1. Data mining method for Traffic accident Severity prediction

ABSTRACT –

The growth of the population volume and the number of vehicles on the road cause congestion (jam) in cities that is one of the main transportation issues. Congestion can lead to negative effects such as increasing accident risks due to the expansion in transportation systems. The smart city concept provides opportunities to handle urban problems, and also to improve the citizens' living environment. In recent years, road traffic accidents (RTAs) have become one of the largest national health issues in the world. Many factors (driver, environment, car, etc.) are related to traffic accidents, some of those factors are more important in determining the accident severity than others. The analytical data mining solutions can significantly be employed to determine and predict such influential factors among human, vehicle and environmental factors and thus to explain RTAs severity. In this research, three classification techniques were applied: Decision trees (Random Forest, Random Tree, J48/C4.5, and CART), ANN (back-propagation), and SVM (polynomial kernel) to detect the influential environmental features of RTAs that can be used to build the prediction model. These techniques were tested using a real dataset obtained from the Department for Transport of the United Kingdom. The experimental results showed that the highest accuracy value was 80.6% using Random Forest followed by 61.4% using ANN then by 54.8% using SVM. A decision system has been build using the model generated by the Random Forest technique that will help decision makers to enhance the decision making process by predicting the severity of the accident.

2. Enhancing the Naive Bayes Spam Filter Through Intelligent Text Modification Detection

ABSTRACT-

Spam emails have been a chronic issue in computer security. They are very costly economically and extremely dangerous for computers and networks. Despite of the emergence of social networks and other Internet based information exchange venues, dependence on email communication has increased over the years and this dependence has resulted in an urgent need to improve spam filters. Although many spam filters have been created to help prevent these spam emails from entering a user's inbox, there is a lack or research focusing on text modifications. Currently, Naive Bayes is one of the most popular methods of spam classification because of its simplicity and efficiency. Naive Bayes is also very accurate; however, it is unable to correctly classify emails when they contain leetspeak or diacritics. Thus, in this proposes, we implemented a novel algorithm for enhancing the accuracy of the Naive Bayes Spam Filter so that it can detect text modifications and correctly classify the email as spam or ham. Our Python algorithm combines semantic based, keyword based, and machine learning algorithms to increase the accuracy of Naive Bayes compared to Spamassassin by over two hundred percent. Additionally, we have discovered a relationship between the length of the email and the spam score, indicating that Bayesian Poisoning, a controversial topic, is actually a real phenomenon and utilized by spammers.

3. Job satisfaction and employee turnover: A firm-level perspective

ABSTRACT-

How companies can use their personnel data and information from job satisfaction surveys to predict employee quits. An important issue discussed at length in the article is how employers can ensure the anonymity of employees in surveys used for management and human resources (HR) analytics. I argue that a simple mechanism whereby the company delegates the implementation of job satisfaction surveys to an external consulting company can be optimal. In the subsequent empirical analysis, I use a unique combination of firm-level data (personnel records) and information from job satisfaction surveys to assess the benefits for companies using data in their decision-making. Moreover, I aim to show how companies can move from a descriptive to a predictive approach.

4. Medical decision making diagnosis system integrating k-means and Naïve Bayes algorithms

ABSTRACT-

Using data mining we can evaluate many patterns which will be use in future to make intelligent systems and decisions By data mining refers to various methods of identifying information or the adoption of solutions based on knowledge and data extraction of these data so that they can be used in various areas such as decision-making, the prediction value for the prediction and calculation. In our days the health industry has collected vast amounts of patient data, which, unfortunately, is not "produced" in order to give some hidden information, and thus to make effective decisions, which are connected with the base of the patient's data and are subject to data mining. This research work has developed a Decision Support in Heart Disease Prediction System (HDPS) using data mining modelling technique, namely, Naïve Bayes and K-means clustering algorithms that are one of the most popular clustering techniques; however, where the initial choice of the centroid strongly influences the final result. Using of medical data, such as age, sex, blood pressure and blood sugar levels, chest pain, electrocardiogram, analyzes of different study patient, etc. graphics can predict the likelihood of the patient. This paper shows the effectiveness of unsupervised learning techniques, which is a k-means clustering to improve teaching methods controlled, which is naive Bayes. It explores the integration of K-means clustering with naive Bayes in the diagnosis of disease patients. It also investigates different methods of initial centroid selection of the K-means clustering such as range, inlier, outlier, random attribute values, and random row methods in the diagnosis of heart disease patients. The results indicate that the integration of the K-means clustering with naïve Bayes with different initial centroid selecting naive Bayesian improve accuracy in diagnosis of the patient.

5. Information extraction methods for text documents in a Cognitive Integrated Management Information System

ABSTRACT-

In contemporary companies unstructured knowledge is essential, mainly due to the possibility to obtain better flexibility and competitiveness of the organization. For example, on the basis of automatic analysis of the experts' opinions, the decision-makers are capable of taking decisions (for example decisions concerning investments). This paper presents issues related to developing and evaluating a methods of information extraction performed by cognitive agent running in integrated management information system. The main advantages of this approach are cognitive agents' ability of including a context of extracted information and its ability of automatic decision-making on the basis of extracted information.

6. Techniques for sentiment analysis of Twitter data: A comprehensive survey

ABSTRACT-

The World Wide Web has intensely evolved a novel way for people to express their views and opinions about different topics, trends and issues. The user-generated content present on different mediums such as internet forums, discussion groups, and blogs serves a concrete and substantial base for decision making in various fields such as advertising, political polls, scientific surveys, market prediction and business intelligence. Sentiment analysis relates to the problem of mining the sentiments from online available data and categorizing the opinion expressed by an author towards a particular entity into at most three preset categories: positive, negative and neutral. In this paper, firstly we present the sentiment analysis process to classify highly unstructured data on Twitter. Secondly, we discuss various techniques to carryout sentiment analysis on Twitter data in detail. Moreover, we present the parametric comparison of the discussed techniques based on our identified parameters.

7. Real-time vehicle detection and tracking

ABSTRACT-

The rapid increase in the number of the automobiles on the highway and urban roads have created many challenges regarding the proper management and control of the traffic. Detection and tracking of vehicles using the traffic surveillance system gives more promising way to manage and control the road traffic. Vehicle surveillance represents a challenging task of moving object segmentation in complex environment. The detection ratio of such algorithms depends upon the quality of the generated foreground mask. Therefore, the aim of this paper is to present an efficient method for detection and tracking of vehicles which focuses on the trajectory of motion of the objects. The proposed method preserves the group of pixels in foreground which can be probable vehicles and discards the rest as noise. Therefore, it selectively rejects the objects which cannot be vehicles at the same time consolidate the candidate vehicles. Here, the foreground mask generation process is improved so that the quality of generated foreground mask better consequently increases the detection ratio. The performance of the proposed method is evaluated by comparing it with other standard methods qualitatively as well as quantitatively. The experimental results have established the superior performance of the proposed method.

8. A novelistic approach to analyse weather conditions and its prediction using deep learning techniques

ABSTRACT - To predict the weather conditions based on the features of the data collected over the past data and to design a model which can allow to predict the future occurence of the event and also gives the accuracy of the different models used.

9. Social &A: An Online Social Network Based Question And Answer System

Abstract—

Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s)he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented SocialQ&A, an online social network based Q&A system. SocialQ&A leverages the social network properties of commoninterest and mutual-trust friend relationship to identify an asker's friends who are most likely to answer the question. We describe the architecture, algorithms and user interface of SocialQ&A, and analyze the Q&A behavior of real users and questions from a small-scale real-world SocialQ&A system. We also conducted comprehensive large-scale simulation to evaluate SocialQ&A in comparison with other methods. Our results suggest that social networks can be leveraged to improve the answer quality and asker's waiting time.

10. Privacy-Preserving Data Encryption Strategy For Big Data In Mobile Cloud Computing

ABSTRACT-

Technology is enhancing each and every day especially in the field of Information Technology and data is very momentous elements. The large volume of data generated through devices is a major obstacle to handle in real time. The accomplishment of data ciphering is a crucial problem during the data progression and dissemination. In order to achieve pursuance, many application disregards data encryption. This paper represents a concern about data privacy and suggests a novel data encryption approach known as Dynamic Data Encryption Strategy (DDES). This accession is drafted to magnify privacy protection scope within liquidate time constraints.

11. Review Spam Detection Using Machine Learning

Abstract:

Prior to buying a product, people usually inform themselves by reading online reviews. To make more profit sellers often try to fake user experience. As customers are being deceived this way, recognizing and removing fake reviews is of great importance. This paper analyzes spam detection methods, based on machine learning, and presents their overview and results.

12.Efficient Processing Of Skyline Queries Using Mapreduce

ABSTRACT-

This research presents an advanced MapReduce-based parallel solution to efficiently address spatial skyline queries on large datasets. In particular, given a set of data points and a set of query points, we first generate the convex hull of the query points in the first Map Reduce phase. Then, we propose a novel concept called independent regions, for parallelizing the process of spatial skyline evaluation. Spatial skyline candidates in an independent region do not depend on any data point in other independent regions. Thus, we calculate the independent regions based on the input data points and the convex hull of the query points in the second phase. With the independent regions, spatial skylines are evaluated in parallel in the third phase, in which data points are partitioned by their associated independent regions in the map functions, and spatial skyline candidates are calculated by reduce functions. The results of the spatial skyline queries are the union of outputs from the reduce functions. Due to high cost of the spatial dominance test, which requires comparing the distance from data points to all convex points, we propose a concept of pruning regions in independent regions. All data points in pruning regions can be discarded without the dominance test. Our experimental results show the efficiency and effectiveness of the proposed parallel spatial skyline solution utilizing Map Reduce on large-scale real-world and synthetic datasets.

13. Data Partitioning In Frequent Itemset Mining On Hadoop Clusters

Abstract –

For mining frequent Itemsets parallel traditional algorithms are used. Existing parallel Frequent Itemsets mining algorithm partition the data equally among the nodes. These parallel Frequent Itemsets mining algorithms have high communication and mining overheads. We resolve this problem by using data partitioning strategy. It is based on Hadoop. The core of Apache Hadoop consists of a storage part, called as Hadoop Distributed File System (HDFS), and a processing part called Map Reduce. Hadoop divides files into large blocks. It distributes them across nodes in a cluster. By using this strategy the performance of existing parallel frequent-pattern increases.

14. Study On Secure Data Deduplication System With Application Awareness Over Cloud Storage Systems

Abstract—

This paper gives study of data deduplication technique and how to secure data on cloud with deduplication scheme. Data deduplication is single instance data storage widely used in cloud storage system to reduce space and upload bandwidth. Duplication-less storage system which de duplicates the data using file level deduplication and block level deduplication. The deduplication is done at source where data is generated and at target where data is stored that is cloud storage system. The deduplication storage system de duplicates data with application awareness index structure. This system consists of two major components, a front-end deduplication and a cloud storage system as back-end. Application aware index structure improves data deduplication efficiency by exploiting application awareness and deduplication time reduction. But for secure deduplication with application awareness we are incorporating the convergent encryption. To deal with security issue metadata has to create first and then encrypt the file to upload to cloud storage.

15. Preserving Mapreduce Based K-Means Clustering Over

Abstract-

Clustering techniques have been widely adopted in many real world data analysis applications, such as customer behavior analysis, targeted marketing, digital forensics, etc. With the explosion of data in today's big data era, a major trend to handle a clustering over large-scale datasets is outsourcing it to public cloud platforms. This is because cloud computing offers not only reliable services with performance guarantees, but also savings on in-house IT infrastructures. However, as datasets used for clustering may contain sensitive information, e.g., patient health information, commercial data, and behavioral data, etc, directly outsourcing them to public cloud servers inevitably raise privacy concerns.

