International Journal on Applicationsin Information and Communication Engineering
Volume 3:Issue 1: January 2017, pp 1 - 5. www.aetsjournal.com ISSN (Online) : 2394 - 6237

DOMINANT TOP K QUERY CONSEQUENCE INTEGRITY VALIDATION USING DATA MINING

Dr.G.Komarasamy , B.Priyachandru , V.GuruMoorthy, M.Karthick

Abstract— Web users and content are increasingly being geo-positioned, and increased focus is being given to serving local content in response to web queries. This development calls for spatial keyword queries that take into account both the locations and textual descriptions of content. We study the efficient, joint processing of multiple top-k spatial keyword in data mining. Such joint processing is attractive during high query loads and also occurs when multiple queries are used to obfuscate a user's true query. We propose a novel algorithm and index structure for the joint processing of top-k spatial in data mining process. Empirical studies show that the proposed solution is efficient on real datasets. We also offer analytical studies on synthetic datasets to demonstrate the efficiency of the proposed solution.

Keywords **--** Textual Descriptions, Top-K Spatial, Multiple Queries, Joint Processing, Synthetic Datasets.

I. INTRODUCTION

In many cases, a set of parties would like to aggregate private key-value lists and find the keys with the highest aggregate values. This problem is ubiquitous having applications in many fields, including network monitoring, security analysis, and distributed databases. Our work is motivated by a simple profiling application. The server computes top-traffic feature reports, such as top client, and we would like to compare local top-*k*reports with aggregate reports corresponding to the data of multiple cooperating collectors. Moreover, organizations maintaining reputation scores for identifiers, like data sent, data received, addresses, can aggregate their scores to find the overall top-scored identifiers. Unfortunately, these and other similar distributed problems lack practical solutions that preserve the privacy of the involved data.

Dr.G.Komarasamy , Asst.Prof (Sr.G) , Department of Computer Science and Engineering , Bannari Amman Institute of Technology, Sathyamangalam, India .

B. Priyachandru , Department of Computer Science and Engineering , Bannari Amman Institute of Technology, Sathyamangalam, India

V. GuruMoorthy , Department of Computer Science and Engineering , Bannari Amman Institute of Technology, Sathyamangalam, India

M. Karthick , Department of Computer Science and Engineering , Bannari Amman Institute of Technology, Sathyamangalam, India

II. LITERATURE REVIEW

A large-scale wireless sensor network constructed in terms of two-tiered architecture, where cloud nodes take charge of storing sensed data and processing queries with respect to the sensing nodes and querists, incurs security breach. This is because the importance of cloud nodes makes them attractive to adversaries and raises concerns about data privacy and query result correctness. To address these problems, we propose an efficient approach, namely EQ (efficient query), which mainly prevents adversaries from gaining the information processed by or stored in cloud nodes, and detects the compromised cloud nodes when they misbehave. EQ can not only achieve the goals of data privacy and integrity preserving but also ensure the secure range query without incurring false positive. For data privacy preserving, EQ presents an order encryption mechanism by adopting stream cipher to encrypt/decrypt all sensed data such that a cloud node can only process issued queries over stored data in the encryption domain.

III. SYSTEM STUDY

A.Existing System

The existing user search does not get perfect queries. The existing queries cannot simultaneously prune the search space using both keyword similarity and spatial distance and this is considering only keyword based queries..

Disadvantage:

1. It cannot simultaneously prune the search space using both keyword similarity and spatial distance.
2. The existing works consider keyword based queries.
3. Client and server communication only available. But cannot find the top n user communication.

B.Proposed System

We propose a novel algorithm and index

structure for the joint processing of top-k spatial keyword queries. Empirical studies show that the proposed solution is efficient on real datasets. We also offer analytical studies on synthetic datasets to demonstrate the efficiency of the proposed solution.

