International Journal of Engineering Research and General Science Volume 3, Issue 5, September-October, 2015
ISSN 2091-2730

LOCATION BASED QUERY ON LOCATION SERVER FOR PRIVACY EFFICIENCY

Mr. Naveen Kumar.Chundy
Computer Science and Engineering
BVSR Engineering College
Chimakurthy A.P. India
naveenchundy@gmail.com

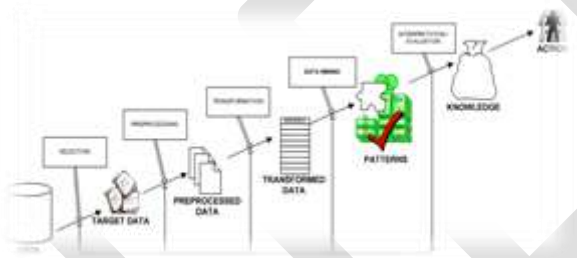
Mrs. K.Suma Anusha M.Tech
Asst.Professor Dept.of CSE
BVSR Engineering College
Chimakurthy A.P. India
Kondaveeti.suma@gmail.com

Abstract— In this paper we present a solution to one of the location-based query problems. This problem is defined as follows: (i) a user wants to query a database of location data, known as Points Of Interest (POIs), and does not want to reveal his/her location to the server due to privacy concerns; (ii) the owner of the location data, that is, the location server, does not want to simply distribute its data to all users. The location server desires to have some control over its data, since the data is its asset. We propose a major enhancement upon previous solutions by introducing a two stage approach, where the first step is based on Oblivious Transfer and the second step is based on Private Information Retrieval, to achieve a secure solution for both parties. The solution we present is efficient and practical in many scenarios. We implement our solution on a desktop machine and a mobile device to assess the efficiency of our protocol. We also introduce a security model and analyse the security in the context of our protocol. Finally, we highlight a security weakness of our previous work and present a solution to overcome it.

Keywords—componentdata mining,computing,users,mobile service providers,Location Server,civil laibility,Black mail,Free storage

1. Introduction

What is Data Mining?



Structure of Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

How Data Mining Works?

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. **Generally, any of four types of relationships are sought:**

- **Classes**: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

- **Clusters**: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

- **Associations**: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

- **Sequential Patterns**: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

1) Extract, transform, and load transaction data onto the data warehouse system.
2) Store and manage the data in a multidimensional database system.
3) Provide data access to business analysts and information technology professionals.
4) Analyze the data by application software.
5) Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

- **Artificial Neural Networks**: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

- **Genetic Algorithms**: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

- **Decision Trees**: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.

- **Nearest Neighbor Method**: A technique that classifies each record in a dataset based on a combination of the classes of the *k* record(s) most similar to it in a historical dataset (where *k*=1). Sometimes called the *k*-nearest neighbor technique.

- **Rule Induction**: The extraction of useful if-then rules from data based on statistical significance.

- **Data Visualization**: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

Characteristics of Data Mining:

- **Large Quantities of Data**: The volume of data so great it has to be analyzed by automated techniques e.g. satellite information, credit card transactions etc.
- **Noisy, Incomplete Data**: Imprecise data is the characteristic of all data collection.
- **Complex Data Structure**: conventional statistical analysis not possible
- **Heterogeneous Data Stored In Legacy Systems**

Benefits of Data Mining:

1) It's one of the most effective services that are available today. With the help of data mining, one can discover precious information about the customers and their behavior for a specific set of products and evaluate and analyze, store, mine and load data related to them
2) An analytical CRM model and strategic business related decisions can be made with the help of data mining as it helps in providing a complete synopsis of customers
3) An endless number of organizations have installed data mining projects and it has helped them see their own companies make an unprecedented improvement in their marketing strategies (Campaigns)
4) Data mining is generally used by organizations with a solid customer focus. For its flexible nature as far as applicability is concerned is being used vehemently in applications to foresee crucial data including industry analysis and consumer buying behaviors
5) Fast paced and prompt access to data along with economic processing techniques have made data mining one of the most suitable services that a company seek.

Advantages of data mining:

1. *Marketing / Retail:*

Data mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaigns such as direct mail, online marketing campaign…etc. Through the results, marketers will have appropriate approach to sell profitable products to targeted customers.