16. Emotion Recognition On Twitter: Comparative Study And Training A Unison Model

Abstract-

Despite recent successes of deep learning in many fields of natural language processing, previous studies of emotion recognition on Twitter mainly focused on the use of lexicons and simple classifiers on bag-of-words models. The central question of our study is whether we can improve their performance using deep learning. To this end, we exploit hashtags to create three large emotion-labeled data sets corresponding to different classifications of emotions. We then compare the performance of several wordand character-based recurrent and convolutional neural networks with the performance on bag-of-words and latent semantic indexing models. We also investigate the transferability of the final hidden state representations between different classifications of emotions, and whether it is possible to build a unison model for predicting all of them using a shared representation. We show that recurrent neural networks, especially character-based ones, can improve over bag-of-words and latent semantic indexing models. Although the transfer capabilities of these models are poor, the newly proposed training heuristic produces a unison model with performance comparable to that of the three single models.

17. Hierarchy-Cutting Model Based Association Semantic For Analyzing Domain Topic On The Web

Abstract –

Traditional text mining techniques for semantic association transform free text into association semantic and Points of Interests (POI) for knowledge discovery. However, it does not consider time and precision while performing linear inseparability from textual web content. Effective relation association semantic technology using machine learning analysis not only can reduce time for knowledge discovery from textual web content, but also can improve the searching accuracy of the related information system. How to realize the semantic relation mining by the machine learning analysis is an important research topic. In this paper, a three-step procedure to mine associations of semantic relations for textual web document content, called, Polynomial Kernelized Maximum Entropy and Support Vector Machine (PKME-SVM) framework is presented. First, Associative Polynomial Kernel-based Maximum Entropy representing semantic relations are extracted from raw text web contents using Polynomial Kernel. The semantic relation extraction process also creates a sentence grammar tree in the form of reduced sentence. Then, Probabilistic Term (i.e. word) Taxonomy (PTT) Framework is applied to discover the probabilistic term on corresponding Web Content Domain. Finally, for pruning the semantic relation from textual web content, the Generalized Association Support Vector Mining algorithm adopts the notion of dual characteristic function for systematic overgeneralization reduction. The objective of PKME-SVM is to obtain accurate result from textual web content using semantic relation mining operations and satisfy the web user specific needs through polynomial kernel. The efficacy of our framework is demonstrated through empirical experiments conducted on Freebase Data Dump. Experiment is conducted on the factors such as personalized information search retrieval rate, computation time and precision ratio.

18. Large-Scale Multi-Modality Attribute Reduction With Multi-Kernel Fuzzy Rough Sets

Abstract-

Fuzzy rough sets have been successfully applied in classification tasks, in particular in combination with OWA operators. There has been a lot of research into adapting algorithms for use with Big Data through parallelisation, but no concrete strategy exists to design a Big Data fuzzy rough sets based classifier. Existing Big Data approaches use fuzzy rough sets for feature and prototype selection, and have often not involved very large datasets. We fill this gap by presenting the first Big Data extension of an algorithm that uses fuzzy rough sets directly to classify test instances, a distributed implementation of FRNN-OWA in Apache Spark. Through a series of systematic tests involving generated datasets, we demonstrate that it can achieve a speedup effectively equal to the number of computing

19.A Secure And Verifiable Access Control Scheme For Big Data Storage In Clouds

ABSTRACT:

Due to the complexity and volume, outsourcing ciphertexts to a cloud is deemed to be one of the most effective approaches for big data storage and access. Nevertheless, verifying the access legitimacy of a user and securely updating a ciphertext in the cloud based on a new access policy designated by the data owner are two critical challenges to make cloud-based big data storage practical and effective. Traditional approaches either completely ignore the issue of access policy update or delegate the update to a third party authority; but in practice, access policy update is important for enhancing security and dealing with the dynamism caused by user join and leave activities. In this paper, we propose a secure and verifiable access control scheme based on the NTRU cryptosystem for big data storage in clouds. We first propose a new NTRU decryption algorithm to overcome the decryption failures of the original NTRU, and then detail our scheme and analyze its correctness, security strengths, and computational efficiency.

20.A Low Cost Web Based Remote Monitoring System With Built-In Security Feature For Vulnerable Environments

ABSTRACT –

`This research paper proposes an intelligent monitoring system based on the embedded technology. Embedded system is a special computer system which combines both hardware and software. The aim of remote monitoring is to measure various environmental phenomena of a targeted area and to alert when the boundary values of the parameters controlling has been exceeded. To alert the authorized personnel, the mode of communication implemented in the system is wireless. When something undesired happens, the system automatically alerts the authorizers by sending SMS through GSM/GPRS modem.This paper also explains how to reduce the latency of the parameters measured which sometimes leads to imperfection and disasters in the large as well as small scale industries or in home automation.

21.A Lightweight Secure Data Sharing Scheme For Mobile Cloud Computing

Abstract –

With the popularity of cloud computing, mobile devices can store/retrieve personal data from anywhere at any time. Consequently, the data security problem in mobile cloud becomes more and more severe and prevents further development of mobile cloud. There are substantial studies that have been conducted to improve the cloud security. However, most of them are not applicable for mobile cloud since mobile devices only have limited computing resources and power. Solutions with low computational overhead are in great need for mobile cloud applications. In this paper, we propose a lightweight data sharing scheme (LDSS) for mobile cloud computing. It adopts CP-ABE, an access control technology used in normal cloud environment, but changes the structure of access control tree to make it suitable for mobile cloud environments. LDSS moves a large portion of the computational intensive access control tree transformation in CP-ABE from mobile devices to external proxy servers. Furthermore, to reduce the user revocation cost, it introduces attribute description fields to implement lazy-revocation, which is a thorny issue in program based CP-ABE systems. The experimental results show that LDSS can effectively reduce the overhead on the mobile device side when users are sharing data in mobile cloud environments.

22. A Non-Invasive Remote Health Monitoring System Using Visible Light Communication

Abstract-

Telemedicine is a rapidly developing application of clinic medicine where medical information is transferred through the phone or other networks for the purpose of consulting and performing remote medical procedures or examinations. Telemedicine can be applied to a greater extend in the field of cardiology where ECG serves as the major tool. This project elaborates the experience; a methodology adopted and highlights various design aspects to be considered for making telemedicine in patient monitoring system effective. In this method, the patient's vital signs like heart beating rate, temperature and glucose level are captured and the values are continually displayed on the doctor's phone using VLC system.

23. Designing A Secure Exam Management System (Sems) For M-Learning Environments.

Abstract—

M-learning has enhanced the e-learning by making the learning process learner-centered. However, enforcing exam security in open environments where each student has his/her own mobile/tablet device connected to a Wi-Fi network through which it is further connected to the Internet can be one of the most challenging tasks. In such environments, students can easily exchange information over the network during exam time. This paper aims to identify various vulnerabilities that may violate exam security in mlearning environments and to design the appropriate security services and countermeasures that can be put in place to ensure exam security. It also aims to integrate the resulting secure exam system with an existing, open-source, and widely accepted Learning Management System (LMS) and its service extension to the m-learning environment, namely “the Moodbile Project”.

24.A Provably Secure General Construction For Key Exchange Protocols Using Smart Card And Password

Abstract:

Quite recently, two smart-card-based passwords authenticated key exchange protocols were proposed by Lee et al. and Hwang et al. respectively. However, neither of them achieves two-factor authentication fully since they would become completely insecure once one factor is broken. To overcome these congenital defects, this study proposes such a secure authenticated key exchange protocol that achieves fully two-factor authentication and provides forward security of session keys. And yet our scheme is simple and reasonably efficient. Furthermore, we can provide the rigorous proof of the security for it.

25. Multi-Party Secret Key Agreement Over State-Dependent Wireless Broadcast Channels

Abstract-

We consider a group of m trusted and authenticated nodes that aim to create a shared secret key K over a wireless channel in the presence of an eavesdropper Eve. We assume that there exists a state dependent wireless broadcast channel from one of the honest nodes to the rest of them including Eve. All of the trusted nodes can also discuss over a cost-free, noiseless and unlimited rate public channel which is also overheard by Eve. For this setup, we develop an information-theoretically secure secret key agreement protocol. We show the optimality of this protocol for "linear deterministic" wireless broadcast channels. This model generalizes the packet erasure model studied in literature for wireless broadcast channels. For "state-dependent Gaussian" wireless broadcast channels, we propose an achievability scheme based on a multi-layer wiretap code. Finding the best achievable secret key generation rate leads to solving a non-convex power allocation problem. We show that using a dynamic programming algorithm, one can obtain the best power allocation for this problem. Moreover, we prove the optimality of the proposed achievability scheme for the reg.

26.Someone In Your Contact List: Cued Recall-Based Textual Passwords

Abstract –

Even today, in many applications textual passwords are used as a traditional approach of authentication. These textual passwords are not secured and could be easily guessed. Moreover textual passwords might be gained with the techniques such as brute force, dictionary attacks, social engineering, and shoulder surfing and spyware attacks. Therefore, information is not secured by using textual passwords as a method of authentication. To overcome these problems, graphical passwords authentication is evolved as an alternative method for textual password. In this graphical password authentication, images are used as passwords in place of text, because images are easy to remember than text. A graphical password consists of clicking images or drag the images, or rotating the images as a password and not typing the text in textual passwords. Generally, graphical password techniques are categorized into three main categories: recall based, recognition based and hybrid based. In Recall based techniques, user has to reproduce a drawing without giving any hint as a password. In recognition based technique, user has to select or recognize the images from different sets of images and that image is same as the image selected at the time of registration phase. In hybrid based technique, the combination of two more techniques involved in recall based or recognition based technique or both. Recall based techniques are classified into two categories: pure recall based technique and cued recall based technique.

27.Authorship Attribution For Social Media Forensics

Abstract—

Authorship verification is the task of analyzing the linguistic patterns of two or more texts to determine whether they were written by the same author or not. The analysis is traditionally performed by experts who consider linguistic features, which include spelling mistakes, grammatical inconsistencies, and stylistics for example. Machine learning algorithms, on the other hand, can be trained to accomplish the same, but have traditionally relied on so-called stylometric features. The disadvantage of such features is that their reliability is greatly diminished for short and topically varied social media texts. In this interdisciplinary work, we propose a substantial extension of a recently published hierarchical Siamese neural network approach, with which it is feasible to learn neural features and to visualize the decision-making process. For this purpose, a new large-scale corpus of short Amazon reviews for text comparison research is compiled and we show that the Siamese network topologies outperform state-of-the-art approaches that were built up on stylometric features. Our linguistic analysis of the internal attention weights of the network shows that the proposed method is indeed able to latch on to some traditional linguistic categories. Index Terms—Authorship verification, similarity learning, forensic text comparison, Siamese network, deep metric learning

28.Text Based Graphical Password System To Obscure Shoulder Surfing

Abstract—

This paper presents a authentication technique where it is having an secure way of hiding the data from hackers and unauthorized user. Here, it will create a different barriers for the hacker so that he may get confused to hack the secure data of the account. There is a critical problem about hacking of data in a networking world. So, to prevent that hacking here it will create a image based password where there is having a random codes below the images in a gallery so that it cannot be copied by any user, and mainly it will prevent the shoulder surfing attack which is very popular nowadays. The images are displayed in a gallery according to the choice of an user interest where he will choose his question and according to that question the relevant images are displayed in it. As it will provide the security to the user's account while logging his account. There is no chances of proxy password in a login page ,it will capture easily the attacker and block the account. As, it is resistance to all types of attacks and mainly it will provide a excellent security to the user's information. It is better than a textual based password and prevents the image gallery attack.

