An Efficient and Secure Way for Key Distribution Using Attribute Based Multiple Keywords Subset Search Method

V.Geetha[#1], Dr.M.V.Srinath[*2], S.Karthiga[*3]

[#]*Head of the Department of Computer Science, Sengamala Thayaar Educational Trust Women's College, Mannargudi, Thiruvarur, Tamil Nadu*

[*]*Department of Computer Science, Sengamala Thayaar Educational Trust Women's College, Mannargudi, Thiruvarur, Tamil Nadu*

[3]senthilkarthi94@gmail.com

Abstract— **A collusion implies a contract between at least two gatherings. In some cases illicit and in this way cryptic, to restrain open rivalry by beguiling, deceiving, or duping others of their lawful rights. In this research, a secure data sharing scheme, which can achieve secure key distribution and data sharing for dynamic group. A safe route for key dissemination with no safe correspondence channels, and the clients can safely get their private keys from assemble administrator. At that point, plan can accomplish fine-grained get to control, any client in the gathering can utilize the source in the cloud and renounce clients can't get to the cloud again after they are repudiated. A proposed escrow free traceable attribute based multiple keywords subset search system with verifiable outsourced decryption approach can protect from the collusion attack, which means that revoked users cannot get the original data file even if they conspire with the untrusted cloud. By utilizing access control polynomial, it is configuration to accomplish efficient get to control for dynamic gatherings. Give a security investigation to demonstrate the security of our plan. The results show that fine-grained access control technology can ensure data privacy in mobile cloud and reduce the overhead on user's side in mobile cloud.**

Index Terms— **Multiple keywords subset search system, Key distribution, Security**

INTRODUCTION

This examination Cloud registering is the consequence of the advancement and appropriation of existing advances and ideal models. The objective of distributed computing is to enable clients to take benefit from these innovations, without the requirement for profound learning about or mastery with every single one of them. The cloud means to cut expenses, and enables the clients to center around their center business as opposed to being hindered by IT hindrances. The principle empowering innovation for distributed computing is virtualization. Virtualization programming isolates a physical registering gadget into at least one "virtual" gadgets, every one of which can be effortlessly utilized and figured out how to perform processing assignments. With working system level virtualization basically making a versatile arrangement of numerous autonomous figuring gadgets, sit without moving processing assets can be distributed and utilized all the more effectively. Virtualization gives the spryness required to accelerate IT tasks, and lessens cost by expanding framework usage.

Cloud is frequently used as a piece of science to delineate a far reaching agglomeration of articles that ostensibly appear from a partition as a cloud and depicts any course of action of things whose points of interest are not additionally reviewed in a given setting. Another clarification is that the old projects that drew arrange schematics encompassed the symbols for servers with a circle, and a bunch of servers in a system chart had a few covering circles, which looked like a cloud. In similarity to the above utilization, the word cloud was utilized as a representation for the Internet and an institutionalized cloud-like shape was utilized to mean a system on communication schematics. Later it was utilized to delineate the Internet in PC organize charts. With this rearrangements, the suggestion is that the specifics of how the end purposes of a system are associated are not pertinent for the reasons for understanding the chart.

Our answer depends on a cautious testing of focuses from the plane and utilizing the closest neighbor prophet. To represent the significance of practicing alert in the approach,

think about the accompanying technique for assessing a measurement of the items: pick an arbitrary point p on the plane and after that utilization the closest neighbor prophet to the question d that is nearest to p and utilize f(d) as an estimator. While the point p are picked consistently aimlessly, shockingly, the subsequent example isn't uniform.

With the expanding measure of excellent organized information on the Web, getting to the information in profound web sources have increased more consideration. The entrance to this information is conceivable through various diverse strategies, for example, creeping and testing. In the calculations connected in these techniques, it is the learning of the information source measure that empowers the calculations to settle on choices in halting the slithering or inspecting forms which can be so exorbitant at times.

