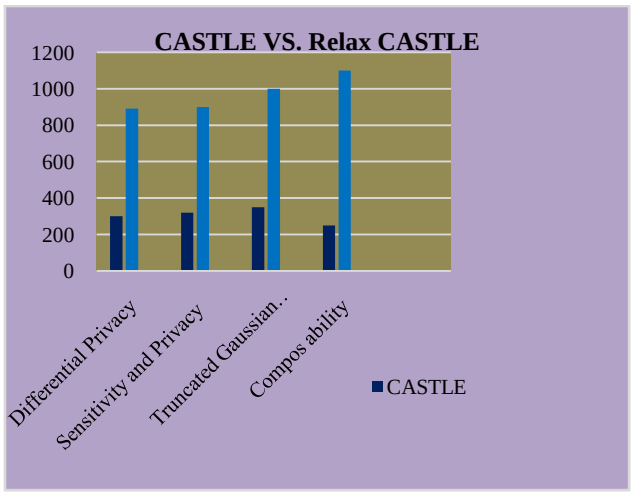

Fig 5.2: Castle vs. Relax CASTLE

TABLE 1: Efficiency of CASTLE.

Running Time(s)		
Data Size	Auditing Time	Setting Up
60	1.17	1780.21
100	4.62	9473.78
300	3148.14	55330.46

TABLE 2: Performance of CASTLE

Sample Size is N2	
Data Size	Auditing Time(s)
400	2.15
10000	27.29

TABLE 3: Performance Analysis

Performance Analysis	Simulatable	Statistical	CASTLE
Parameter Selection	Average	Good	Excellent
Collusion Resistance	Average	Good	Excellent
Denial Threat Removal	Average	Good	Excellent
Query Type	Average	Good	Excellent
Extension to more Attributes	Average	Good	Excellent
Dynamic Dataset Consideration	Average	Good	Excellent

VI. Conclusion

This paper presents a completely unique question auditing argument referred toward as CASTLE that authorizations auditing difference queries while not denial thread threats and enhancing the utility with slight perturbation. The key strategy is acknowledging the authentic information that query denials leak information. Upon every question denial, we have a tendency to derive info escape and treat it as a region of the antagonist information once auditing later queries.

Our theme present a lot of comprehensive and general privacy definition that relies on the safe zone and considers the correlation among the dataset. It achieves sensible performance in terms of auditing potency. We have a tendency to conjointly propose a relaxed version that improves the utility where as satisfying differential privacy. Moreover, we have a tendency to propose A Next finished theme for auditing amalgamated equality queries while not denial threats. The experimental study shows that our theme provides the utmost information effectiveness to users, as the privacy boundaries area unit reached for every case.

VII. Reference

[1] Haibing Lu, Jaideep Vaidya, VijayalakshmiAtluri, and Yingjiu Li, "Statistical database auditing without query denial threat," INFORMS JOC, vol. 27, no. 1, pp. 20–34, 2014.
[2] Taeho Jung, Xiang-Yang Li, Wenchao Huang, Jianwei Qian, Linlin Chen, Junze Han, JiahuiHou, and Cheng Su, "Accounttrade: Accountable protocols for big data trading against dishonest consumers," in INFOCOM 2017-IEEE Conference on Computer Communications, IEEE. IEEE, 2017, pp. 1–9.
[3] PrasangUpadhyaya, Nick R Anderson, Magdalena Balazinska, Bill Howe, Raghav Kaushik, Ravishankar Ramamurthy, and Dan Suciu, "Stop that query! the need for managing data use.," in CIDR, 2013
[4] Nabil R Adam and John C Worthmann, "Security-control methods for statistical databases: a comparative study," ACM CSUR, vol. 21, no. 4, pp. 515–556, 1989.
[5] Jianwei Qian, Xiang-Yang Li, et al., "Social network de-anonymization and privacy inference with knowledge graph model," IEEE
[6] Transactions on Dependable and Secure Computing, Pp.1-14, 2017.
[7] Adam NR, Wortmann JC (1989) Security-control methods for statistical databases: A comparative study. ACM Comput. Surveys 21:515–556
[8] Chin FYL, Özsoyoglu G (1981) Statistical database design. ACM Trans. Database Systems 6:113–139.
[9] Friedman AD, Hoffman LJ (1980) Towards a fail-safe approach to secure databases. IEEE Sympos. Security and Privacy, Oakland, CA.
[10] Castro J (2007) A shortest-paths heuristic for statistical data protection in positive tables. INFORMS J. Comput. 19:520–533.
[11] Kadane JB, Krishnan R, Shmueli G (2006) A data disclosure policy for count data based on the COM-Poisson distribution. Management Sci. 52:1610–1617.
[12] Kenthapadi K, Mishra N, Nissim K (2005) Simulatable auditing. Proc. Twenty-Fourth ACM Sympos. Principles Database Systems (ACM, New York), 118–127.
[13] [12] Malvestuto FM, Moscarini M (2006) Auditing sum-queries to make a statistical database secure. ACM Trans. Inform. System Security 33:451–464.
[14] [13] Steven P Reiss, "Security in databases: A combinatorial study," JACM, vol. 26, no. 1, pp. 45–57, 1979.
[15] [14] Andrew C Berry, "The accuracy of the gaussian approximation to the sum of independent variates," Transactions of the AMS, vol. 49, no. 1, pp. 122–136, 1941.
[16] [15] Cynthia Dwork, Frank McSherry, KobbiNissim, and Adam Smith, "Calibrating noise to sensitivity in private data analysis," in Theory of cryptography, pp. 265–284. Springer, 2006.
[17] [16] Cynthia Dwork, "Differential privacy: A survey of results," in TAMC. Springer, 2008, pp. 1–19.
[18] [17] Ben Cousins and Santosh Vempala, "A practical volume algorithm," Mathematical Programming Computation, vol. 8, no. 2, pp. 133–160, 2016.
[19] [18] Cunjing Ge and Feifei Ma, "A fast and practical method to estimate volumes of convex polytopes," in International Workshop on Frontiers in Algorithmics. Springer, 2015, pp. 52–65.
[20] Krishnaram Kenthapadi, Nina Mishra, and Kobbi Nissim, "Simulatable auditing," in PODS. ACM, 2005, pp. 118–127.
[21] Francis Y Chin and Gultekin Ozsoyoglu, "Statistical database design," ACM TODS, vol. 6, no. 1, pp. 113–139, 1981.