Advantage:

- It can get simultaneous search using both keyword similarity and spatial distance.
- More than 3 Clients can communicate with the user. If u need another communication contact the server part.
- The proposed system offers location privacy by hiding the true query among multiple fake queries.
- Top-k processing techniques are classified according to the query model they assume. Some techniques assume a selection query model, where scores are attached directly to base tuples. Other techniques assume a join query model, where scores are computed over join results.
- Top-k processing techniques are classified according to the data access methods they assume to be available in the underlying data sources. For example, some techniques assume the availability of random access, while others are restricted to only sorted access.
- Some techniques produce exact answers, while others allow for approximate answers, or deal with uncertain data.
- Top-k processing techniques are classified based on the restrictions they impose on the underlying ranking (scoring) function.
- Top-k processing techniques are classified according to their level of integration with database systems. For example, some techniques are implemented in an application layer on top of the database system, while others are implemented as query operators.
- Top-k processing techniques are classified based on the uncertainty involved in their data and query models.

IV. TECHNOLOGIES USED

Around the time the First Person project was floundering in consumer electronics, a new craze was gaining momentum in America; the craze was called "Web surfing." The World Wide Web, a name applied to the Internet's millions of linked HTML documents was suddenly becoming popular for use by the masses. The reason for this was the introduction of a graphical Web browser called Mosaic, developed by ncSA. The browser simplified Web browsing by combining text and graphics into a single interface to eliminate the need for users to learn manyconfusing UNIX and DOS commands. Navigating around the Web was much easier using Mosaic.

It has only been since 1994 that Oak technology has been applied to the Web. In 1994, two Sun developers created the first version of Hot Java, and then called Web Runner, which is agraphical browser for the Web that exists today. The browser was coded entirely in the Oak language, by this time called Java. Soon after, the Java compiler was rewritten in the Java language from its original C code, thus proving that Java could be used effectively as an application language. Sun introduced Java in May 1995 at the Sun World 95 convention.

Web surfing has become an enormously popular practice among millions of computer users. Until Java, however, the content of information on the Internet has been a bland series of HTML documents. Web users are hungry for applications that are interactive, that users can execute no matter what hardware or software platform they are using, and that travel across heterogeneous networks and do not spread viruses to their computers. Java can create such applications. Number of affiliations, the final affiliation will be centered on the page ; all previous will be in two columns.

4.1 Working of Java

For those who are new to object-oriented programming, the concept of a class will be new to you. Simplistically, a class is the definition for a segment of code that can contain both data (called attributes) and functions (called methods).

When the interpreter executes a class, it looks for a particular method by the name of main, which will sound familiar to C programmers. The main method is passed as a parameter an array of strings (similar to the argv[] of C), and is declared as a static method.

To output text from the program, we execute the println method of System.out, which is java's output stream. UNIX users will appreciate the thoery behind such a stream, as it is actually standard output. For those who are instead used to the Wintel platform, it will write the string passed to it to the user's program.

V. IMPLEMENTATION

5.1 MODULES

1. Grouping/partitioning a large set of queries.
2. Caching historical information for future queries.

3. Techniques for the efficient processing of one group of queries.

(1) Grouping/partitioning a large set of queries

To use a space filling curve to group queries so as to improve the overall performance of processing all queries. The processing of multiple nearest neighbor queries; they propose R-tree-based solutions and heuristics for the grouping of queries. However, the joint query we consider is actually one group of queries. We thus aim to compute one group of queries efficiently. Any grouping/partitioning approach can be applied before our algorithm.

(2) Caching historical information for future queries

Cache the result objects as well as the index that supports these objects as the results so as to optimize query response time. Consider a scenario where objects are stationary while queries are mobile. They develop a semantic caching scheme that records a cached item (object) as well as its valid range (Voronoi cell). However, their approach cannot be directly applied to our problem because their valid range concept ignores the query keywords in our problem. In our experiments, traditional caching, i.e., LRU buffering, is considered as a competitor to our proposals.

(3) Techniques for the efficient processing of one group of queries

The sub-expressions shared among a set of SQL queries can be evaluated once and then subsequently reused. This reduces the cost when compared to processing each query separately. None of these existing works consider keyword-based queries. However, we adopt the same general philosophy and aim to share computation across multiple queries.