I. BACKGROUND

A distributed processing is a kind of Internet-based enrolling that gives shared planning resources and data to PCs and diverse devices on ask. It is a model for enabling inescapable, on-ask for access to a typical pool of configurable figuring resources (e.g., frameworks, servers, amassing, applications and organizations), which can be immediately provisioned and released with unimportant organization effort. Disseminated registering and limit courses of action give customers and endeavors diverse capacities to store and process their data in pariah server ranches. It depends on sharing of assets to accomplish lucidness and economy of scale, like an utility (like the power framework) over a system. Promoters guarantee that distributed computing enables organizations to maintain a strategic distance from forthright foundation expenses, and spotlight on examines that separate their organizations rather than on framework. Safeguards also affirm that circulated registering empowers endeavors to get their applications up and running speedier, with improved sensibility and less upkeep, and engages IT to all the more quickly modify assets to meet fluctuating and capricious business request. Cloud suppliers normally utilize a "pay as go" display. This will prompt out of the blue high charges if executives don't adjust to the cloud valuing model. The present openness of high-confine frameworks, negligible exertion PCs and limit devices and what's more the unfathomable gathering of gear virtualization, advantage organized outline, and autonomic and utility enlisting have provoked an advancement in conveyed processing. Associations can scale up as figuring needs increase and a short time later cut back again as solicitations reduce. Distributed computing has turned into a

profoundly requested administration or utility because of the upsides of high processing power, shoddy cost of administrations, superior, adaptability, openness and in addition accessibility. Some cloud sellers are encountering development rates of half every year, except being still in a phase of early stages, it has entanglements that should be routed to make distributed computing administrations more dependable and easy to understand.

Disseminated frameworks are gatherings of arranged PCs, which have a similar objective for their work. The expressions "simultaneous figuring", "parallel registering", and "conveyed processing" have a considerable measure of cover, and no reasonable qualification exists between them. A similar framework might be portrayed both as "parallel" and "conveyed"; the processors in a regular circulated framework run simultaneously in parallel. Parallel figuring might be viewed as a specific firmly coupled type of appropriated processing, and disseminated registering might be viewed as an approximately coupled type of parallel registering. All things considered, it is conceivable to generally characterize simultaneous frameworks as "parallel" or "appropriated" utilizing the accompanying criteria. In parallel registering, all processors may approach a common memory to trade data between processors. In conveyed figuring, every processor has its own private memory (dispersed memory). Data is traded by passing messages between the processors. On the privilege shows the distinction amongst appropriated and parallel frameworks. Schematic perspective of a run of the mill circulated framework; of course, the framework is spoken to as a system topology in which every hub is a PC and each line interfacing the hubs is a correspondence connect. Demonstrates the same circulated framework in more detail: every PC has its own particular nearby memory, and data can be traded just by passing messages starting with one hub then onto the next by utilizing the accessible correspondence joins. Demonstrates a parallel framework in which every processor has an immediate access to a common memory.

II. RELATED WORK

A. Aggregate Estimation in Hidden Databases with Checkbox Interfaces

In this research, create novel procedures for assessing and following different sorts of total inquiries, e.g., COUNT and SUM, over powerful web databases that are taken cover behind restrictive inquiry interfaces and as often as possible changed. Shrouded Web Databases: Many web databases are "covered up" be-rear prohibitive pursuit

interfaces that enable a client to determine the coveted qualities for one or a couple of characteristics (i.e., shape a conjunctive hunt inquiry), and come back to the client a modest number (limited by a steady k which can be 50 or 100) of tuples that match the client indicated question, chose and positioned by an exclusive scoring capacity. Cases of such databases incorporate Yahoo! Automobiles, Amazon.com, eBay.com, CareerBuilder.com, and so forth. Issue Motivations: The issue consider. The examination is the means by which an outsider can utilize the prohibitive web interface to gauge and track total inquiry replies over a dynamic web database. Total inquiries are the most widely recognized sort of questions in choice emotionally supportive networks as they empower powerful examination to gather bits of knowledge from the information. The examination tends to a novel issue where checkboxes exist in the web interface of a shrouded database. To empower the guess handling of total inquiries and creates calculation fair-minded weighted-slither which performs arbitrary bore downs on a novel structure of questions which allude to as a left-profound tree and furthermore propose weight modification and low likelihood creep to enhance estimation precision. This examination played out a far reaching set of tests on manufactured and genuine datasets with changing database sizes (from 5000 to 100000), number of traits (from 20 to 50) and best k confinement (from k = 10 to 30). found that, as anticipated by the hypothetical investigation, the relative blunder diminishes when the quantity of inquiries issued increments.