29.Privacy-Preserving Location Sharing Services For Social Networks

Abstract—

A common functionality of many location-based social networking applications is a location sharing service that allows a group of friends to share their locations. With a potentially untrusted server, such a location sharing service may threaten the privacy of users. Existing solutions for Privacy-Preserving Location Sharing Services (PPLSS) require a trusted third party that has access to the exact location of all users in the system or rely on expensive algorithms or protocols in terms of computational or communication overhead. Other solutions can only provide approximate query answers. To overcome these limitations, we propose a new encryption notion, called Order-Retrievable Encryption (ORE), for PPLSS for social networking applications. The distinguishing characteristics of our PPLSS are that it (1) allows a group of friends to share their exact locations without the need of any third party or leaking any location information to any server or users outside the group, (2) achieves low computational and communication cost by allowing users to receive the exact location of their friends without requiring any direct communication between users or multiple rounds of communication between a user and a server, (3) provides efficient query processing by designing an index structure for our ORE scheme, (4) supports dynamic location updates, and (5) provides personalized privacy protection within a group of friends by specifying a maximum distance where a user is willing to be located by his/her friends. Experimental results show that the computational and communication cost of our PPLSS is much better than the state-of-the-art solution.

30. Prism: Privacy-Aware Interest Sharing And Matching In Mobile Social Networks

Abstract—

In a profile matchmaking application of mobile social networks, users need to reveal their interests to each other in order to find the common interests. A malicious user may harm a user by knowing his personal information. Therefore, mutual interests need to be found in a privacy preserving manner. In this paper, we propose an efficient privacy protection and interests sharing protocol referred to as PRivacy-aware Interest Sharing and Matching (PRISM). PRISM enables users to discover mutual interests without revealing their interests. Unlike existing approaches, PRISM does not require revealing the interests to a trusted server. Moreover, the protocol considers attacking scenarios that have not been addressed previously and provides an efficient solution. The inherent mechanism reveals any cheating attempt by a malicious user. PRISM also proposes the procedure to eliminate Sybil attacks. We analyze the security of PRISM against both passive and active attacks. Through implementation, we also present a detailed analysis of the performance of PRISM and compare it with existing approaches. The results show the effectiveness of PRISM without any significant Performance degradation.

31. Illusionpin: Shoulder-Surfing Resistant Authentication Using Hybrid Images

Abstract-

Shoulder surfing attack is the direct observation of user from far distance by hacker. Traditional methods use personal identification number (pin) consists of a sequence of digits for authentication. This method is used for Digital Authentication of touch screen devices. The applications of touch screen devices include ATM machines, Smart phones and Kiosk. Shoulder surfing attack suffers from various issues, Challenges and limitations like security and privacy. There are various algorithms and techniques have been proposed in the literature to overcome these difficulties and still needs improvement. Hence in this work a novel algorithm using illusion pin with hybrid images for shoulder surfing attack authentication scheme has been proposed. This proposed method using Illusion-pin (I-pin) blends of two keypads with different ordering digits using hybrid images. The user keypads is shuffled in every authentication attempt. This method is used to restrict the shoulder surfing attack by implementing this visibility algorithm. Hence hackers are unable to recognize or learning the user pin which provides more security and authentication.

32. Reversible Data Hiding In Encrypted Images Using Interpolation-Based Distributed Space Reservation

Abstract –

With the development of cloud storage and privacy protection, reversible data hiding in encrypted images (RDHEI) has attracted increasing attention as a technology that can embed additional data in the encryption domain. In general, an RDHEI method embeds secret data in an encrypted image while ensuring that the embedded data can be extracted error-free and the original image can be restored lossless. In this paper, A high-capacity RDHEI algorithm is proposed. At first, the Most Significant Bits (MSB) of each pixel was predicted adaptively and marked by Huffman coding in the original image. Then, the image was encrypted by a stream cipher method. At last, the vacated space can be used to embed additional data. Experimental results show that our method achieved higher embedding capacity while comparing with the state-of-the-art methods.

33. Vehicular Cloud Data Collection For Intelligent Transportation Systems

Abstract—

The Internet of Things (IoT) envisions to connect billions of sensors to the Internet, in order to provide new applications and services for smart cities. IoT will allow the evolution of the Internet of Vehicles (IoV) from existing Vehicular Ad hoc Networks (VANETs), in which the delivery of various services will be offered to drivers by integrating vehicles, sensors, and mobile devices into a global network. To serve VANET with computational resources, Vehicular Cloud Computing (VCC) is recently envisioned with the objective of providing traffic solutions to improve our daily driving. These solutions involve applications and services for the benefit of Intelligent Transportation Systems (ITS), which represent an important part of IoV. Data collection is an important aspect in ITS, which can effectively serve online travel systems with the aid of Vehicular Cloud (VC). In this paper, we involve the new paradigm of VCC to propose a data collection model for the benefit of ITS. We show via simulation results that the participation of low percentage of vehicles in a dynamic VC is sufficient to provide meaningful data collection.

34.Cost Minimization Algorithms For Data Center Management

Abstract:

Due to the increasing usage of cloud computing applications, it is important to minimize energy cost consumed by a data center, and simultaneously, to improve quality of service via data center management. One promising approach is to switch some servers in a data center to the idle mode for saving energy while to keep a suitable number of servers in the active mode for providing timely service. In this paper, we design both online and offline algorithms for this problem. For the offline algorithm, we formulate data center management as a cost minimization problem by considering energy cost, delay cost (to measure service quality), and switching cost (to change servers's active/idle mode). Then, we analyze certain properties of an optimal solution which lead to a dynamic programming based algorithm. Moreover, by revising the solution procedure, we successfully eliminate the recursive procedure and achieve an optimal offline algorithm with a polynomial complexity.

35. Multi-Party Secret Key Agreement Over State-Dependent Wireless Broadcast Channels

Abstract –

We consider a group of m trusted and authenticated nodes that aim to create a shared secret key K over a wireless channel in the presence of an eavesdropper Eve. We assume that there exists a state dependent wireless broadcast channel from one of the honest nodes to the rest of them including Eve. All of the trusted nodes can also discuss over a cost-free, noiseless and unlimited rate public channel which is also overheard by Eve. For this setup, we develop an information-theoretically secure secret key agreement protocol. We show the optimality of this protocol for "linear deterministic" wireless broadcast channels. This model generalizes the packet erasure model studied in literature for wireless broadcast channels. For "state-dependent Gaussian" wireless broadcast channels, we propose an achievability scheme based on a multi-layer wiretap code. Finding the best achievable secret key generation rate leads to solving a non-convex power allocation problem. We show that using a dynamic programming algorithm, one can obtain the best power allocation for this problem. Moreover, we prove the optimality of the proposed achievability scheme for the regime of high-SNR and large-dynamic range over the channel states in the (generalized) degrees of freedom sense.

36.Optimizing Cloud-Service Performance: Efficient Resource Provisioning Via Optimal Workload Allocation

Abstract—

Cloud computing is being widely accepted and utilized in the business world. From the perspective of businesses utilizing the cloud, it is critical to meet their customers' requirements by achieving service-level-objectives. Hence, the ability to accurately characterize and optimize cloud-service performance is of great importance. In this paper a stochastic multi-tenant framework is proposed to model the service of customer requests in a cloud infrastructure composed of heterogeneous virtual machines. Two cloud-service performance metrics are mathematically characterized, namely the percentile and the mean of the stochastic response time of a customer request, in closed form. Based upon the proposed multi-tenant framework, a workload allocation algorithm, termed max-min-cloud algorithm, is then devised to optimize the performance of the cloud service. A rigorous optimality proof of the max-min-cloud algorithm is also given. Furthermore, the resource-provisioning problem in the cloud is also studied in light of the max-min-cloud algorithm. In particular, an efficient resource-provisioning strategy is proposed for serving dynamically arriving customer requests. These findings can be used by businesses to build a better.

37. Cost Minimization Algorithms For Data Center Management

Abstract:

Due to the increasing usage of cloud computing applications, it is important to minimize energy cost consumed by a data center, and simultaneously, to improve quality of service via data center management. One promising approach is to switch some servers in a data center to the idle mode for saving energy while to keep a suitable number of servers in the active mode for providing timely service. In this paper, we design both online and offline algorithms for this problem. For the offline algorithm, we formulate data center management as a cost minimization problem by considering energy cost, delay cost (to measure service quality), and switching cost (to change servers's active/idle mode). Then, we analyze certain properties of an optimal solution which lead to a dynamic programming based algorithm. Moreover, by revising the solution procedure, we successfully eliminate the recursive procedure and achieve an optimal offline algorithm with a polynomial complexity.

38. Attribute-Based Storage Supporting Secure Deduplication Of Encrypted Data In Cloud

Abstract—

Attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Compared with the prior data deduplication systems, our system has two advantages. Firstly, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Secondly, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under other access policies without revealing the underlying plaintext.

39. Efficient Resource Constrained Scheduling Using Parallel Two-Phase Branch-And-Bound Heuristics

Abstract:

Due to the fact that project scheduling under resource constraints problem is one of the most intractable problems in Operations Research, it has recently become a popular area for the latest optimization techniques, including virtually all local search paradigms. The sheer diversity and momentum of activity has made developments in project scheduling increasingly difficult to track and assimilate. This paper provides a high- level bibliography, structured overview and limited critique of nine project scheduling under resource constraints problems; resource-constrained project scheduling problem (RCPSP), preemptive resource-constrained project scheduling problem (PRCPSP), generalized resource-constrained project scheduling problem (GRCPSP) , resource-constrained project scheduling problem with generalized precedence relations (RCPSP-GPR), Time/cost trade-off problems (TCTP) , Discrete time/resource trade-off problems (DTRP), Multi-mode resource-constrained project scheduling problems (MRCPSP), Resource levelling problems (RLP) and Resource-constrained project scheduling with discounted cash flows (RCPSDC). The current developments, strengths, and weaknesses of the scheduling approaches to the stated problems are considered.

40. Identity based data out sourcing with comprehensive e auditing in clouds

Abstract—

Cloud storage system provides facilitative file storage and sharing services for distributed clients. To address integrity, controllable outsourcing and origin auditing concerns on outsourced files, we propose an identity-based data outsourcing (IBDO) scheme equipped with desirable features advantageous over existing proposals in securing outsourced data. First, our IBDO scheme allows a user to authorize dedicated proxies to upload data to the cloud storage server on her behalf, e.g., a company may authorize some employees to upload files to the company's cloud account in a controlled way. The proxies are identified and authorized with their recognizable identities, which eliminates complicated certificate management in usual secure distributed computing systems. Second, our IBDO scheme facilitates comprehensive auditing, i.e., our scheme not only permits regular integrity auditing as in existing schemes for securing outsourced data, but also allows to audit the information on data origin, type and consistence of outsourced files. Security analysis and experimental evaluation indicate that our IBDO scheme provides strong security with desirable efficiency.

41. Joint Policy- And Network-Aware Vm Management For Cloud Data Centers

Abstract—

Policies play an important role in network configuration and therefore in offering secure and high performance services especially over multi-tenant Cloud Data Center (DC) environments. At the same time, elastic resource provisioning through virtualization often disregards policy requirements, assuming that the policy implementation is handled by the underlying network infrastructure. This can result in policy violations, performance degradation and security vulnerabilities. In this paper, we define PLAN, a PoLicy-Aware and Networkaware VM management scheme to jointly consider DC communication cost reduction through Virtual Machine (VM) migration while meeting network policy requirements. We show that the problem is NP-hard and derive an efficient approximate algorithm to reduce communication cost while adhering to policy constraints. Through extensive evaluation, we show that PLAN can reduce topology-wide communication cost by 38% over diverse aggregate traffic and configuration policies.

