PRESERVING LOCATION PRIVACY IN GEOSOCIAL APPLICATIONS

S.Ramya,(M.Sc)
Department ofComputerScience
Kamban College of Arts and Science for Women, Tiruvannamalai.

Prof R.AngelinPreethi,M.Sc.,M.Phil(P.hD)
Department of Computer Science
Kamban College of Arts andScience for Women, Tiruvannamalai.

Abstract

Using geosocial applications, such as Four Square, millions of people interact with their surroundings through their friends and their recommendations. Without adequate privacy protection, however, these systems can be easilymisused, for example, to track users or target them for home invasion. In this paper, we introduce LocX, a novelalternativethatprovidessignificantlyimprovedlocationprivacywithoutaddinguncertaintyintoqueryresults or relying on strong assumptions about server security. Our key insight is to apply secure user-specific, distance- preserving coordinate transformations to all location data shared with the server. The friends of a user share this user's secrets so they can apply the same transformation. This allows all location queries to be evaluated correctly by the server, but our privacy mechanisms guarantee that servers are unable to see or infer the actual location data from the transformed data or from the data access. We show that LocX provides privacy even against a powerful adversary model, and we use prototype measurements to show that it provides privacy with very little performance overhead, making it suitable for today's mobiledevices.

1. INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it intouseful information-informationthatcanbeusedtoincrease 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 evolvingseparatetransactionandanalyticalsystems, 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 aresought:**

- **Classes**: Stored data is used to locate data in predetermined groups. For example, arestaurant 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 dailyspecials.

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

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

- **Sequentialpatterns**:Dataisminedtoanticipate 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 hikingshoes.

Data mining consists of five major elements:

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

Different levels of analysis areavailable:

- **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 whileCHAIDsegmentsusingchisquaretests to create multi-way splits. CART typically requires less data preparation thanCHAID.

- **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 neighbortechnique.

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

- **Data visualization**: The visual interpretation of complex relationships inmultidimensional 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 transactionsetc.
- **Noisy, incomplete data**: Imprecise data is the characteristic of all datacollection.
- **Complex data structure**: conventional statistical analysis notpossible
- **Heterogeneous data stored in legacysystems 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

customersandtheirbehaviorforaspecificset of products and evaluate and analyze, store, mine and load data related tothem

2) AnanalyticalCRMmodelandstrategicbusiness related decisions can be made with the help of data mining as it helps in providing a complete synopsis ofcustomers
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 beingusedvehementlyinapplicationstoforesee crucial data including industry analysis and consumer buyingbehaviors
5) Fast paced and prompt access to data along with economic processing techniques have madedata mining one of the most suitable services that a companyseek.

Advantages of Data Mining:

1. **Marketing /Retail:**

Data mining helps marketing companies build models based on historical data to predict who will respond to thenewmarketingcampaignssuchasdirectmail,online marketing campaign…etc. Through the results, marketers will have appropriate approach to sell profitable products to targetedcustomers.

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 desiredquality.

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 behaviors.

6. Researchers:

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

LITERATURESURVEY

Juels and B.S. Kaliski(2007) A POR scheme enables an archive or back-up service(prover) to produce a concise proof that a user(verifier) can retrieve a target file F, that is, that the archive retains and reliably transmits file datasufficient for the user to recover F in its entirety. The goal of a PORistoaccomplishthesechecks without users having to download files themselves. A POR can also provide quality-of-service guarantees, i.e. show that a file is retrievablewithinacertaintimebound.APORprotocol in which the verifier stores only a single cryptographic key—irrespective of the size and number of the files whose retrievability it seeks to verify as well asa amount of dynamic state for each file. Insymmetric-keycryptographyandefficienterror-coding,wetrust that our sentinel-based POR protocol is amenable to real-world application. As storage-as-a-service spreads and users rely on external agents to store critical information, the privacy and integrity guaranteesofconventionalcryptographywillbenefit from

extensions in to POR-based assurances around dataavailability.

H.Shacham,B.Waters(2008) In a proof-of- retrievability system, a data storage center must prove to verifier that is actually storing all of a clients data. The first proof-of-retrievability schemes with full proofs of security againstarbitrary adversaries in the strongest model. First scheme, built from BLS signatures and secure in the random oracle model, as the shortest query and response of any proof-of-retrievability with public verifiability. Second scheme, which builds elegantly on pseudo random functions (PRFs) and is secure in the standard model, as the shortest response of any proof-of-retrievability with privateverifiability.