B. Enhance Estimation in Hidden Databases Using Cache Memory with Left Deep Tree

In this research, They find that, with the end goal of information examination, such checkbox spoke to properties contrast on a very basic level from the downright/numerical ones that Ire generally considered. The examination, I address the issue of information investigation over shrouded databases with checkbox interfaces. Broad tests on both engineered and genuine datasets exhibit exactness and effectiveness of our proposed calculations. Shrouded databases are information stores "taken cover behind" i.e., just available through a prohibitive IB look interface. Information capacities gave by such IB interface go from a basic watchword look textbox (e.g., Google) to an intricate blend of textboxes, dropdown controls, checkboxes, and so forth. Once a client determines an inquiry question of enthusiasm through the information interface, the shrouded database chooses and restores a set number (i.e., top-k) of

tuples fulfilling the client indicated look conditions (frequently as per a restrictive positioning capacity), where k is typically a little whole number, for example, 50 or 100. Truth be told, numerous IB concealed databases convey their best k comes about for an inquiry with a few ib pages.

C. Aggregate Estimations over Location Based Services

In this research, The database is basically "covered up", and get to is ordinarily constrained to a confined open web inquiry interface or API by which one can indicate a self-assertive area as a question, which returns at most k closest tuples to the inquiry point (where k is regularly a modest number, for example, 10 or 50). For instance, in Google maps it is conceivable to indicate a subjective area and get the ten closest Starbucks. Accordingly, the question interfaces of these administrations might be dynamically displayed as a "closest neighbor" KNN interface over a database of two dimensional focuses on a plane: given a self-assertive inquiry point, the framework restores the k focuses in the database that are closest to the question point. Moreover, there are essential contrasts among the administrations in view of the kind of data that is returned alongside the k tuples. A few administrations (e.g., Google maps) restore the areas (i.e., the x and y arranges) of the k returned tuples. Allude to such administrations as Location-Returned LBS (LR-LBS). Different administrations (e.g., Chat, Sina Weibo) restore a positioned rundown of k closest tuples, yet stifle the area of each tuple, returning just the tuple ID and maybe some different characteristics, (for example, tuple name).refer to such administrations as Location-Not-Returned LBS (LNR-LBS). The two sorts of administrations force extra questioning confinements, the most essential being a for each client/IP restrain on the quantity of inquiries one can issue over a given time period (e.g., as a matter of course, Google delineate forces an inquiry rate breaking point of 10,000 for every client for every day). In investigate the issue of total estimation over area based administrations that are progressively mainstream. Presented a scientific categorization of LBS with k-NN inquiry interface in view of whether area of the tuple is returned (LR-LBS) or not (LNRLBS). For the previous, proposed a productive calculation and different mistake decrease methodologies that beats earlier work. The start think about into the last by proposing powerful calculations for total and deriving the situation of tuple to discretionary exactness which may be of free intrigue. Checked the viability of our calculations by utilizing a far reaching set of examinations on a substantial certifiable

geographic dataset and online exhibits on prominent genuine sites.