Data mining brings a lot of benefits to retail companies in the same way as marketing. Through market basket analysis, a store can have an appropriate production arrangement in a way that customers can buy frequent buying products together with pleasant. In addition, it also helps the retail companies offer certain discounts for particular products that will attract more customers.

2. *Finance / Banking*

Data mining gives financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner.

3. *Manufacturing*

By applying data mining in operational engineering data, manufacturers can detect faulty equipments and determine optimal control parameters. For example semi-conductor manufacturers has a challenge that even the conditions of manufacturing environments at different wafer production plants are similar, the quality of wafer are lot the same and some for unknown reasons even has defects. Data mining has been applying to determine the ranges of control parameters that lead to the production of golden wafer. Then those optimal control parameters are used to manufacture wafers with desired quality.

4. *Governments*

Data mining helps government agency by digging and analyzing records of financial transaction to build patterns that can detect money laundering or criminal activities.

5. Law enforcement:

Data mining can aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviours.

6. Researchers:

Data mining can assist researchers by speeding up their data analyzing process; thus, allowing those more time to work on other projects.

What is Secure Computing?

Computer security (Also known as cyber security or IT Security) is information security as applied to computers and networks. The field covers all the processes and mechanisms by which computer-based equipment, information and services are protected from unintended or unauthorized access, change or destruction. Computer security also includes protection from unplanned events and natural disasters. Otherwise, in the computer industry, the term security -- or the phrase computer security -- refers to techniques for ensuring that data stored in a computer cannot be read or compromised by any individuals without authorization. Most computer security measures involve data encryption and passwords. Data encryption is the translation of data into a form that is unintelligible without a deciphering mechanism. A password is a secret word or phrase that gives a user access to a particular program or system.

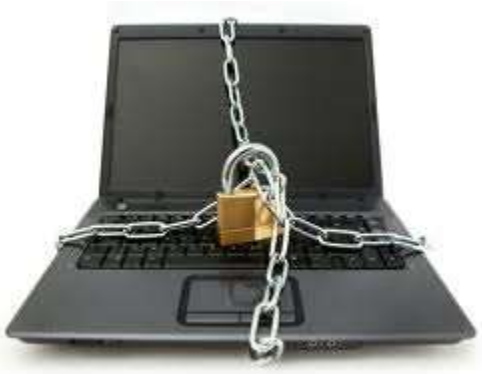


Diagram clearly explain the about the secure computing

Working conditions and basic needs in the secure computing:

If you don't take basic steps to protect your work computer, you put it and all the information on it at risk. You can potentially compromise the operation of other computers on your organization's network, or even the functioning of the network as a whole.

1. Physical security:

Technical measures like login passwords, anti-virus are essential. (More about those below) However, a secure physical space is the first and more important line of defense.

Is the place you keep your workplace computer secure enough to prevent theft or access to it while you are away? While the Security Department provides coverage across the Medical center, it only takes seconds to steal a computer, particularly a portable device like a laptop or a PDA. A computer should be secured like any other valuable possession when you are not present.

Human threats are not the only concern. Computers can be compromised by environmental mishaps (e.g., water, coffee) or physical trauma. Make sure the physical location of your computer takes account of those risks as well.

2. Access passwords:

The University's networks and shared information systems are protected in part by login credentials (user-IDs and passwords). Access passwords are also an essential protection for personal computers in most circumstances. Offices are usually open and shared spaces, so physical access to computers cannot be completely controlled.

To protect your computer, you should consider setting passwords for particularly sensitive applications resident on the computer (e.g., data analysis software), if the software provides that capability.

3. Prying eye protection:

Because we deal with all facets of clinical, research, educational and administrative data here on the medical campus, it is important to do everything possible to minimize exposure of data to unauthorized individuals.

4. Anti-virus software:

Up-to-date, properly configured anti-virus software is essential. While we have server-side anti-virus software on our network computers, you still need it on the client side (your computer).

5. Firewalls:

Anti-virus products inspect files on your computer and in email. Firewall software and hardware monitor communications between your computer and the outside world. That is essential for any networked computer.

6. Software updates:

It is critical to keep software up to date, especially the operating system, anti-virus and anti-spyware, email and browser software. The newest versions will contain fixes for discovered vulnerabilities.

Almost all anti-virus have automatic update features (including SAV). Keeping the "signatures" (digital patterns) of malicious software detectors up-to-date is essential for these products to be effective.