Kevin D.Bowers, Alina Oprea (2009) A proof of retrievability (POR) is a file system (prover) to a client (verifier) that a target file F is intact, in the sense that the client can fully recoverit. A therotical framework for the design of PORs. Our framework improves the previously proposed POR constructions of Juels-Kaliski and Shacham-Waters, and also sheds light on the conceptual limitations of previous therotical models for PORs. Cloud computing, the trend toward loosely coupled networking of computing resources, is unmooring data local storage platforms. It is worth exploring further optimizations in our implementation to enhance the encoding throughput. An interesting practical problems to design different encoding technique with a minimal numbers of disk accesses for very large files. We leave the open the problems ofdesigningefficientPORprotocolsthatsupportfile updates, as well as publicly verifiablePORs.

M. A. Shah, R. Swaminathan(2010) Agrowingnumberofonlineservicessuchasgoogle, yahoo !are starting to charge users for their storage. Customers often use these services to store valuable data such as email, family photos and videos, and disk backups. To make storage services accountable for data loss, that allows a protocol for third party auditor to periodically verify the data stored by a service and assist in returning the data intact to the customer.Thestorageservicesmustexporthooksfor challenges to the response queries and compute extensive functions for responses. To avoid these overheads we can batch many files together into single key and check that fill all at once. Hybrid as mentioned but is still imposes the end and overheads that the storage service and customer experience with our protocols. It will eliminate the encryption techniques and extend the formal definition of provable data possession and proof of retrievability. Auditor to be trusted and hat collude with either party.

G. Ateniese, R.D. Pietro (2008)

Provable Data Possession (PDP) is recently appearedin the research literature. The main issue is how to frequently, efficiently and securely verify that storage server is faithfully storing its client's (potentially very large)outsourceddata.Thestorageserverisassumedto be untrusted in terms of both security andreliability

.Prior work has addressed this problem using either public key cryptography or requiring the client to outsource its data in encrypted form .A highly efficient and provably secure PDP technique based entirely on symmetric key cryptography, while not requiring any bulk encryption. The concept of third-party data warehousing and, more generally, data outsourcing has become quite popular. Early work concentrated on data authentication and integrity. The central goal in PDP is allow a client to efficiently, frequently and securely verify that a server-who purportedly stores clients potentiallyverylargeamountofdata-isnotcheatingthe client.

Q. Wang, K. Ren, W. Lou(2009) A distributed data storage as gained increasing popularity for efficient and robust data management in wireless sensor networks (WSNs). But the distributed architecture also makes it challenging to build a highly secure and dependable yet lightweight data storage system. To address the challenges, we propose adependable and secure data storage scheme with dynamic integrity assurance. Based on the principle of secret sharing and erasure coding it propose a hybrid share generation and distribution scheme to achieve reliable and fault-tolerant initial datastorage by providing redundancy for original data components. To the best of our knowledge, distributeddatastorageandaccesssecurityasafairly new area has received limited attentions. Data Integrity and availability is an important and necessary component of secure data storage for distributed sensor networks. Our goal is to provide various mechanisms for ensuring and maintaining the security and dependability of sensed network data under the aforementioned adversary model. To ensure the integrity of data shares, an efficient dynamic data integrity checking scheme is constructed based on the principle of algebraic signatures. In existing approaches more desirable properties and advantages are achieved. In our schemeweapproachhighlysecureandefficient,thus can be implemented in the current generation of sensornetworks.

R.Curtmola, O.Khan (2011) As storage-outsourcing services and resource- sharing networks have become popular, the problem of efficiently proving the integrity of data stored at untrusted servers has received increased attention.In the provable data possession (PDP) model the client preprocesses the data and then sends it to an untrusted server for storage,

while keeping a small amount of meta-data. A definitional framework and efficient constructions for dynamic provable data possession(DPDP),whichextendsthePDPmodelto support provable updates to stored data. In cloud storage systems, the server (or peer) that stores the client's data is not necessarily trusted. The main contributionofthisareprovidethefirstefficientfully dynamic PDP solution next present a rank-based authenticated dictionary built over a skip list. This constructionyieldsaDPDPschemewithlogarithmic computation and communication and the same detectionprobabilityastheoriginalPDPscheme.We give an alternative construction of a rank-based authenticated dictionary using an RSA tree. This construction results in a DPDP scheme with improved detection probability but higher server computation. To evaluate the performance of our DPDP scheme interms of communication and computational overhead, in order to determine the price of dynamism over static PDP. In version controltoevaluateanapplicationthatsuitsour scheme's ability to efficiently handle and prove updates to versioned, hierarchical resources.