D. Aggregate Estimation over Dynamic Hidden Web Databases

In this research, the database is basically "covered up", and get to is ordinarily constrained to a confined open web inquiry interface or API by which one can indicate a self-assertive area as a question, which returns at most k closest tuples to the inquiry point (where k is regularly a modest number, for example, 10 or 50). For instance, in Google maps it is conceivable to indicate a subjective area and get the ten closest Starbucks. Accordingly, the question interfaces of these administrations might be dynamically displayed as a "closest neighbor" KNN interface over a database of two dimensional focuses on a plane: given a self-assertive inquiry point, the framework restores the k focuses in the database that are closest to the question point. Moreover, there are essential contrasts among the administrations in view of the kind of data that is returned alongside the k tuples. A few administrations (e.g., Google maps) restore the areas (i.e., the x and y arranges) of the k returned tuples. Allude to such administrations as Location-Returned LBS (LR-LBS). Different administrations (e.g., Chat, Sina Weibo) restore a positioned rundown of k closest tuples, yet stifle the area of each tuple, returning just the tuple ID and maybe some different characteristics, (for example, tuple name). Refer to such administrations as Location-Not-Returned LBS (LNR-LBS). The two sorts of administrations force extra questioning confinements, the most essential being a for each client/IP restrain on the quantity of inquiries one can issue over a given time period (e.g., as a matter of course, Google delineate forces an inquiry rate breaking point of 10,000 for every client for every day). In investigate the issue of total estimation over area based administrations that are progressively mainstream. Presented a scientific categorization of LBS with k-NN inquiry interface in view of whether area of the tuple is returned (LR-LBS) or not (LNRLBS). For the previous, proposed a productive calculation and different mistake decrease methodologies that beats earlier work. The start think about into the last by proposing powerful calculations for total and deriving the situation of tuple to discretionary exactness which may be of free intrigue. Checked the viability of our calculations by utilizing a far reaching set of examinations on a substantial certifiable geographic dataset and online exhibits on prominent genuine sites.

E. HDBTracker: Monitoring the Aggregates on Dynamic Hidden Web Databases

In this research, various web databases, e.g., amazon.com, eBay.com, are "covered up" behind (i.e., open just through) their prohibitive hunt and perusing interfaces. This exhibit grandstands HDB Tracker, an online framework that uncovers and tracks (the progressions of) client indicated total inquiries over such concealed web databases, particularly those that are habitually refreshed, by issuing few inquiry questions through people in general web interfaces of these databases. The capacity to track and screen totals has applications over a wide assortment of areas - e.g., government organizations can track tally of openings at online occupation chasing sites to comprehend key monetary pointers, while organizations can track the AVG cost of an item finished a bushel of web based business sites to comprehend the focused scene as well as material expenses. A key strategy utilized as a part of HDB Tracker is RS estimator, the primary calculation that can productively screen changes to total inquiry replies over a concealed web database. Propose to demo HDB Tracker, a prototypal framework worked for observing the continuous changes of different kinds of totals, e.g., COUNT, SUM, and AVG questions with or without choice conditions, over every now and again changed web databases that are taken cover behind restrictive hunt and additionally perusing interfaces.

F. Interactive Pattern Mining on Hidden Data: A Sampling-based Solution

In this research, Mining incessant examples from a concealed dataset is a vital undertaking with different genuine applications. In this exploration, propose an answer for this issue depends on Markov Chain Monte Carlo (MCMC) examining of successive examples. Rather than restoring all the incessant examples, the proposed worldview restores a little arrangement of arbitrarily chose designs so the family destiny of the dataset can be kept up. Our answer additionally permits intuitive inspecting, so that the examined examples can satisfy the client's prerequisite successfully. Show exploratory outcomes from a few genuine datasets to approve the ability and helpfulness of our answer; specifically, demonstrate cases that by utilizing our proposed arrangement, an online business commercial center can permit congratulate tern mining on client session information without unveiling the information to the general population; such a mining worldview helps the merchants of the commercial center, which inevitably support the commercial center's Own income. Visit design mining assumes a key part in exploratory information examination. In the course of the

most recent two decades, specialists in-vented different productive calculations for mining examples of changing degrees of complexity, for example, sets. Accessibility of successive things of client session questions likewise helps the venders in picking an educational title for their item to encourage powerful coordinating of dealer's item with the purchaser's inquiry inevitably, boosting the commercial center's income. Since various venders might be keen on various arrangements of inquiries, the key test for the commercial center in this errand is to discover a component to help a dealer by giving regular thing sets of questions that would. Advantage that particular merchant the most all without taking a chance with the spillage of huge piece of the session information. The novel research issue and our work is only the starting, so the chance of future work is bottomless. For instance, in this work have considered a semi-legit show, though, all things considered, a malignant model might be required.

