IGeekS Technologies

Bridging Technology.

For: - B. E | B. Tech | M. E | M. Tech | MCA | BCA | Diploma |MS |M. Sc |

IEEE
REAL TIME PROJECTS& TRAINING GUIDE
SOFTWARE & EMBEDDED

Java | Microsoft .NET | android | hadoop | INTERNET of THINGS | embedded | ns-2 NETWORK SIMULATOR | MATLAB | VLSI | ARM Cortex-A15 | php MySQL | python | raspberry pi

FINAL YEAR PROJECTS

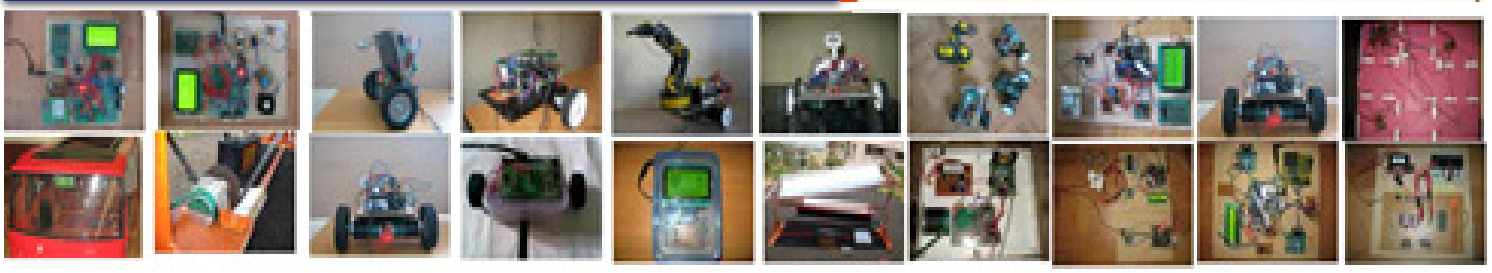
IEEE & Application Projects

BE, Diploma, BSc
M Tech, MCA, BCA

CS | Mechanical
E & C | Electrical

www.makefinalyearproject.com

7019280372/9590544567



JAVA ML, DM, CLOUD, NETWORKING, AI, DS PROJECT ABSTRACTS FOR 2019 – 2020

#19, MN Complex, 2nd Cross, Sampige Main Road, Malleswaram, Bangalore – 560003
Call Us: 9590544567 / 7019280372 **www.makefinalyearproject.com**
www.igeekstechnologies.com **Land Mark: Opposite Joyalukkas Gold Showroom, Near to Mantri Mall**

Title: Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges

Abstract—Prior to the innovation of information communication technologies (ICT), social interactions evolved within small cultural boundaries such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user-generated information, online human networks, and rich human behavior-related data. However, the misuse of social technologies such as social media (SM) platforms, has introduced a new form of aggression and violence that occurs exclusively online. A new means of demonstrating aggressive behavior in SM websites are highlighted in this paper. The motivations for the construction of prediction models to fight aggressive behavior in SM are also outlined. We comprehensively review cyberbullying prediction models and identify the main issues related to the construction of cyberbullying prediction models in SM. This paper provides insights on the overall process for cyberbullying detection and most importantly overviews the methodology. Though data collection and feature engineering process has been elaborated, yet most of the emphasis is on feature selection algorithms and then using various machine learning algorithms for prediction of cyberbullying behaviors.

Title: DCCR: Deep Collaborative Conjunctive Recommender for Rating Prediction

Abstract—Recently, collaborative filtering combined with various kinds of deep learning models is appealing to recommender systems, which have shown a strong positive effect in an accuracy improvement. However, many studies related to deep learning model rely heavily on abundant information to improve prediction accuracy, which has stringent data requirements in addition to raw rating data. Furthermore, most of them ignore the interaction effect between users and items when building the recommendation model. To address these issues, we propose DCCR, a deep collaborative conjunctive recommender, for rating prediction tasks that are solely based on the raw ratings. A DCCR is a hybrid architecture that consists of two different kinds of neural network models (i.e., an autoencoder and a multilayered perceptron). The main function of the autoencoder is to extract the latent features from the perspectives of users and items in parallel, while the multilayered perceptron is used to represent the interaction between users and items based on fusing the user and item latent features. To further improve the performance of DCCR, an advanced activation function is proposed, which can be specified with input vectors. The extensive experiments conducted with two well-known real-world datasets and performances of the DCCR with varying settings are analyzed. The results demonstrate that our DCCR model outperforms other state-

of-art methods. We also discuss the performance of the DCCR with additional layers to show the extensibility of our model.