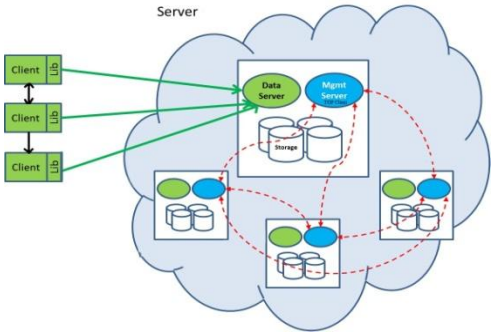

Fig.1 System Architecture

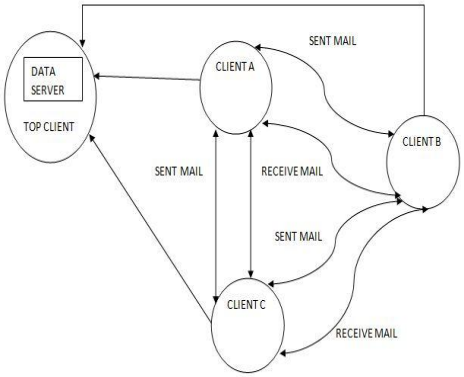

Fig.2 Dataflow Diagram

5.2 Coding Standards

Coding standards are guidelines to programming that focuses on the physical structure and appearance of the program. They make the code easier to read, understand and maintain. This phase of the system actually implements the blueprint developed during the design phase. The coding specification should be in such a way that any programmer must be able to understand the code and can bring about changes whenever felt necessary. Some of the standard needed to achieve the above-mentioned objectives are as follows:

- Program should be simple, clear and easy to understand.
- Naming conventions
- Value conventions
- Script and comment procedure
- Message box format
- Exception and error handling

5.3 Naming Conventions

Naming conventions of classes, data member, member functions, procedures etc., should be **self-descriptive**. One should even get the meaning and scope of the variable by its name. The conventions are adopted for **easy understanding** of the intended message by the user. So it is customary to follow the conventions. These conventions are as follows:

Class names

Class names are problem domain equivalence and begin with capital letter and have mixed cases.

Member Function and Data Member name

Member function and data member name begins with a lowercase letter with each subsequent letters of the

new words in uppercase and the rest of letters in lowercase.

VALUE CONVENTIONS

Value conventions ensure values for variable at any point of time. This involves the following:

- Proper default values for the variables.
- Proper validation of values in the field.
- Proper documentation of flag values.

SCRIPT WRITING AND COMMENTING STANDARD

Script writing is an art in which indentation is utmost important. Conditional and looping statements are to be properly aligned to facilitate easy understanding. Comments are included to minimize the number of surprises that could occur when going through the code.

MESSAGE BOX FORMAT

When something has to be prompted to the user, he must be able to understand it properly. To achieve this, a specific format has been adopted in displaying messages to the user. They are as follows:

- X – User has performed illegal operation.
- ! – Information to the user.

VI. SOFTWARE TESTING

(1) Testing

Testing is a process used to help identify the correctness, completeness and quality of developed computer software. With that in mind, testing can never completely establish the correctness of computer software.

There are many approaches to software testing, but effective testing of complex products is essentially a process of investigation, not merely a matter of creating and following rote procedure. One definition of testing is "the process of questioning a product in order to evaluate it", where the "questions" are things the tester tries to do with the product, and the product answers with its behavior in reaction to the probing of the tester. Although most of the intellectual processes of testing are nearly identical to that of review or inspection, the word testing is connoted to mean the dynamic analysis of the product—putting the product through its paces.

Testing objectives include

1. Testing is a process of executing a program with the intent of finding an error.
2. A good test case is one that has a high probability of finding an as yet undiscovered error.
3. A successful test is one that uncovers an as yet undiscovered error.

Testing should systematically uncover different classes of errors in a minimum amount of time and with a minimum amount of effort. A secondary benefit of testing is that it demonstrates that the software appears to be working as stated in the specifications. The data collected through testing can also provide an indication of the software's reliability and quality. But, testing cannot show the absence of defect -- it can only show that software defects are present.