IV. METHODOLOGY

A Confidential Data Distribution System for Dynamic Cloud

Give a safe method to key circulation with no safe correspondence channels. The clients can safely acquire their private keys from amass administrator with no Certificate Authorities because of the check for people in general key of the client. Our plan can accomplish fine-grained get to control, with the assistance of the gathering client list, any client in the gathering can utilize the source in the cloud and denied clients can't get to the cloud again after they are renounced. Propose a safe information sharing plan which can be shielded from intrigue assault. The denied clients can not have the capacity to get the first information records once they are renounced regardless of whether they plot with the untrusted cloud. Our plan can accomplish secure client repudiation with the assistance of polynomial capacity. Our arrangement can reinforce dynamic social events capably, when another customer partakes in the get-together or a customer is denied from the get-together, the private keys of interchange customers don't ought to be recomputed and revive.
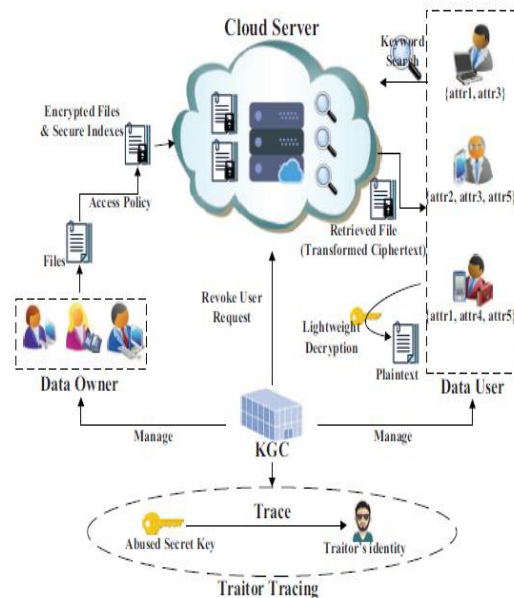

Fig. 1 Schematic Diagram of Protected Data Distribution System

Key Generation Algorithm

Input: Select two random distinct prime numbers w and x.

Output: Find Public Key (U), Private Key (R) and Modulus (j)

Begin

Procedure (w, x, U, R and j)

    1. $j \leftarrow w * x$

    2. Calculate Euler $\emptyset$ () of j $\emptyset$ (j) $\leftarrow$ (w-1) * (x-1)

    3. Generate a public key U, such that, gcd (U, $\emptyset$ (j)) = 1, 1< U < $\emptyset$ (x)

    4. Calculate the private key q, such that, $R \leftarrow U^{-1}$ mod ($\emptyset$ (j))

    End Procedure

    End

Encryption Algorithm

Input: Select Plain text (T1), Public key (U) and Modulus (j).

    Output: Find Cipher text (C1).

    Begin

    Procedure (T1, U, j and C1)

    $C1 \leftarrow T^P$ mod j

    End Procedure

    End

Decryption algorithm

Input: Select Cipher text (C1), Private key (R) and Modulus (j).

Output: Find Plain text (T1).