42. Race: Robust And Auditable Access Control With Multiple Attribute Authorities For Public Cloud Storage

Abstract:

Data access control is a challenging issue in public cloud storage systems. Ciphertext-Policy Attribute-Based Encryption (CP-ABE) has been adopted as a promising technique to provide flexible, fine-grained and secure data access control for cloud storage with honest-but-curious cloud servers. However, in the existing CP-ABE schemes, the single attribute authority must execute the time-consuming user legitimacy verification and secret key distribution, and hence it results in a single-point performance bottleneck when a CP-ABE scheme is adopted in a large-scale cloud storage system. Users may be stuck in the waiting queue for a long period to obtain their secret keys, thereby resulting in low-efficiency of the system. Although multiauthority access control schemes have been proposed, these schemes still cannot overcome the drawbacks of single-point bottleneck and low efficiency, due to the fact that each of the authorities still independently manages a disjoint attribute set. In this paper, we propose a novel heterogeneous framework to remove the problem of single-point performance bottleneck and provide a more efficient access control scheme with an auditing mechanism. Our framework employs multiple attribute authorities to share the load of user legitimacy verification. Meanwhile, in our scheme, a CA (Central Authority) is introduced to generate secret keys for legitimacy verified users. Unlike other multiauthority access control schemes, each of the authorities in our scheme manages the whole attribute set individually. To enhance security, we also propose an auditing mechanism to detect which AA (Attribute Authority) has incorrectly or maliciously performed the legitimacy verification procedure. Analysis shows that our system not only guarantees the security requirements but also makes great performance improvement on key generation.

43. Live Data Analytics With Collaborative Edge And Cloud Processing In Wireless Iot Networks

ABSTRACT:

Recently, big data analytics has received important attention in a variety of application domains including business, finance, space science, healthcare, telecommunication and Internet of Things (IoT). Among these areas, IoT is considered as an important platform in bringing people, processes, data and things/objects together in order to enhance the quality of our everyday lives. However, the key challenges are how to effectively extract useful features from the massive amount of heterogeneous data generated by resource-constrained IoT devices in order to provide real-time information and feedback to the end-users, and how to utilize this data-aware intelligence in enhancing the performance of wireless IoT networks. Although there are parallel advances in cloud computing and edge computing for addressing some issues in data analytics, they have their own benefits and limitations. The convergence of these two computing paradigms, i.e., massive virtually shared pool of computing and storage resources from the cloud and real-time data processing by edge computing, could effectively enable live data analytics in wireless IoT networks. In this regard, we propose a novel framework for coordinated processing between edge and cloud computing/processing by integrating advantages from both the platforms. The proposed framework can exploit the network wide knowledge and historical information available at the cloud center to guide edge computing units towards satisfying various performance requirements of heterogeneous wireless IoT networks. Starting with the main features, key enablers and the challenges of big data analytics, we provide various synergies and distinctions between cloud and edge processing. More importantly, we identify and describe the potential key enablers for the proposed edge-cloud collaborative framework, the associated key challenges and some interesting future research directions.

44.A general frame work for edited video and raw video summarization

Abstract:

Video summaries provide condensed and succinct representations of the content of a video stream through a combination of still images, video segments, graphical representations and textual descriptors. This paper presents a conceptual framework for video summarisation derived from the research literature and used as a means for surveying the research literature. The framework distinguishes between video summarisation techniques (the methods used to process content from a source video stream to achieve a summarisation of that stream) and video summaries (outputs of video summarisation techniques). Video summarisation techniques are considered within three broad categories: internal (analyse information sourced directly from the video stream), external (analyse information not sourced directly from the video stream) and hybrid (analyse a combination of internal and external information). Video summaries are considered as a function of the type of content they are derived from (object, event, perception or feature based) and the functionality offered to the user for their consumption (interactive or static, personalised or generic). It is argued that video summarisation would benefit from greater incorporation of external information, particularly user based information that is unobtrusively sourced, in order to overcome longstanding challenges such as the semantic gap and providing video summaries that have greater relevance to individual users.

45.A semi-Automatic and trust worthys cheme for continuous cloud service Certification Constraints

Abstract-

Cloud computing offers the prospect of on-demand, elastic computing, provided as a utility service, and it is revolutionizing many domains of computing. Compared with earlier methods of processing data, cloud computing environments provide significant benefits, such as the availability of automated tools to assemble, connect, configure and reconfigure virtualized resources on demand. These make it much easier to meet organizational goals as organizations can easily deploy cloud services. However, the shift in paradigm that accompanies the adoption of cloud computing is increasingly giving rise to security and privacy considerations relating to facets of cloud computing such as multi-tenancy, trust, loss of control and accountability. Consequently, cloud platforms that handle sensitive information are required to deploy technical measures and organizational safeguards to avoid data protection breakdowns that might result in enormous and costly damages. Sensitive information in the context of cloud computing encompasses data from a wide range of different areas and domains. Data concerning health is a typical example of the type of sensitive information handled in cloud computing environments, and it is obvious that most individuals will want information related to their health to be secure. Hence, with the growth of cloud computing in recent times, privacy and data protection requirements have been evolving to protect individuals against surveillance and data disclosure. Some examples of such protective legislation are the EU Data Protection Directive (DPD) and the US Health Insurance Portability and Accountability Act (HIPAA), both of which demand privacy preservation for handling personally identifiable information.

46.A Cloud-Integrated, Multi layered Agent-Based Cyber-Physical System Architecture

Abstract—

Advances in digital electronics have led to a significant increase in the number of systems that couple the digital (cyber) systems with the physical world, namely what have become known as the Cyber-Physical System (CPS). The design of CPS requires a significant amount of reasoning with respect to unique challenges and complex functional, reliability, and performance requirements. This paper presents the suitability of Multi-Agent System technology towards resolving some challenges in the design of CPS. In the proposed method, challenges that are common to most CPS design and the attributes and behaviour characteristics of agents in MultiAgent System to tackle such challenges are outlined. The correctness of this approach is shown from the perspective of Agent-Oriented Programming (AOP) with Java Agent Development (JADE) Platform where agents exhibit autonomous and interactive behaviour. The capability of agents in MAS to interact with one another and with their environments; their flexibility and freedom to monitor, control and change their behaviours are considered appropriate solutions to the prominent CPS design challenges.

47.A Collision-Mitigation Cuckoo Hashing Scheme For Large-Scale Storage Systems

Abstract—

With the rapid growth of the amount of information, cloud computing servers need to process and analyze large amounts of high-dimensional and unstructured data timely and accurately. This usually requires many query operations. Due to simplicity and ease of use, cuckoo hashing schemes have been widely used in real-world cloud-related applications. However, due to the potential hash collisions, the cuckoo hashing suffers from endless loops and high insertion latency, even high risks of re-construction of entire hash table. In order to address these problems, we propose a cost-efficient cuckoo hashing scheme, called MinCounter. The idea behind Min Counter is to alleviate the occurrence of endless loops in the data insertion by selecting un busy kicking-out routes. MinCounter selects the “cold” (infrequently accessed), rather than random, buckets to handle hash collisions. We further improve the concurrency of the Min Counter scheme to pursue higher performance and adapt to concurrent applications. Min Counter has the salient features of offering efficient insertion and query services and delivering high performance of cloud servers, as well as enhancing the experiences for cloud users. We have implemented Min Counter in a large-scale cloud testbed and examined the performance by using three real world traces. Extensive experimental results demonstrate the efficacy and efficiency of Min Counter.

48. Optimizing Green Energy, Cost, And Availability In Distributed Data Centers

Abstract:

Integrating renewable energy and ensuring high availability are among two major requirements for geo distributed data centers. Availability is ensured by provisioning spare capacity across the data centers to mask data center failures (either partial or complete). We propose a mixed integer linear programming formulation for capacity planning while minimizing the total cost of ownership (TCO) for highly available, green, distributed data centers. We minimize the cost due to power consumption and server deployment, while targeting a minimum usage of green energy.Solving our model shows that capacity provisioning considering green energy integration, not only lowers carbon footprint but also reduces the TCO. Results show that upto 40% green energy usage is feasible with marginal increase in the TCO compared to the other cost-aware models.

49. Identity-Based Remote Data Integrity Checking With Perfect Data Privacy Preserving For Cloud Storage

Abstract—

Remote data integrity checking (RDIC) enables a data storage server, such as a cloud server, to prove to a verifier that it is actually storing a data owner's data honestly. To date, a number of RDIC protocols have been proposed in the literature, but almost all the constructions suffer from the issue of a complex key management, that is, they rely on the expensive public key infrastructure (PKI), which might hinder the deployment of RDIC in practice. In this paper, we propose a new construction of identity-based (ID-based) RDIC protocol by making use of key-homomorphic cryptographic primitive to reduce the system complexity and the cost for establishing and managing the public key authentication framework in PKI based RDIC schemes. We formalize ID-based RDIC and its security model including security against a malicious cloud server and zero knowledge privacy against a third party verifier. We then provide a concrete construction of ID-based RDIC scheme which leaks no information of the stored files to the verifier during the RDIC process. The new construction is proven secure against the malicious server in the generic group model and achieves zero knowledge privacy against a verifier. Extensive security analysis and implementation results demonstrate that the proposed new protocol is provably secure and practical in the real-world applications.

50. Privacy-Preserving Data Encryption Strategy For Big Data In Mobile Cloud Computing

ABSTRACT:

Technology is enhancing each and every day especially in the field of Information Technology and data is very momentous elements. The large volume of data generated through devices is a major obstacle to handle in real time. The accomplishment of data ciphering is a crucial problem during the data progression and dissemination. In order to achieve pursuance, many application disregards data encryption. This paper represents a concern about data privacy and suggests a novel data encryption approach known as Dynamic Data Encryption Strategy (DDES). This accession is drafted to magnify privacy protection scope within liquidate time constraints.

51. Two-Factor Data Access Control With Efficient Revocation For Multi authority Cloud Storage Systems

Abstract:

Attribute-based encryption, particularly for figure content arrangement trait based encryption, can satisfy the usefulness of fine-grained get to control in distributed storage frameworks. Since client's properties might be issued by different characteristic experts, multi-specialist figure content strategy property based encryption is a rising cryptographic crude for upholding trait construct get to control in light of outsourced information. Be that as it may, the majority of the current multi-specialist property based frameworks are either unreliable in characteristic level disavowal or absence of proficiency in correspondence overhead and calculation cost. In this paper, we propose a quality based access control plot with two-factor security for multi-specialist distributed storage frameworks. In our proposed conspire, any client can recuperate the outsourced information if and just if this client holds adequate property mystery keys as for the entrance arrangement and approval enter with respect to the outsourced information. Moreover, the proposed plot appreciates the properties of consistent size figure content and little calculation cost. Other than supporting the property level denial, our proposed conspire enables information proprietor to complete the client level repudiation. The security investigation, execution examinations, and test comes about demonstrate that our proposed plot isn't just secure yet in addition down to earth.

52. Vehicular Cloud Data Collection For Intelligent Transportation Systems

Abstract—

The Internet of Things (IoT) envisions to connect billions of sensors to the Internet, in order to provide new applications and services for smart cities. IoT will allow the evolution of the Internet of Vehicles (IoV) from existing Vehicular Ad hoc Networks (VANETs), in which the delivery of various services will be offered to drivers by integrating vehicles, sensors, and mobile devices into a global network. To serve VANET with computational resources, Vehicular Cloud Computing (VCC) is recently envisioned with the objective of providing traffic solutions to improve our daily driving. These solutions involve applications and services for the benefit of Intelligent Transportation Systems (ITS), which represent an important part of IoV. Data collection is an important aspect in ITS, which can effectively serve online travel systems with the aid of Vehicular Cloud (VC). In this paper, we involve the new paradigm of VCC to propose a data collection model for the benefit of ITS. We show via simulation results that the participation of low percentage of vehicles in a dynamic VC is sufficient to provide meaningful data collection.

53.E-School Techno Bus Based On Cloud Computing Services

Abstract—

This paper surveys the emerging paradigm of cloud mobile media. We start with two alternative perspectives for cloud mobile media networks: an end-to-end view and a layered view. Summaries of existing research in this area are organized according to the layered service framework: i) cloud resource management and control in infrastructure-as-a-service (IaaS), ii) cloud-based media services in platform-as-a-service (PaaS), and iii) novel cloud-based systems and applications in software-as-a-service (SaaS). We further substantiate our proposed design principles for cloud-based mobile media using a concrete case study: a cloud-centric media platform (CCMP) developed at Nanyang Technological University. Finally, this paper concludes with an outlook of open research problems for realizing the vision of cloud-based mobile media.