2. EXISTINGSYSTEM

QOS (Quality of Service) is usually defined as a set of non-functional properties, such as

- Responsetime
- Throughput
- Reliability and soon

QOS-based Web service discovery and selection has garnered much attention from both academia and industry.

DRAWBACKS

It is impractical for a user to acquire QOS information by invoking all of the service candidates. And some QOS properties (e.g., reputation and reliability) are difficult to be evaluated, since they require both long observation duration and a large number ofinvocations.

PROPOSED SYSTEM

We proposed an enhanced measurement for computing QOS similarity between different users and between different services. The measurement takes into account the personalized deviation of Web services' QOS and users' QOS experiences, in order to improve the accuracy of similarity computation. Based on the above enhanced similarity measurement, we proposed a location-aware CF-based Web service QOS prediction method for service recommendation.

FEATURES

Our location-aware QOS prediction method has a solid basis, because of the strong relation between the locations of users (or Web services) and the Web services'QOSperceivedbytheusers.Weconductedan experiment to evaluate the impact of data sparseness on the prediction coverage, in which, our proposed methodswerecomparedwiththetraditionalCFmethod.

3. DESCRIPTION OFMODULES

- DataSegment
- UserRegistration
- Locationgrouping
- Service providerschema
- Report

i. DATASEGMENT

Users will be clustered into different regions according to their locations and historical QOS records. At the beginning, we retrieve users' approximate locations by their IP addresses. The location information reveals a user's country, city, latitude/longitude, ISP and domain name. Then users from the same city will be grouped together to form initial regions. All the data are grouped into one segment in database.

ii. USERREGISTRATION

A new user has been registered with the correct details. He/she has been logged into the services. So he/she able to search the location.

Before the search has been start he must be the registered user.

iii. LOCATIONGROUPING

Clustering Web services can help Location recipient to find potential similar services. Different from retrieving user location from an IP address, LoRec directly clusters Web services based on their QOS similarity. All the location has been grouped once the non-registered locationhasbeensearchbytheuser.SoGrouping play the major role here to provide thedetails.

iv. SERVICE PROVIDERSCHEMA

The first two phase aggregate users and Web services into a certain number of clusters based on their respective similarities. QOS predictions can be generated from both service regions and user regions. With the compressed QOS data, searching neighbors

and making Web service QOS predictions for an active user can be computed faster than conventional methods.

First, when a user searches Web services using LoRec, predicted QOS values will be shown next to each candidate service, and the one with the best predicted value will be highlighted in the search result for the activeuser.

v. REPORT

The report show the correct details of each user with the location process.

CONCLUSION

To ensure cloud data storage security, it is critical to enable a third party auditor to evaluate the service quality from an objective and independent perspective. Public auditability also allows client to delegate the integrity verification tasks to TPA while they themselves can be reliable or not be able to commit necessary computation resources performing continuousverifications.Itsupportsbatchauditingupon delegations from multi-users. Based on RSAAlgorithm instantiation, computation cost of server and verifier will be reduced. Our construction is deliberately designed to meet an efficient securityprovided by TPA. Extensive security and performance analysis show that the proposed scheme is highly efficient and provably secure.

4. REFERENCES

[1] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enablingpublic verifiabilityanddatadynamicsforstoragesecurity in cloud computing," in Proc. Of ESORICS'09. Saint Malo, France: SpringerVerlag, 2009, pp. 355–370.

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrustedstores," in Proc. of CCS'07. New York, NY, USA: ACM, 2007, pp. 598–609.

[3] A. Juels and B. S. Kaliski, Jr., "Pors: proofs of retrievabilityfor large files," in Proc. of CCS'07. New York, NY, USA: ACM, 2007, pp. 584–597.

[4] H. Shacham and B. Waters, "Compact proofs of retrievability,"in Proc. of ASIACRYPT'08. Melbourne, Australia: Springer-Verlag, 2008, pp. 90–107.

[5] K. D. Bowers, A. Juels, and A. Oprea, "Proofs of retrievability: Theory and implementation," Cryptology ePrint Archive,Report 2008/175, 2008.

[6] M.A.Shah,R.Swaminathan,"Privacy-preserving audit and extraction of digita contents," CryptologyePrintArchive, Report 2008/186, 2008.

[7] A. Oprea, M. K. Reiter, "Remote integrity check withdishonest storage server," in Proc. of ESORICS'08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 223–237.