Begin

Procedure (T1, U, j and C1)
T1 ← C1 R mod j
End Procedure

A polynomial is an articulation that can be worked from constants and images brought in determinates or factors by methods for expansion, increase and exponentiation to a non-negative power. Two such articulations that might be changed, one to the next, by applying the typical properties of commutativity, associativity and distributivity of expansion and augmentation are considered as characterizing a similar polynomial. A polynomial in a solitary uncertain x can simply be composed (or revamped) in the frame are constants and {\display style x} x is the vague. "Indeterminate" implies that {\display style x} x speaks to no specific esteem, albeit any esteem might be substituted for it. The mapping that partners the consequence of this substitution to the substituted esteem is a capacity, called a polynomial capacity. That is, a polynomial can either be zero or can be composed as the entirety of a limited number of non-zero terms. Each term comprises of the result of a number called the coefficient of the term and a limited number of in determinates, raised to nonnegative whole number forces. The example on an uncertain in a term is known as the level of that vague in that term; the level of the term is the entirety of the degrees of the in determinates in that term, and the level of a polynomial is the biggest level of any one term with nonzero coefficient. Since x = x1, the level of a vague without a composed type is one. A term and a polynomial with no in determinates are called, separately, a consistent term and a steady polynomial. The level of a consistent term and of a nonzero steady polynomial is 0. The level of the zero polynomial, 0, (which has no terms by any means) is for the most part regarded as not characterized.

A key generation centre is capable to create general society parameter for the framework and people in general/mystery key sets for the clients. Once the client's mystery key is spilled for benefits or different purposes, KGC runs follow calculation to locate the malevolent client. After the double crosser is followed, KGC sends client repudiation demand to cloud server to disavow the client's hunt benefit. Cloud server (CS). Cloud server has enormous storage room and ground-breaking registering capacity, which gives on-request administration to the framework. Cloud server is mindful to store the information proprietor's scrambled records and react on information client's hunt inquiry. Data proprietor. Information proprietor uses the distributed storage administration to store the documents. Prior to the information outsourcing, the information proprietor separates watchword set from the record and scrambles it into secure list. The record is additionally scrambled to ciphertext. Amid the encryption procedure, the entrance strategy is determined and implanted into the ciphertext to acknowledge finegrained get to control. Data client. Every datum client has ascribe set to portray his attributes, for example, educator, software engineering school, senior member, and so on. The characteristic set is inserted into client's mystery key. Utilizing the mystery key, information client can look on the scrambled records put away in the cloud, i.e., picks a watchword set that he needs to seek. At that point, the watchword is encoded to a trapdoor utilizing client's mystery key. On the off chance that the client's trait set fulfills the entrance strategy characterized in the scrambled documents, the cloud server reacts on client's inquiry question and finds the match records. Something else, the hunt inquiry is rejected. After the match documents are restored, the client runs unscrambling calculation to recoup the plaintext.

V. RESULT AND DISCUSSION

To assess the execution, EF-TAMKS-VOD are recreated utilizing the Stanford Pairing-Based Crypto (PBC) library. The analyses on these plans are directed on a PC running Windows 7 task framework with the accompanying settings: CPU: Intel center i5 CPU at 2.5GHz; physical memory: DDR3 4GB 1333MHz. The sort An elliptic bend parameter is chosen for test. It gives 1024-piece discrete log security quality identically with the gathering request of 160-piece. Sort A parings are developed on the bend $y2 = x3 + x$ over the field Zp for some prime p = 3 (mod 4). Which is given in PBC library. The center calculations are executed on the trial workbench to test the transmission and calculation overheads of the plans and EFTAMKS-VOD. As indicated by the chose parameter in the examination, we have jZ = 160 bits, jGj = 1024 bits and jGT j = 1024 bits. The number l1 of the catchphrase set is settled to be 5 to do the tests.

High Performance

In these result compare with performance. HS-MPP gives result based on performance 40%. UA gives result based on performance 25%. LPF gives result based on performance 50%. EF-TAMKSVOD gives result based on performance 80%.
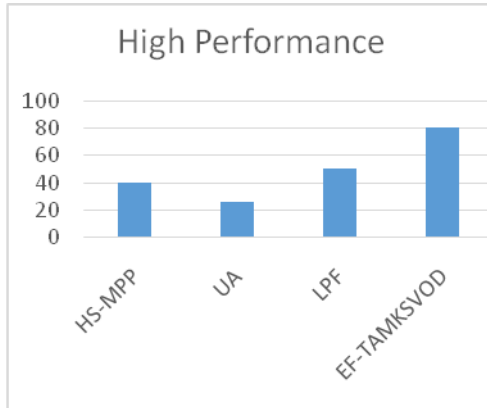
Fig.2   Performance analysis