54. Practical Privacy-Preserving Map reduce Based K-Means Clustering Over Large-Scale Dataset

Abstract—

Clustering techniques have been widely adopted in many real world data analysis applications, such as customer behavior analysis, targeted marketing, digital forensics, etc. With the explosion of data in today's big data era, a major trend to handle a clustering over large-scale datasets is outsourcing it to public cloud platforms. This is because cloud computing offers not only reliable services with performance guarantees, but also savings on in-house IT infrastructures. However, as datasets used for clustering may contain sensitive information, e.g., patient health information, commercial data, and behavioral data, etc, directly outsourcing them to public cloud servers inevitably raise privacy concerns. In this paper, we propose a practical privacy-preserving Kmeans clustering scheme that can be efficiently outsourced to cloud servers. Our scheme allows cloud servers to perform clustering directly over encrypted datasets, while achieving comparable computational complexity and accuracy compared with clusterings over unencrypted ones. We also investigate secure integration of MapReduce into our scheme, which makes our scheme extremely suitable for cloud computing environment. Thorough security analysis and numerical analysis carry out the performance of our scheme in terms of security and efficiency. Experimental evaluation over a 5 million objects dataset further validates the practical performance of our scheme.

55. Fine-Grained Two-Factor Access Control For Web-Based Cloud Computing Services

ABSTRACT—

In this paper, we introduce a new fine-grained two-factor authentication (2FA) access control system for web-based cloud computing services. Specifically, in our proposed 2FA access control system, an attribute-based access control mechanism is implemented with the necessity of both user secret key and a lightweight security device. As a user cannot access the system if s/he does not hold both, the mechanism can enhance the security of the system, especially in those scenarios where many users share the same computer for web-based cloud services. In addition, attribute-based control in the system also enables the cloud server to restrict the access to those users with the same set of attributes while preserving user privacy, i.e., the cloud server only knows that the user fulfils the required predicate, but has no idea on the exact identity of the user. Finally, we also carry out a simulation to demonstrate the practicability of our proposed 2FA system.

56.A Secure And Dynamic Multi-Keyword Ranked Search Scheme Over Encrypted Cloud Data

ABSTRACT –

The benefit of storage as a service several enterprises area unit moving their valuable information to the cloud, since it prices less, simply scalable and can be accessed from anyplace any time. However, delicate information like email, personal health record, Gov. record ought to be encoded before outsourcing for security prerequisites, which obsoletes information utilization like keyword based document retrieval. During this paper, we present Multiuser Multi-keyword Ranked Search scheme over Encrypted Cloud using MHR Tree and KP-ABE. The MHR tree (Markle hash tree) algorithm can achieve logarithmic search time and deal with the deletion and insertion of documents flexibly. Keyword Policy-Attribute based encryption (KP-ABE) is secure encryption technique, it provide the fine grained access control and high security on document collection.

57.An Efficient Privacy-Preserving Ranked Keyword Search Method

Abstract—

Cloud data owners prefer to outsource documents in an encrypted form for the purpose of privacy preserving. Therefore it is essential to develop efficient and reliable ciphertext search techniques. One challenge is that the relationship between documents will be normally concealed in the process of encryption, which will lead to significant search accuracy performance degradation. Also the volume of data in data centers has experienced a dramatic growth. This will make it even more challenging to design ciphertext search schemes that can provide efficient and reliable online information retrieval on large volume of encrypted data. In this paper, a hierarchical clustering method is proposed to support more search semantics and also to meet the demand for fast ciphertext search within a big data environment. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed in this paper. Experiments have been conducted using the collection set built from the IEEE Xplore. The results show that with a sharp increase of documents in the dataset the search time of the proposed method increases linearly whereas the search time of the traditional method increases exponentially. Furthermore, the proposed method has an advantage over the traditional method in the rank privacy and relevance of retrieved documents.

58.Differentially Private Online Learning For Cloud-Based Video Recommendation With Multimedia Big Data In Social Networks

Abstract—

With the rapid growth in multimedia services and the enormous offers of video contents in online social networks, users have difficulty in obtaining their interests. Therefore, various personalized recommendation systems have been proposed. However, they ignore that the accelerated proliferation of social media data has led to the big data era, which has greatly impeded the process of video recommendation. In addition, none of them has considered both the privacy of users' contexts (e,g., social status, ages and hobbies) and video service vendors' repositories, which are extremely sensitive and of significant commercial value. To handle the problems, we propose a cloud-assisted differentially private video recommendation system based on distributed online learning. In our framework, service vendors are modeled as distributed cooperative learners, recommending videos according to user's context, while simultaneously adapting the video-selection strategy based on user-click feedback to maximize total user clicks (reward). Considering the sparsity and heterogeneity of big social media data, we also propose a novel geometric differentially private model, which can greatly reduce the performance (recommendation accuracy) loss. Ou r simulation shows the proposed algorithms outperform other existing methods and keep a delicate balance between computing accuracy and privacy preserving level.

59. Secure optimization computation out sourcing in cloud computing: A case Study Of Linear Programming

ABSTRACT—

How to protect the data that is processed and generated by the customers, is becoming the major concern in the present day situation. Various engineering, computing and optimization techniques are being used to solve this problem. The investigation has been performed for secure outsourcing of problem for the large-scale systems. In this paper, the essential terms involved in the cloud security has been presented. Whereas, the privacy cheating discouragement ”Seclude”, is used for achieving the greater aspects of security. Although the cloud computing is being used to outsource large-scale computations to the cloud, data privacy has become a major issue. In this paper, the modern cryptographic techniques in secure outsourcing along with the research work, which has been proposed in past years, has been presented. Based on some drawback measures, the identification of the problem in the current scenario has been done. This paper also discusses about the motivation towards the problem and our future research directions.

60. Dual-Server Public-Key Encryption With Keyword Search For Secure Cloud Storage.

Abstract –

Searchable encryption is of increasing interest for protecting the data privacy in secure searchable cloud storage. In this paper, we investigate the security of a wellknown cryptographic primitive, namely, public key encryption with keyword search (PEKS) which is very useful in many applications of cloud storage. Unfortunately, it has been shown that the traditional PEKS framework suffers from an inherent insecurity called inside keyword guessing attack (KGA) launched by the malicious server. To address this security vulnerability, we propose a new PEKS framework named dual-server PEKS (DS-PEKS). As another main contribution, we define a new variant of the smooth projective hash functions (SPHFs) referred to as linear and homomorphic SPHF (LH-SPHF). We then show a generic construction of secure DS-PEKS from LH-SPHF. To illustrate the feasibility of our new framework, we provide an efficient instantiation of the general framework from a Decision Diffie–Hellman-based LH-SPHF and show that it can achieve the strong security against inside the KGA.

61. Deduplicatable Dynamic Proof Of Storage For Multi-User Environments

Abstract—

Dynamic Proof of Storage (PoS) is a useful cryptographic primitive that enables a user to check the integrity of outsourced files and to efficiently update the files in a cloud server. Although researchers have proposed many dynamic PoS schemes in singleuser environments, the problem in multi-user environments has not been investigated sufficiently. A practical multi-user cloud storage system needs the secure client-side cross-user deduplication technique, which allows a user to skip the uploading process and obtain the ownership of the files immediately, when other owners of the same files have uploaded them to the cloud server. To the best of our knowledge, none of the existing dynamic PoSs can support this technique. In this paper, we introduce the concept of deduplicatable dynamic proof of storage and propose an efficient construction called DeyPoS, to achieve dynamic PoS and secure cross-user deduplication, simultaneously. Considering the challenges of structure diversity and private tag generation, we exploit a novel tool called Homomorphic Authenticated Tree (HAT). We prove the security of our construction, and the theoretical analysis and experimental results show that our construction is efficient in practice.

62. Division and replication of data incloud for optimal performance And Security

BSTRACT:

The issues related to security and operation overcomes by the Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS). In cloud computing, the information is stored on third party space which results in a security concerns. The user and lymph gland within cloud may compromise the data. Therefore, to protect data within the cloud senior high school security measures are required. Divide a data file into fragments, and replicate the fragmented data over the cloud lymph gland is done in DROPS methodological analysis. Only a fragment of a particular data file can be stored by each of the node that ensures that no meaningful information is revealed to the aggressor even in case of a successful attack.

63.A Map reduce Implementation With New enhancements

ABSTRACT –

A prominent parallel data processing tool Map Reduce is gaining significant momentum from both industry and academia as the volume of data to analyze grows rapidly. While Map Reduce is used in many areas where massive data analysis is required, there are still debates on its performance, efficiency per node, and simple abstraction. This survey intends to assist the database and open source communities in understanding various technical aspects of the Map Reduce framework. In this survey, we characterize the MapReduce framework and discuss its inherent pros and cons. We then introduce its optimization strategies reported in the recent literature. We also discuss the open issues and challenges raised on parallel data analysis with Map Reduce.

64. Point-Of-Interest Recommendation For Location Promotion In Location-Based Social Networks

Abstract –

With the rapid development of location-based social networks (LBSNs), point of interests (POI) recommendation has become an important means to help people discover attractive and interesting locations from billions of locations globally. However, this recommendation is very challenging compared to the traditional recommender systems. A user may visit only a limited number of POIs, leading to a very sparse user-item matrix. This matrix becomes even sparser when the user travels to a distant place as most of the items visited by a user are usually located within a short distance from the user's home. Moreover, user interests and behavior patterns may vary dramatically across different time and different geographical regions. On the other hand, in reality, human movement exhibits sequential patterns. Thus, how to predict users' next move based on her previous visited locations is important and challenging in LBSNs. Our project focuses on offering a more accurate and efficient recommender system by overcoming the aforementioned challenges, and it contains the following three parts: In the first part, we design ST-SAGE, a spatial-temporal sparse additive generative model for POI recommendation. ST-SAGE considers both personal interests of the users and the preferences of the crowd in the target region at the given time by exploiting both the co-occurrence patterns of POIs and the content of POIs. To further alleviate the data sparsity issue, ST-SAGE exploits the geographical correlation by smoothing the crowd's preferences over a well-designed spatial index structure called spatial pyramid. To speed up the training process of ST-SAGE, we implement a parallel version of the model inference algorithm on the GraphLab framework.

65. Social q &A: An Online Social Network Based Question And Answer System

Abstract—

Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s)he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented SocialQ&A, an online social network based Q&A system. SocialQ&A leverages the social network properties of common-interest and mutual-trust friend relationship to identify an asker through friendship who are most likely to answer the question, and enhance the user security. We also improve SocialQ&A with security and efficiency enhancements by protecting user privacy and identifies, and retrieving answers automatically for recurrent questions. We describe the architecture and algorithms, and conducted comprehensive large-scale simulation to evaluate SocialQ&A in comparison with other methods. Our results suggest that social networks can be leveraged to improve the answer quality and asker's waiting time. We also implemented a real prototype of SocialQ&A, and analyse the Q&A behaviour of real users and questions from a small-scale real-world SocialQ&A system.