(2) UNIT TESTING

The procedure level testing is made first. By giving improper inputs, the errors occurred are noted and eliminated. Then the other testing level is made.

(3) INTEGRATION TESTING

Testing is done for each module. After testing all the modules, the modules are integrated and testing of the final system is done with the test data, specially designed to show that the system will operate successfully in all its aspects conditions. Thus the system testing is a confirmation that all is correct and an opportunity to show the user that the system works.

(4) VALIDATION TESTING

The final step involves Validation testing, which determines whether the software function as the user expected. The end-user rather than the system developer conduct this test most software developers as a process called “Alpha and Beta Testing” to uncover that only the end user seems able to find. The compilation of the entire project is based on the full satisfaction of the end users. In the project, validation testing is made in various forms.

In question entry form, the correct answer only will be accepted in the answer box. The answers other than the four given choices will not be accepted.

REFERENCES

[1] Boldyreva, N. Chenette, and A. O’neill, “Order-preserving encryption revisited: Improved security analysis and solutions,” in Proc. Int.Cryptol. Conf. CRYPTO, 2011.
[2] Boldyreva, N. Chenette, Y. Lee, and A. O’neill, “Order-preserving symmetric encryption,” in Proc. Annu. Int. Conf. Theory Appl. ryptograph.Tech., 2009.
[3] M. Burkhart and X. Dimitropoulos, “Fast privacy preserving top-k queries using secret sharing,” in Proc. 19th ICCCN, 2010.
[4] D. Boneh, C. Gentry, B. Lynn, and H. Shacham, “Aggregate and veriably encrypted signatures from bilinear maps,” in Proc. Int. Conf. Theory Appl. Cryptograph. Tech. Adv. Cryptol., 2003.
[5] M. Bellare and S. K. Miner, “A forward-secure digital signature scheme, in Proc. 19th Annu. Int. Cryptol. Conf. Adv. Cryptol.
[6] F. Chen and A. X. Liu, “SafeQ: Secure and efficient query processing in sensor networks,” in Proc. 24th IEEE Conf. Comput. Commun.,

Mar. 2010.
[7] Dwork, "Differential privacy," in Proc. ICALP, 2006.
[8] J. Vaidya and C. Clifton, "Privacy-preserving top-k queries," in Proc. IEEE 21st ICDE, Apr. 2005, pp. 545–546.
[9] Wu, J. Xu, X. Tang, and W.-C. Lee, "Top-k monitoring in wireless sensor networks," IEEE Trans. Knowl. Data Eng., vol. 19, no. 7, pp. 962–976, Jul. 2007.
[10] M. Yu, C.-S. Lu, and S.-Y. Kuo, "Non interactive pairwise key establishment for sensor networks," IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 556–569, Sep. 2010.
[11] F. Ye, H. Luo, S. Lu, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," in Proc. IEEE 23rd Annu. Joint Conf. Comput. Commun. INFOCOM, Mar. 2004, pp. 2446–2457.
[12] M. Yu, G.-K. Ni, I.-Y. Chen, E. Gelenbe, and S.-Y. Kuo, "Top-k query result completeness verification in sensor networks," in Proc. IEEE Int. ICC Workshops, Jun. 2013, pp. 1026–1030.
[13] T.-H. You, W.-C. Peng, and W.-C. Lee, "Protecting moving trajectories with dummies," in Proc. Int. Conf. Mobile Data Manag., May 2007, pp. 278–282.
[14] C.-M. Yu, Y.-T. Tsou, C.-S. Lu, and S.-Y. Kuo, "Practical and secure multidimensional query framework in tiered sensor networks," IEEE Trans. Inf. Forensics Security, vol. 6, no. 2, pp. 241–255, Jun. 2011.
[15] R. Zhang, J. Shi, Y. Liu, and Y. Zhang, "Verifiable fine-grained top- k queries in tiered sensor networks," in Proc. 24th IEEE INFOCOM, Mar. 2010, pp. 1–9.
[16] Zhang, J. Shi, and Y. Zhang, "Secure multidimensional range queries in sensor networks," in Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput., 2009, pp. 197–206.