More Reliability

In these result compare with reliability. HS-MPP gives result based on reliability 60%. UA gives result based on reliability 35%. LPF gives result based on reliability 70%. EF-TAMKSVOD gives result based on reliability 75%.
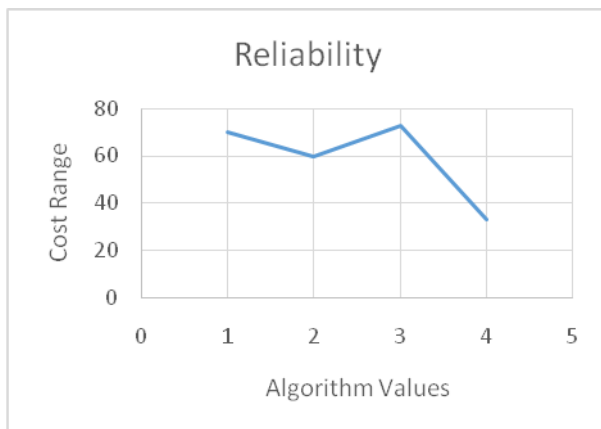

Fig.3   Reliability analysis

Low Cost

In these result compare with reliability. HS-MPP gives result based on reliability 60%. UA gives result based on reliability 35%. LPF gives result based on reliability 70%. EF-TAMKSVOD gives result based on reliability 20%.
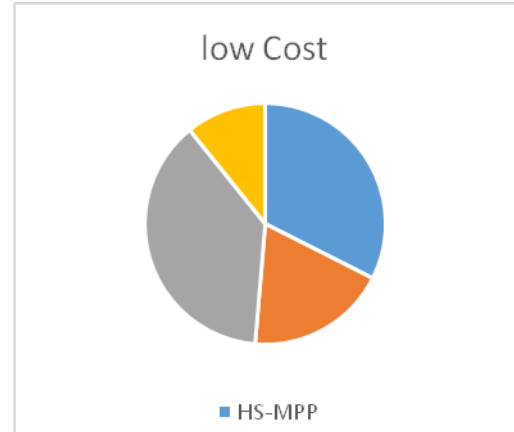

Fig.4 Cost analysis

VI. CONCLUSION

In this proposition, proposed a secure data sharing scheme plan, which can accomplished secure key appropriation and information sharing for dynamic gathering. A protected path for key dissemination with no safe correspondence channels, and the clients can safely got their private keys from gather director. At that point, our plan can accomplished fine-grained get to control, any client in the gathering can utilized source in the cloud and denied clients can't got to the cloud again after they are disavowed. Our plan can shielded from the arrangement assault, which implies that renounced clients can't get the first information record regardless of whether they plot with the untrusted cloud. Our approach utilizing polynomial capacity. By utilizing access control polynomial, it is configuration to accomplished efficient get to control for dynamic gatherings. Given security investigation to demonstrate the security of our plan. In future persists, will improve the performance of leveraging access control polynomial. High security and then improve feasibility.

REFERENCES

[1] Weimo Liuy, Saravanan Thirumuruganathan, Nan Zhangy, "Aggregate Estimation over Dynamic Hidden Web Databases", vol 7 IEEE Year: Sep 2014.

[2] Weimo Liu1, Saar Bin Suhaim1, Saravanan Thirumuruganathan, "HDB Tracker: Monitoring the Aggregates on Dynamic Hidden Web Databases", 2nd Conference Year: Sep 2014.

[3] Mansurul Bhuiyan, Synthesis Mukhopadhyay, "Interactive Pattern Mining on Hidden Data: A Sampling-based Solution", IEEE Year: Nov 2012.

[4] C. Sheng, N. Zhang, Y. Tao, and X. Jin, "Optimal algorithms for crawling a hidden database in the web", Proc. VLDB Endowment, vol. 5, no. 11, pp. 1112–1123, 2012.

[5] Mohammed Al-Kata and Bunge Suk Lee, "Load Shedding for Temporal Queries over Data Streams", IEEE Year: Nov 2011.