66. Modelling Urban Behaviour By Mining Geotagged Social Data

Abstract—

Data generated on location-based social networks provide rich information on the whereabouts of urban dwellers. Specifically, such data reveal who spends time where, when, and on what type of activity (e.g., shopping at a mall, or dining at a restaurant). That information can, in turn, be used to describe city regions in terms of activity that takes place therein. For example, the data might reveal that citizens visit one region mainly for shopping in the morning, while another for dining in the evening. Furthermore, once such a description is available, one can ask more elaborate questions. For example, one might ask what features distinguish one region from another – some regions might be different in terms of the type of venues they host and others in terms of the visitors they attract. As another example, one might ask which regions are similar across cities. In this paper, we present a method to answer such questions using publicly shared Foursquare data. Our analysis makes use of a probabilistic model, the features of which include the exact location of activity, the users who participate in the activity, as well as the time of the day and day of week the activity takes place. Compared to previous approaches to similar tasks, our probabilistic modeling approach allows us to make minimal assumptions about the data – which relieves us from having to set arbitrary parameters in our analysis (e.g., regarding the granularity of discovered regions or the importance of different features). We demonstrate how the model learned with our method can be used to identify the most likely and distinctive features of a geographical area, quantify the importance features used in the model, and discover similar regions across different cities. Finally, we perform an empirical comparison with previous work and discuss insights obtained through our findings.

67.A Workflow Management System For Scalable Data Mining On Clouds

Abstract—

The extraction of useful information from data is often a complex process that can be conveniently modeled as a data analysis workflow. When very large data sets must be analyzed and/or complex data mining algorithms must be executed, data analysis workflows may take very long times to complete their execution. Therefore, efficient systems are required for the scalable execution of data analysis workflows, by exploiting the computing services of the Cloud platforms where data is increasingly being stored. The objective of the paper is to demonstrate how Cloud software technologies can be integrated to implement an effective environment for designing and executing scalable data analysis workflows. We describe the design and implementation of the Data Mining Cloud Framework (DMCF), a data analysis system that integrates a visual workflow language and a parallel runtime with the Software-as-a-Service (SaaS) model. DMCF was designed taking into account the needs of real data mining applications, with the goal of simplifying the development of data mining applications compared to generic workflow management systems that are not specifically designed for this domain. The result is a high-level environment that, through an integrated visual workflow language, minimizes the programming effort, making easier to domain experts the use of common patterns specifically designed for the development and the parallel execution of data mining applications. The DMCF's visual workflow language, system architecture and runtime mechanisms are presented. We also discuss several data mining workflows developed with DMCF and the scalability obtained executing such workflows on a public Cloud.

68. Fidoop-Dp: Data Partitioning In Frequent Itemset Mining On Hadoop Clusters

Abstract:

Data partitioning improves the convenience of data utilization. It is complicated to extract information from a collection of data in a fast manner, so this research proposed a hierarchical clustering based improved data partitioning approach. General cloud architectures support and enforce partitioning to offer fast search results. The purpose of this study is to build up a partitioning method based on the amalgamation of cosine and soft cosine similarities to improve the data partitioning performance with better portioning speed and accuracy. Threshold-based partitioning methods are considered to have vertical and horizontal partitioning, where the basis is cosine and soft cosine similarity. The main focus of this research is to develop a hybrid approach for data partitioning in vertical as well as in horizontal manner. To assess the proposed work, Quality of Service (QoS) parameters such as accuracy, recall, true positive, false positive, true negative, and false negative are calculated and compared to data partitioning using a cosine-like algorithm. A comparison has been drawn between H. Guo et al. and K. Korjus et al. with the proposed work. 2.28% enhancement of recall and 39.94 % of enhancement of precision has been noticed with H. Guo etc.

69. A Subword-Based Deep Learning Approach For Sentiment Analysis Of Political Tweets

Abstract –

Deep learning has emerged as a powerful machine learning technique that learns multiple layers of representations or features of the data and produces state-of-the-art prediction results. Along with the success of deep learning in many other application domains, deep learning is also popularly used in sentiment analysis in recent years. This paper first gives an overview of deep learning and then provides a comprehensive survey of its current applications in sentiment analysis.

70.Inverted Linear Quadtree: Efficient Top K Spatial Keyword Search

Abstract—

With advances in geo-positioning technologies and geo-location services, there are a rapidly growing amount of spatio-textual objects collected in many applications such as location based services and social networks, in which an object is described by its spatial location and a set of keywords (terms). Consequently, the study of spatial keyword search which explores both location and textual description of the objects has attracted great attention from the commercial organizations and research communities. In the paper, we study the problem of top k spatial keyword search (TOPK-SK), which is fundamental in the spatial keyword queries. Given a set of spatio-textual objects, a query location and a set of query keywords, the top k spatial keyword search retrieves the closest k objects each of which contains all keywords in the query. Based on the inverted index and the linear quadtree, we propose a novel index structure, called inverted linear quadtree (IL-Quadtree), which is carefully designed to exploit both spatial and keyword based pruning techniques to effectively reduce the search space. An efficient algorithm is then developed to tackle top k spatial keyword search. In addition, we show that the IL-Quadtree technique can also be applied to improve the performance of other spatial keyword queries such as the direction-aware top k spatial keyword search and the spatiotextual ranking query. Comprehensive experiments on real and synthetic data clearly demonstrate the efficiency of our methods.

71.Truth Discovery In Crowdsourced Detection Of Spatial Events

Abstract—

The ubiquity of smartphones has led to the emergence of mobile crowdsourcing tasks such as the detection of spatial events when smartphone users move around in their daily lives. However, the credibility of those detected events can be negatively impacted by unreliable participants with low-quality data. Consequently, a major challenge in mobile crowdsourcing is truth discovery, i.e., to discover true events from diverse and noisy participants' reports. This problem is uniquely distinct from its online counterpart in that it involves uncertainties in both participants' mobility and reliability. Decoupling these two types of uncertainties through location tracking will raise severe privacy and energy issues, whereas simply ignoring missing reports or treating them as negative reports will significantly degrade the accuracy of truth discovery. In this paper, we propose two new unsupervised models, i.e., Truth finder for Spatial Events (TSE) and Personalized Truth finder for Spatial Events (PTSE), to tackle this problem. In TSE, we model location popularity, location visit indicators, truths of events, and three-way participant reliability in a unified framework. In PTSE, we further model personal location visit tendencies. These proposed models are capable of effectively handling various types of uncertainties and automatically discovering truths without any supervision or location tracking. Experimental results on both real-world and synthetic datasets demonstrate that our proposed models outperform existing state-of-the-art truth discovery approaches in the mobile crowdsourcing environment.

72.A Sequential Personalized Spatial Item Recommender System

Abstract—

With the rapid development of location-based social networks (LBSNs), spatial item recommendation has become an important way of helping users discover interesting locations to increase their engagement with location-based services. Although human movement exhibits sequential patterns in LBSNs, most current studies on spatial item recommendations do not consider the sequential influence of locations. Leveraging sequential patterns in spatial item recommendation is, however, very challenging, considering 1) users' check-in data in LBSNs has a low sampling rate in both space and time, which renders existing prediction techniques on GPS trajectories ineffective; 2) the prediction space is extremely large, with millions of distinct locations as the next prediction target, which impedes the application of classical Markov chain models; and 3) there is no existing framework that unifies users' personal interests and the sequential influence in a principled manner. In light of the above challenges, we propose a sequential personalized spatial item recommendation framework (SPORE) which introduces a novel latent variable topic-region to model and fuse sequential influence with personal interests in the latent and exponential space. The advantages of modeling the sequential effect at the topic-region level include a significantly reduced prediction space, an effective alleviation of data sparsity and a direct expression of the semantic meaning of users' spatial activities. Furthermore, we design an asymmetric Locality Sensitive Hashing (ALSH) technique to speed up the online top-k recommendation process by extending the traditional LSH. We evaluate the performance of SPORE on two real datasets and one large-scale synthetic dataset. The results demonstrate a significant improvement in SPORE's ability to recommend spatial items, in terms of both effectiveness and efficiency, compared with the state-of-the-art methods.

73.Detecting Malicious Facebook Applications

Abstract—

With 20 million installs a day [1], third-party apps are a major reason for the popularity and addictiveness of Facebook. Unfortunately, hackers have realized the potential of using apps for spreading malware and spam. The problem is already significant, as we find that at least 13% of apps in our dataset are malicious. So far, the research community has focused on detecting malicious posts and campaigns. In this paper, we ask the question: Given a Facebook application, can we determine if it is malicious? Our key contribution is in developing FRAppE—Facebook's Rigorous Application Evaluator—arguably the first tool focused on detecting malicious apps on Facebook. To develop FRAppE, we use information gathered by observing the posting behavior of 111K Facebook apps seen across 2.2 million users on Facebook. First, we identify a set of features that help us distinguish malicious apps from benign ones. For example, we find that malicious apps often share names with other apps, and they typically request fewer permissions than benign apps. Second, leveraging these distinguishing features, we show that FRAppE can detect malicious apps with 99.5% accuracy, with no false positives and a high true positive rate (95.9%). Finally, we explore the ecosystem of malicious Facebook apps and identify mechanisms that these apps use to propagate. Interestingly, we find that many apps collude and support each other; in our dataset, we find 1584 apps enabling the viral propagation of 3723 other apps through their posts. Long term, we see FRAppE as a step toward creating an independent watchdog for app assessment and ranking, so as to warn Facebook users before installing apps.

74.A Novel Recommendation Model Regularized With User Trust And Item Ratings

Abstract—

We propose Trust SVD, a trust-based matrix factorization technique for recommendations. Trust SVD integrates multiple information sources into the recommendation model in order to reduce the data sparsity and cold start problems and their degradation of recommendation performance. An analysis of social trust data from four real-world data sets suggests that not only the explicit but also the implicit influence of both ratings and trust should be taken into consideration in a recommendation model. Trust SVD therefore builds on top of a state-of-the-art recommendation algorithm, SVD++ (which uses the explicit and implicit influence of rated items), by further incorporating both the explicit and implicit influence of trusted and trusting users on the prediction of items for an active user. The proposed technique is the first to extend SVD++ with social trust information. Experimental results on the four data sets demonstrate that Trust SVD achieves better accuracy than other ten counterparts recommendation techniques.

75.Automatically Mining Facets For Queries From Their Search Results

Abstract –

Web search engines have stored in their logs information about users since they started to operate. This information often serves many purposes. The primary focus of this survey is on introducing to the discipline of query mining by showing its foundations and by analyzing the basic algorithms and techniques that are used to extract useful knowledge from this (potentially) infinite source of information. We show how search applications may benefit from this kind of analysis by analyzing popular applications of query log mining and their influence on user experience. We conclude the paper by, briefly, presenting some of the most challenging current open problems in this field.

76. Cross-Domain Sentiment Classification Using Sentiment Sensitive Embeddings

Abstract—

Unsupervised Cross-domain Sentiment Classification is the task of adapting a sentiment classifier trained on a particular domain (source domain), to a different domain (target domain), without requiring any labeled data for the target domain. By adapting an existing sentiment classifier to previously unseen target domains, we can avoid the cost for manual data annotation for the target domain. We model this problem as embedding learning, and construct three objective functions that capture: (a) distributional properties of pivots (i.e. common features that appear in both source and target domains), (b) label constraints in the source domain documents, and, (c) geometric properties in the unlabeled documents in both source and target domains. Unlike prior proposals that first learn a lower-dimensional embedding independent of the source domain sentiment labels, and next a sentiment classifier in this embedding, our joint optimisation method learns embeddings that are sensitive to sentiment classification. Experimental results on a benchmark dataset show that by jointly optimising the three objectives we can obtain better performances in comparison to optimising each objective function separately, thereby demonstrating the importance of task-specific embedding learning for cross-domain sentiment classification. Among the individual objective functions, the best performance is obtained by (c). Moreover, the proposed method reports cross-domain sentiment classification accuracies that are statistically comparable to the current state-of-the-art embedding learning methods for cross domain sentiment classification.