7. Keep secure backups:

Even if you take all these security steps, bad things can still happen. Be prepared for the worst by making backup copies of critical data, and keeping those backup copies in a separate, secure location. For example, use supplemental hard drives, CDs/DVDs, or flash drives to store critical, hard-to-replace data.

8. **Report problems:**

If you believe that your computer or any data on it has been compromised, your should make a information security incident report. That is required by University policy for all data on our systems, and legally required for health, education, financial and any other kind of record containing identifiable personal information.

Benefits of secure computing:

- **Protect yourself - Civil liability**:
  You may be held legally liable to compensate a third party should they experience financial damage or distress as a result of their personal data being stolen from you or leaked by you.
- **Protect your credibility - Compliance**:
  You may require compliancy with the Data Protection Act, the FSA, SOX or other regulatory standards. Each of these bodies stipulates that certain measures be taken to protect the data on your network.
- **Protect your reputation – Spam:**
  A common use for infected systems is to join them to a bot net (a collection of infected machines which takes orders from a command server) and use them to send out spam. This spam can be traced back to you, your server could be blacklisted and you could be unable to send email.
- **Protect your income - Competitive advantage:**
  There are a number of "hackers-for-hire" advertising their services on the internet selling their skills in breaking into company's servers to steal client databases, proprietary software, merger and acquisition information, personnel detail set al.
- **Protect your business – Blackmail**:
  A seldom-reported source of income for "hackers" is to· break into your server, change all your passwords and lock you out of it. The password is then sold back to you. Note: the "hackers" may implant a backdoor program on your server so that they can repeat the exercise at will.
- **Protect your investment - Free storage:**
  Your server's hard drive space is used (or sold on) to house the hacker's video clips, music collections, pirated software or worse. Your server or computer then becomes continuously slow and your internet connection speeds deteriorate due to the number of people connecting to your server in order to download the offered wares.

2. System Analysis

Existing system:

The Location Server (LS), which offers some LBS, spends its resources to compile information about various interesting POIs. Hence, it is expected that the LS would not disclose any information without fees. Therefore the LBS has to ensure that LS's data is not accessed by any unauthorized user. During the process of transmission the users should not be allowed to discover any information for which they have not paid. It is thus crucial that solutions be devised that address the privacy of the users issuing queries, but also prevent users from accessing content to which they do not have authorization.

Disadvantages of existing system:

- Among many challenging barriers to the wide deployment of such application, privacy assurance is a major issue
- The user can get answers to various location based queries,

Proposed system:

- In this paper, we propose a novel protocol for location based queries that has major performance improvements with respect to the approach by Ghinita at el. And. Like such protocol, our protocol is organized according to two stages. In the first stage, the user privately determines his/her location within a public grid, using oblivious transfer. This data contains both the ID and associated symmetric key for the block of data in the private grid. In the second stage, the user executes a communicational efficient PIR, to retrieve the appropriate block in the private grid. This block is decrypted using the symmetric key obtained in the previous stage.
- Our protocol thus provides protection for both the user and the server. The user is protected because the server is unable to determine his/her location. Similarly, the server's data is protected since a malicious user can only decrypt the block of data obtained by PIR with the encryption key acquired in the previous stage. In other words, users cannot gain any more data than what they have paid for. We remark that this paper is an enhancement of a previous work.

Advantages of proposed system:

- ✓ Redesigned the key structure.
- ✓ Added a formal security model.

Implemented the solution on both a mobile device and desktop machine. □□□□□□□□□□□□□ □□□ □□□

3. Implementation

MODULES:

1. Users
2. Mobile Service Provider
3. Location Server

Modules Description:

Users:

The users in our model use some location-based service provided by the location server LS. For example, what is he nearest ATM or restaurant? The purpose of the mobile service provider SP is to establish and maintain the communication between the location server and the user. The location server LS owns a set of POI records ri for $1 \leq ri \leq \rho$. Each record describes a POI, giving GPS coordinates to its location (xgps,ygps), and a description or name about what is at the location.

Mobile Service Provider:

We reasonably assume that the mobile service provider SP is a passive entity and is not allowed to collude with the LS. We make this assumption because the SP can determine the whereabouts of a mobile device, which, if allowed to collude with the LS, completely subverts any method for privacy. There is simply no technological method for preventing this attack. As a consequence of this assumption, the user is able to either use GPS (Global Positioning System) or the mobile service provider to acquire his/her coordinates.