[6] Raven kumar MCA, Mohamed Faseel.VK, "Aggregate Estimation in Hidden Databases with Checkbox Interfaces", IEEE 29th Conference on Computational Complexity Year: Apr 2016.

[7] S.Nandagopal1, S.Jegadeesan, "Enhance Estimation in Hidden Databases Using Cache Memory with Left Deep Tree", IEEE Year: March 2016.

[8] Md Farhadur Rahmanz, Saravanan Thirumuruganathan, "Aggregate Estimations over Location Based Services", IET conference Year: Sep, 2015.

[9] Fan Wang, Gaga Agrawal, "Effective and Efficient Sampling Methods for Deep Web Aggregation Queries", IEEE 21th Conference on Computational Year: March 2011.

[10] M. Benedikt, G. Gottlob, and P. Senellart, "Determining relevance of accesses at runtime,‖ in Proc. 30th ACM SIGMOD-SIGACT- SIGART Symp. Principles Database Syst., 2011, pp. 211–222.

[11] L. Barbosa and J. Freire, "Siphoning hidden-web data through keyword-based interfaces", JIDM, 1(1):133–144, 2010.

[12] N. Afrati a, Paraskos's V. Lekeas b, 1, Chen Li c, "Adaptive-sampling algorithms for answering aggregation Queries on Web sites", IEEE Year: Sep 2007.

[13] Jiying Wang, Ji-Rong, "Wen Instance-based Schema Matching for Web Databases by Domain-specific Query Probing" , International conference Year: 2004.

[14] B. Henry and K. Chang, "Statistical schema matching across web query interfaces. In SIGMOD", VOL 6, 2003.

[15] E. Agichtein, P. G. Ipeirotis, and L. Gravano, "Modeling query-based access to text databases. In WebDB", conference 2003.

[16] N. Bruno, L. Gravano, and A. Marian, "Evaluating top-k queries over web-accessible databases", In ICDE, 2002.

[17] J. Callan and M. Connell, "Query-based sampling of text databases. ACM TOIS", 19(2):97–130, 2001.

[18] A. Dasgupta, X. Jin, B. Jewell, N. Zhang, and G. Das, "Unbiased estimation of size and other aggregates over hidden web databases",‖ in Proc. Int. Conf.Manage. Data, 2010, pp. 855–866.

[19] M. Benedikt, P. Bourhis, and C. Ley, "Querying schemas with access restrictions,‖ Proc. VLDB Endowment", vol. 5, no. 7,pp. 634–645, 2012

[20] R. Khare, Y. An, and I.-Y. Song, ―Understanding deep web search interfaces: A survey,‖ ACM SIGMOD Rec., vol. 39, no. 1, pp. 33– 40, 2010.

[21] A. Dasgupta, X. Jin, B. Jewell, N. Zhang, and G. Das, "Unbiased estimation of size and other aggregates over hidden web databases", In SIGMOD, pages 855–866, 2010.

[22] Beeri, C., S. Naqvi, R. Ramakrishnan,Shmueli, S. Tsur, "Sets and Negation in a Logic Database Language", Proc. 6th ACM SIGMOD-SIGACT Symposium on Principles of Database Systems, 1987.

[23] Bancilhon, F., D, Maier, Y. Sagiv and Ullman, Magic Sets and other Strange Ways to Implements Logic Programs, Proc. S-th ACM SIGMOD-SIGACT Symposium on Principles of Database Systems, pp. 1-16, 1986.

[24] Bancilhon, F., and R. Ramakrishan, An Amateur's Introduction to Recursive Query Processing Strategies, Proc. 1986 ACM-SIGMOD Intl. Conf. on Mgt. of Data, pp. 16-52, 1986.

[25] Krishnamurthy, R., Boral, H., Zaniolo, C. Optimization of Nonrecursive Queries, Proc. of 12th VLDB, Kyoto, Japan, 1986.

[26] Ait-Kaci, H. and R. Nasr, "Residuation: a Paradigm for Integrating Logic and Func- tional Programming," submitted for publication 1986.