77.Connecting Social Media To E-Commerce: Cold-Start Product Recommendation Using Microblogging Information

Abstract—

In recent years, the boundaries between ecommerce and social networking have become increasingly blurred. Many e-commerce websites support the mechanism of social login where users can sign on the websites using their social network identities such as their Facebook or Twitter accounts. Users can also post their newly purchased products on microblogs with links to the e-commerce product web pages. In this paper we propose a novel solution for cross-site coldstart product recommendation, which aims to recommend products from e-commerce websites to users at social networking sites in "cold start" situations, a problem which has rarely been explored before. A major challenge is how to leverage knowledge extracted from social networking sites for cross-site cold-start product recommendation. We propose to use the linked users across social networking sites and e-commerce websites (users who have social networking accounts and have made purchases on e-commerce websites) as a bridge to map users' social networking features to another feature representation for product recommendation. In specific, we propose learning both users' and products' feature representations (called user embeddings and product embeddings, respectively) from data collected from ecommerce websites using recurrent neural networks and then apply a modified gradient boosting trees method to transform users' social networking features into user embeddings. We then develop a feature-based matrix factorization approach which can leverage the learnt user embeddings for cold-start product recommendation. Experimental results on a large dataset constructed from the largest Chinese microblogging service SI N A WE I B O and the largest Chinese B2C e-commerce website JI N G DO N G have shown the effectiveness of our proposed framework.

78. Predict The Diagnosis Of Heart Disease Patients Using Classification Mining Techniques

Abstract—

Associative classification is a recent and rewarding technique which integrates association rule mining and classification to a model for prediction and achieves maximum accuracy. Associative classifiers are especially fit to applications where maximum accuracy is desired to a model for prediction. There are many domains such as medical where the maximum accuracy of the model is desired. Heart disease is a single largest cause of death in developed countries and one of the main contributors to disease burden in developing countries. Mortality data from the registrar general of India shows that heart disease are a major cause of death in India, and in Andhra Pradesh coronary heart disease cause about 30%of deaths in rural areas. Hence there is a need to develop a decision support system for predicting heart disease of a patient. In this paper we propose efficient associative classification algorithm using genetic approach for heart disease prediction. The main motivation for using genetic algorithm in the discovery of high level prediction rules is that the discovered rules are highly comprehensible, having high predictive accuracy and of high interestingness values. Experimental Results show that most of the classifier rules help in the best prediction of heart disease which even helps doctors in their diagnosis decisions.

79. Point-Of-Interest Recommendation For Location Promotion In Location-Based Social Networks

Abstract –

The wide spread use of location based social networks (LBSNs) has enabled the opportunities for better location based services through Point-of-Interest (POI) recommendation. Indeed, the problem of POI recommendation is to provide personalized recommendations of places of interest. Unlike traditional recommendation tasks, POI recommendation is personalized, locationaware, and context depended. In light of this difference, this paper proposes a topic and location aware POI recommender system by exploiting associated textual and context information. Specifically, we first exploit an aggregated latent Dirichlet allocation (LDA) model to learn the interest topics of users and to infer the interest POIs by mining textual information associated with POIs. Then, a Topic and Location-aware probabilistic matrix factorization (TL-PMF) method is proposed for POI recommendation. A unique perspective of TL-PMF is to consider both the extent to which a user interest matches the POI in terms of topic distribution and the word-of-mouth opinions of the POIs. Finally, experiments on real-world LBSNs data show that the proposed recommendation method outperforms state-of-the-art probabilistic latent factor models with a significant margin. Also, we have studied the impact of personalized interest topics and word-of-mouth opinions on POI recommendations.

80.Cyberbullying Detection Based On Semantic-Enhanced Marginalized Denoising Auto-Encoder

Abstract:

As a side effect of increasingly popular social media, cyberbullying has emerged as a serious problem afflicting children, adolescents and young adults. Machine learning techniques make automatic detection of bullying messages in social media possible, and this could help to construct a healthy and safe social media environment. In this meaningful research area, one critical issue is robust and discriminative numerical representation learning of text messages. In this paper, a new representation learning method is introduced to tackle this problem. The method named Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) is developed via semantic extension of the popular deep learning model stacked denoisingautoencoder. The semantic extension consists of semantic dropout noise and sparsity constraints, where the semantic dropout noise is designed based on domain knowledge and the word embedding technique. The proposed method is able to exploit the hidden feature structure of bullying information and learn a robust and discriminative representation of text.

81. Cooperative Query Answer Authentication Scheme Over Anonymous Sensing Data

Abstract—

In cloud service over crowd-sensing data, the data owner (DO) publishes the sensing data through the cloud server, so that the user can obtain the information of interest on demand. But the cloud service providers (CSP) are often untrustworthy. The privacy and security concerns emerge over the authenticity of the query answer and the leakage of the DO identity. To solve these issues, many researchers study the query answer authentication scheme for cloud service system. The traditional technique is providing DO's signature for the published data. But the signature would always reveal DO's identity. To deal with this disadvantage, this paper proposes a cooperative query answer authentication scheme, based on the ring signature, the Merkle hash tree (MHT) and the non-repudiable service protocol. Through the cooperation among the entities in cloud service system, the proposed scheme could not only verify the query answer but also protect the DO's identity. First, it picks up the internal nodes of MHT to sign, as well as the root node. Thus, the verification computation complexity could be significantly reduced from O(log2N) to O(log2N0.5) in the best case. Then it improves an existing ring signature to sign the selected nodes. Furthermore, the proposed scheme employs the non-repudiation protocol during the transmission of query answer and verification object (VO) to protect trading behavior between the CSP and users. The security and performance analysis prove the security and feasibility of the proposed scheme. Extensive experimental results demonstrate its superiority of verification efficiency and communication overhead.

82.Geo mob–A geo location based browser for secured mobile banking

Abstract –

With the increasing personalization of the Web, many websites allow users to create their own personal accounts. This has resulted in Web users often having many accounts on different websites, to which they need to authenticate in order to gain access. Unfortunately, there are several security problems connected to the use and re-use of passwords, the most prevalent authentication method currently in use, including eavesdropping and replay attacks. Several alternative methods have been proposed to address these shortcomings, including the use of hardware authentication devices. However, these more secure authentication methods are often not adapted for mobile Web users who use different devices in different places and in untrusted environments, such as public Wi-Fi networks, to access their accounts. We have designed a method for comparing, evaluating and designing authentication solutions suitable for mobile users and untrusted environments. Our method leverages the fact that mobile users often bring their own cell phones, and also takes into account different levels of security adapted for different services on the Web. Another important trend in the authentication landscape is that an increasing number of websites use third-party authentication. This is a solution where users have an account on a single system, the identity provider, and this one account can then be used with multiple other websites. In addition to requiring fewer passwords, these services can also in some cases implement authentication with higher security than passwords can provide

83. Knowledge-Enhanced Mobile Video Broadcasting (Kmv-Cast) Framework With Cloud Support

Abstract—

The convergence of mobile communications and cloud computing facilitates the cross-layer network design and content-assisted communication. Mobile video broadcasting can benefit from this trend by utilizing joint source-channel coding and strong information correlation in clouds. In this paper, a knowledge-enhanced mobile video broadcasting (KMV-Cast) is proposed. The KMV-Cast is built on a linear video transmission instead of traditional digital video system, and exploits the hierarchical Bayesian model to integrate the correlated information into the video reconstruction at the receiver. The correlated information is distilled to obtain its intrinsic features, and the Bayesian estimation algorithm is used to maximize the video quality. The KMV-Cast system consists of both likelihood broadcasting and prior knowledge broadcasting. The simulation results show that the proposed KMV-Cast scheme outperforms the typical linear video transmission scheme called Softcast, and achieves 8dB more of the peak signal-to-noise ratio (PSNR) gain at low-SNR channels (i.e., -10dB), and 5dB more of PSNR gain at high-SNR channels (i.e., 25dB). Compared to traditional digital video system, the proposed scheme has 7dB more of PSNR gain than JPEG2000+802.11a scheme at 10dB channel SNR.

84. Efficient And Privacy-Preserving Min And K-Th Min Computations In Mobile Sensing Systems

Abstract—

Protecting the privacy of mobile phone user participants is extremely important for mobile phone sensing applications. In this paper, we study how an aggregator can expeditiously compute the minimum value or the k-th minimum value of all users' data without knowing them. We construct two secure protocols using probabilistic coding schemes and a cipher system that allows homomorphic bitwise XOR computations for our problems. Following the standard cryptographic security definition in the semi-honest model, we formally prove our protocols' security. The protocols proposed by us can support time-series data and need not to assume the aggregator is trusted. Moreover, different from existing protocols that are based on secure arithmetic sum computations, our protocols are based on secure bitwise XOR computations, thus are more efficient.

85.A Lightweight Secure Data Sharing Scheme For Mobile Cloud Computing

Abstract –

With the popularity of cloud computing, mobile devices can store/retrieve personal data from anywhere at any time. Consequently, the data security problem in mobile cloud becomes more and more severe and prevents further development of mobile cloud. There are substantial studies that have been conducted to improve the cloud security. However, most of them are not applicable for mobile cloud since mobile devices only have limited computing resources and power. Solutions with low computational overhead are in great need for mobile cloud applications. In this paper, we propose a lightweight data sharing scheme (LDSS) for mobile cloud computing. It adopts CP-ABE, an access control technology used in normal cloud environment, but changes the structure of access control tree to make it suitable for mobile cloud environments. LDSS moves a large portion of the computational intensive access control tree transformation in CP-ABE from mobile devices to external proxy servers. Furthermore, to reduce the user revocation cost, it introduces attribute description fields to implement lazy-revocation, which is a thorny issue in program based CP-ABE systems. The experimental results show that LDSS can effectively reduce the overhead on the mobile device side when users are sharing data in mobile cloud environments.

86.Privacy-Preserving Location-Proximity For Mobile Apps

Abstract—

Location Based Services (LBS) have seen alarming privacy breaches in recent years. While there has been much recent progress by the research community on developing privacy-enhancing mechanisms for LBS, their evaluation has been often focused on the privacy guarantees, while the question of whether these mechanisms can be adopted by practical LBS applications has received limited attention. This paper studies the applicability of Privacy-Preserving Location Proximity (PPLP) protocols in the setting of mobile apps. We categorize popular location social apps and analyze the tradeoffs of privacy and functionality with respect to PPLP enhancements. To investigate the practical performance trade-offs, we present an in-depth case study of an Android application that implements InnerCircle, a state-of-the-art protocol for privacypreserving location proximity. This study indicates that the performance of the privacy-preserving application for coarsegrained precision is comparable to real applications with the same feature set.

87.A User-Oriented Behavior-Based Malware Variants Detection System For Android

Abstract—

Android, the most popular mobile OS, has around 78 % of the mobile market share. Due to its popularity, it attracts many malware attacks. In fact, people have discovered around one million new malware samples per quarter [1], and it was reported [2] that over 98 % of these new malware samples are in fact “derivatives” (or variants) from existing malware families. In this paper, we first show that runtime behaviors of malware’s core functionalities are in fact similar within a malware family. Hence, we propose a framework to combine “runtime behavior” with “static structures” to detect malware variants. We present the design and implementation of MONET, which has a client and a backend server module. The client module is a lightweight, indevice app for behavior monitoring and signature generation, and we realize this using two novel interception techniques. The backend server is responsible for large scale malware detection. We collect 3723 malware samples and top 500 benign apps to carry out extensive experiments of detecting malware variants and defending against malware transformation. Our experiments show that MONET can achieve around 99 % accuracy in detecting malware variants. Furthermore, it can defend against 10 different obfuscation and transformation techniques, while only incurs around 7 % performance overhead and about 3 % battery overhead. More importantly, MONET will automatically alert users with intrusion details so to prevent further malicious behaviors.