Location Server:

We are assuming that the mobile service provider SP is trusted to maintain the connection, we consider only two possible adversaries. Each and every one for individual communication direction. We consider the case in which the user is the adversary and tries to obtain more than he/she is allowed. Next we consider the case in which the location server LS is the adversary, and tries to uniquely associate a user with a grid coordinate.

4. Conclusion

In this paper we have presented a location based query solution that employs two protocols that enables a user to privately determine and acquire location data. The first step is for a user to privately determine his/her location using oblivious transfer on a public grid. The second step involves a private information retrieval interaction that retrieves the record with high communication efficiency. We analysed the performance of our protocol and found it to be both computationally and communication ally more efficient than the solution by Ghinita *et al.*, which is the most recent solution. We implemented a software prototype using a desktop machine and a mobile device. The software prototype demonstrates that our protocol is within practical limits.

Future work will involve testing the protocol on many different mobile devices. The mobile result we provide may be different than other mobile devices and software environments. Also, we need to reduce the overhead of the primality test used in the private information retrieval based protocol. Additionally, the problem concerning the LS supplying misleading data to the client is also interesting. Privacy preserving reputation techniques seem a suitable approach to address such problem. Once suitable strong solutions exist for the general case, they can be easily integrated into our approach.

REFERENCES:

[1] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan,"Private queries in location based services: anonymizers are not necessary," in Proc. 2008 ACM SIGMOD Int. Conf. Management of Data, New York, NY, USA, 2008, pp. 121–132, ser. SIGMOD'08, ACM.

[2] E. H. Moore, "On certain crinkly curves," Trans. Amer. Math. Soc., vol. 1, pp. 72–90, Jan. 1900.

[3] H. Sagan, Space-Filling Curves. New York, NY, USA: Springer- Verlag, 1994.

[4] M. Naor and B. Pinkas, "Oblivious transfer with adaptive queries," Proc. CRYPTO'99, 1999, vol. 1666, pp. 791 - 791.

[5] M. Bellare and S. Micali, "Non-interactive oblivious transfer and applications," Proc. CRYPTO'89. 1990, pp. 547 - 557.

[6] M. Mokbel, "Towards privacy-aware location-based database servers," in Proc. 22nd Int. Conf. Data Engineering Workshops, 2006, pp. 93–102.

[7] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location- based identity inference in anonymous spatial queries," IEEE Trans. Knowl. Data Eng., vol. 19, no. 12, pp. 1719–1733, Dec. 2007.

[8] A.-A. Hossain, A. Hossain, H.-K. Yoo, and J.-W. Chang, "H-star: Hilbert-order based star network expansion cloaking algorithm in road networks," in Proc. IEEE 14th Int. Conf. Computational Science and Engineering (CSE), Aug. 2011, pp. 81–88.

[9] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inform. Theory*, vol. 31, no. 4, pp. 469–472, Jul. 1985.

[10] B. Gedik and L. Liu, "Location privacy in mobile systems: A personalizedanonymization model," in *Proc. ICDCS*, Columbus, OH, USA, 2005, pp. 620–629.

[11] C. Gentry and Z. Ramzan, "Single-database private information retrieval with constant communication rate," in *Proc. ICALP*, L. Caires, G. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, Eds., Lisbon, Portugal, 2005, pp. 803–815, LNCS 3580.

[12] Marco Gruteser and Dirk Grunwald. Anonymous usage of locationbased services through spatial and temporal cloaking. In Proceedings of the 1st international conference on Mobile systems, applications and services, MobiSys '03, pages 31–42, New York, NY, USA, 2003. ACM.

[13] Ling Liu Bugra Gedik. A customizable k-anonymity model for protecting location privacy. Technical Report GIT-CERCS-04-15, Georgia Institute of Technology, April 2004. [14] Chi-Yin Chow, Mohamed F. Mokbel, and Xuan Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems, GIS '06, pages 171–178, New York, NY, USA, 2006. ACM.

[15] Reza Shokri, Julien Freudiger, Murtuza Jadliwala, and Jean-Pierre Hubaux. A distortion-based metric for location privacy. In Proceedings of the 8th ACM workshop on Privacy in the electronic society, WPES '09, pages 21–30, New York, NY, USA, 2009. ACM