88. Efficient Multi-Factor Authenticated Key Exchange Scheme For Mobile Communications

Abstract –

Authenticated key exchange is one of the most important applications in applied cryptography, the user interacts with a server to set up a session key to pre-registered information, authentication factor, like password or biometrics of the user is stored. Singlefactor AKE is widely used in practice. Higher security concerns call for MFAKE schemes, e.g. combining passwords, biometrics and device simultaneously. Casually designed schemes, security is even weakened in the sense that leakage of one authentication factor will defeat the whole MFAKE protocol. An inevitable by-product arises that the usability of the protocol often drops greatly. To summarize, the existing multi-factor protocols did not provide enough security and efficiency simultaneously. It proposes a very efficient MFAKE protocol. It defines the security model and gives the according to security analysis. It also implements our proposed method as textual, graphical, biometric and device password to access the user accounts. The theoretic comparisons and the experimental results show that this scheme achieves both security and usability operating conditions is demonstrated through simulation results using MATLAB/Simulink followed by an experimental validation.

89.My Privacy My Decision: Control Of Photo Sharing On Online Social Network

Abstract –

Photo sharing refers to the transfer or publishing of a user's digital photos online and the website which provides such acquaintances over services such as hosting, uploading, sharing and managing of photos through online system. This function provides the upload and display of images through both websites and applications. The photo sharing term can be set up and managed by individual users for the usage of online photo galleries including photo blogs. It means that other users can view but not essentially download the photos, users being able to select different copy-right options for their photos. Unfortunately, it may reveal users privacy if they are permitted to post, comment, and tag a photo liberally. To address this problem, this project proposes an efficient facial recognition system that can recognize everyone in the photo. Online photo sharing applications have become popular as it provides users various new and innovative alternatives to share photos with a range of people. The photo sharing feature is incorporated in many social networking sites which allow users to post photo for their loving ones, families and friends. For users of social networking sites such as Facebook, this system focuses on the privacy concerns and needs of the users, at the same time explores ideas for privacy protection mechanisms. By considering users current concerns and behaviors, the tool can be designed as per the user's desire which they can adopt and then can be motivated to use.

90.P-Lint: A Permission Smell Detector For Android Applications

Abstract—

Permissions are one of the most fundamental components for protecting an Android user's privacy and security. Unfortunately, developers frequently misuse permissions by requiring too many or too few permissions, or by not adhering to permission best practices. These permission-related issues can negatively impact users in a variety of ways, ranging from creating a poor user experience to severe privacy and security implications. To advance the understanding permission-related issues during the app's development process, we conducted an empirical study of 574 GitHub repositories of open-source Android apps. We analyzed the occurrences of four types of permission-related issues across the lifetime of the apps. Our findings reveal that (i) permission-related issues are a frequent phenomenon in Android apps, (ii) the majority of issues are fixed within a few days after their introduction, (iii) permission-related issues can frequently linger inside an app for an extended period of time, which can be as high as several years, before being fixed, and (iv) both project newcomers and regular contributors exhibit the same behaviour in terms of number of introduced and fixed permission-related issues per commit. Index Terms—Mobile Permissions, Android, Mobile Software Engineering, Software Repository Mining

91.Zapdroid: Managing Infrequently Used Applications On Smartphones

ABSTRACT—

User surveys have shown that a typical user has over a hundred apps on her smartphone, but stops using many of them. We conduct a user study to identify such unused apps, which we call zombies, and show via experiments that zombie apps consume significant resources on a user's smartphone and access her private information. We then design and build ZapDroid, which enables users to detect and silo zombie apps in an effective way to prevent their undesired activities. If and when the user wishes to resume using such an app, Zap Droid restores the app quickly and effectively. Our evaluations show that: Zap Droid saves twice the energy from unwanted zombie app behaviours as compared to apps from the Play Store that kill background unwanted processes, and it effectively prevents zombie apps from using undesired permissions. In addition, Zap Droid is energy efficient, consuming < 4% of the battery per day.

92.Privacy And Secure Medical Data Transmission And Analysis For Wireless Sensing Healthcare System

Abstract—

The convergence of Internet of Things (IoT), cloud computing and wireless body-area networks (WBANs) has greatly promoted the industrialization of e-/m-healthcare (electronic-/mobile-healthcare). However, the further flourishing of e-/m-Healthcare still faces many challenges including information security and privacy preservation. To address these problems, a healthcare system (HES) framework is designed that collects medical data from WBANs, transmits them through an extensive wireless sensor network infrastructure and finally publishes them into wireless personal area networks (WPANs) via a gateway. Furthermore, HES involves the GSRM (Groups of Send-Receive Model) scheme to realize key distribution and secure data transmission, the HEBM (Homomorphic Encryption Based on Matrix) scheme to ensure privacy and an expert system able to analyze the scrambled medical data and feed back the results automatically. Theoretical and experimental evaluations are conducted to demonstrate the security, privacy and improved performance of HES compared with current systems or schemes. Finally, the prototype implementation of HES is explored to verify its feasibility.

93. Sbvlc: Secure Barcode-Based Visible Light Communication For Smartphones

Abstract:

Short range communication technologies are the emerging technology in mobile phone based applications. One of the technologies used in short range communication is Near Field Communication (NFC). These short range communication technologies are used in mobile advertisement, data sharing and contactless payments. For this purposes 2D barcodes are being introduced in mobile applications. Since every front camera enabled mobile phones can able to scan 2D barcodes. In this paper we are going to implement 2D barcodes in mobile phone applications which are very helpful for mobile payments and personal identification. With the help of visual light communication 2D barcodes are displayed on the smart phone screens when we use barcodes. In this paper we designed a color 2D barcode for providing better security features with the help of visible light communication (VLC) between smart phones. In this paper we are going to develop a secure 2D color barcode which secure private information during private data sharing process. Finally our experimental result shows that our proposed system provides better security when compared to normal 2D barcode schemes.

94.Mobile Attendance Using Near Field Communication And One-Time Password

Abstract—

To maintain the attendance records, several government organisations and educational institutions in many countries still depend on the paper-based attendance approach. This approach has presented several disadvantages such as time-consuming and wastage of environmental resources. There is a necessity to change these traditional methods of attendance recording with more efficient ones. Thus, many works have been done in this direction. Moreover, this study aims to analyse the most recent studies on automated attendance systems regarding the timeline. Our critical review has highlighted studies in the existing literature concerning technology, application domain, and main findings. Moreover, shed light on most numerous studies on any of three previous aspects.

95.Stamp: Enabling Privacy-Preserving Location Proofs For Mobile Users

ABSTRACT:

Location-based services are quickly becoming immensely popular. In addition to services based on users' current location, many potential services rely on users' location history, or their spatial-temporal provenance. Malicious users may lie about their spatial-temporal provenance without a carefully designed security system for users to prove their past locations. In this paper, we present the SpatialTemporal provenance Assurance with Mutual Proofs (STAMP) scheme. STAMP is designed for ad-hoc mobile users generating location proofs for each other in a distributed setting. However, it can easily accommodate trusted mobile users and wireless access points. STAMP ensures the integrity and nontransferability of the location proofs and protects users' privacy. A semi-trusted Certification Authority is used to distribute cryptographic keys as well as guard users against collusion by a light-weight entropy-based trust evaluation approach. Our prototype implementation on the Android platform shows that STAMP is low-cost in terms of computational and storage resources. Extensive simulation experiments show that our entropy-based trust model is able to achieve high collusion detection accuracy

96.Android Meets Led Bulbs In Smart-Home Automation.

Abstract:

This paper provides a low cost-effective and flexible home control and monitoring system with the aid of an integrated micro-web server with IP connectivity for access to and control of equipment and devices remotely using Android-based smartphone app. The proposed system does not require a dedicated server PC with respect to similar systems and offers a new communication protocol for monitoring and controlling the home environment with more than just switching functionality. Smart home interfaces and device definitions to ensure interoperability between ZigBee devices from various manufacturers of electrical equipment, meters and Smart Energy enables products to allow manufactured. We introduced the proposed home energy control systems design intelligent services for users and provides, we show their implementation, with smartphone.

97.Green House Monitoring And Control Using Smart Phone

Abstract:

The remote appliances control system based on the Android smart phone GUI is designed on Android Smartphone. A user logs into the smart Android phone interface, and clicks the buttons gently to send message commands from the GUI which will be transmitted to home information center through the GSM network. Then the AVR ATmega processor recognizes the specified command, and controls the home appliance switches in the wireless radio frequency manner to achieve remote control of appliances ultimately. This seminar focuses on the design of Android terminal, the communication between PIC and GSM module, the realization of the wireless module device's driver, the difficulty in supplying the appropriate low-voltage DC for MCU and wireless module just by a single live wire. The users can manipulate appliances anytime, anywhere, letting our houses become more and more automated and intelligent. There are some problems in the PC monitor terminal, such as its great bulk, inconvenience to carry, high cost, limited monitoring range and so on. Therefore, it's a good choice to design a terminal based on phone.

98.Android Based Device Control.

Abstract –

Automation of the surrounding environment of a modern human being allows increasing his work efficiency and comfort. There has been a significant development in the area of an individual's routine tasks and those can be automated. In the present times, we can find most of the people clinging to their mobile phones and smart devices throughout the day. Hence with the help of his companion – a mobile phone, some daily household tasks can be accomplished by personifying the use of the mobile phone. Analyzing the current smart phone market, novice mobile users are opting for Android based phones. It has become a second name for a mobile phone in layman terms. Home Automation System (HAS) has been designed for mobile phones having Android platform to automate an 8 bit Bluetooth interfaced microcontroller which controls a number of home appliances like lights, fans, bulbs and many more using on/off relay. This paper presents the automated approach of controlling the devices in a household that could ease the tasks of using the traditional method of the switch. The most famous and efficient technology for short range wireless communication- Bluetooth is used here to automate the system. The HAS system for Android users is a step towards the ease of the tasks by controlling one to twenty four different appliances in any home environment.

99. Development And Controlling Of 5 In 1 Multipurpose Agricultural Robot Using Smart Phones

Abstract: -

The paper aims on the design, development and the fabrication of the robot which can dig the soil, leveler to close the mud and sprayer to spray water, these whole systems of the robot works with the battery and the solar power. More than 40% of the population in the world chooses agriculture as the primary occupation, in recent years the development of the autonomous vehicles in the agriculture has experienced increased interest. The vehicle is controlled by Relay switch through IR sensor input. The language input allows a user to interact with the robot which is familiar to most of the people. The advantages of these robots are hands-free and fast data input operations. In the field of agricultural autonomous vehicle, a concept is been developed to investigate if multiple small autonomous machine could be more efficient than traditional large tractors and human forces.

100. Development Of Wsn Based Water Level Monitoring And Control System Using Android

Abstract:

As indicated by Human Rights Watch, twenty million individuals in our nation are as yet drinking water defiled with arsenic. The World wellbeing Organization (WHO) has likewise expressed this emergency as "the biggest mass harming of a populace ever". To diminish the water related ailments and avoid water populace, we need to quantify water parameters, for example, ph, turbidity, conductivity, temperature and so on. Conventional approach of water observing requires gathering information from different sources physically. A while later examples will send lab for testing and breaking down. So as to spare time utilization and diminishing manual exertion my testing supplies will be put in any water source. Thus, this model can distinguish contamination remotely and take essential activities. The primary objective of this paper to assemble a Sensor-based Water Quality Monitoring System. Arduino Mega 2560 go about as a base station and information from sensor hubs will be send to it. For the scholastic reason, this paper exhibits a little model of sensor systems comprising of temperature, water level, stream and ph. At that point ph and temperature sensor esteems were sent cloud stage (ARTIK cloud) and showed as a graphical portrayal on a neighborhood PC. In addition, GSM shield (SIM808) is associated with Arduino Mega which thinks about sensor esteems to edge esteems and sends a text-based notification to the operator if the got esteem is above or underneath the edge esteem. The aftereffects of this undertaking are talked about in the outcome area of the paper. We tried three water tests from three diverse water sources, (for example, modern water, faucet water and pool water). Three water tests gathered from three distinctive swimming pools.(Except one example) Ph esteem found in rest of the examples were in typical range (temperature esteem between 26-27'C). Result segment (in page 20) clarifies our venture discoveries in subtleties.