Title: Authentication by Encrypted Negative Password

Abstract—Secure password storage is a vital aspect in systems based on password authentication, which is still the most widely used authentication technique, despite some security flaws. In this paper, we propose a password authentication framework that is designed for secure password storage and could be easily integrated into existing authentication systems. In our framework, first, the received plain password from a client is hashed through a cryptographic hash function (e.g., SHA-256). Then, the hashed password is converted into a negative password. Finally, the negative password is encrypted into an encrypted negative password (ENP) using a symmetric-key algorithm (e.g., AES), and multi-iteration encryption could be employed to further improve security. The cryptographic hash function and symmetric encryption make it difficult to crack passwords from ENPs. Moreover, there are lots of corresponding ENPs for a given plain password, which makes precomputation attacks (e.g., lookup table attack and rainbow table attack) infeasible. The algorithm complexity analyses and comparisons show that the ENP could resist lookup table attack and provide stronger password protection under dictionary attack. It is worth mentioning that the ENP does not introduce extra elements (e.g., salt); besides this, the ENP could still resist precomputation attacks. Most importantly, the ENP is the first password protection scheme that combines the cryptographic hash function, the negative password, and the symmetric-key algorithm, without the need for additional information except the plain password.

Title: Enabling Efficient and Geometric Range Query with Access Control over Encrypted Spatial Data

Abstract—As a basic query function, range query has been exploited in many scenarios such as SQL retrieves, location-based services, and computational geometry. Meanwhile, with explosive growth of data volume, users are increasingly inclining to store data on the cloud for saving local storage and computational cost. However, a long-standing problem is that the user's data may be completely revealed to the cloud server because it has full data access right. To cope with this problem, a frequently-used method is to encrypt raw data before outsourcing them, but the availability and operability of data will be reduced significantly. In this paper, we propose an efficient and geometric range query scheme (EGRQ) supporting searching and data access control over encrypted spatial data. We employ secure KNN computation, polynomial fitting technique, and order-preserving encryption to achieve secure, efficient, and accurate geometric range query over cloud data. Then, we propose a novel spatial data access control strategy to refine user's rights in our EGRQ. To improve the efficiency, R-tree is adopted to reduce the searching space and matching times in whole search process. Finally, we theoretically prove the security of our proposed scheme in

terms of confidentiality of spatial data, privacy protection of index and trapdoor, and the unlinkability of trapdoors. In addition, extensive experiments demonstrate the high efficiency of our proposed model compared with existing schemes.

Title: Hidden Ciphertext Policy Attribute-Based Encryption with Fast Decryption for Personal Health Record System

Abstract—Since cloud computing has been playing an increasingly important role in real life, the privacy protection in many fields has been paid more and more attention, especially, in the field of personal health record (PHR). The traditional ciphertext-policy attribute-based encryption (CP-ABE) provides the fine-grained access control policy for encrypted PHR data, but the access policy is also sent along with ciphertext explicitly. However, the access policy will reveal the users' privacy, because it contains too much sensitive information of the legitimate data users. Hence, it is important to protect users' privacy by hiding access policies. In most of the previous schemes, although the access policy is hidden, they face two practical problems: 1) these schemes do not support large attribute universe, so their practicality in PHR is greatly limited and 2) the cost of decryption is especially high since the access policy is embedded in the cipher text. To address these problems, we construct a CP-ABE scheme with efficient decryption, where both the size of public parameters and the cost of decryption are constant. Moreover, we also show that the proposed scheme achieves full security in the standard model under static assumptions by using the dual system encryption method.

Title: Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data

Abstract—Road traffic speed prediction is a challenging problem in intelligent transportation system (ITS) and has gained increasing attentions. Existing works are mainly based on raw speed sensing data obtained from infrastructure sensors or probe vehicles, which, however, are limited by expensive cost of sensor deployment and maintenance. With sparse speed observations, traditional methods based only on speed sensing data are insufficient, especially when emergencies like traffic accidents occur. To address the issue, this paper aims to improve the road traffic speed prediction by fusing traditional speed sensing data with new-type "sensing" data from cross domain sources, such as tweet sensors from social media and trajectory sensors from map and traffic service platforms. Jointly modeling information from different datasets brings many challenges, including location uncertainty of low-resolution data, language ambiguity of traffic description in texts, and heterogeneity of cross-domain data. In response to these challenges, we present a unified probabilistic framework, called Topic-Enhanced Gaussian Process Aggregation Model (TEGPAM), consisting of three components, i.e., location disaggregation model, traffic topic model, and traffic speed Gaussian Process model, which integrate new-type data with traditional data. Experiments on real world data from two large cities validate the effectiveness and efficiency of our model.

Title: Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites

Abstract—Online reviews have become an important source of information for users before making an informed purchase decision. Early reviews of a product tend to have a high impact on the subsequent product sales. In this paper, we take the initiative to study the behavior characteristics of early reviewers Through their posted reviews on two real-world large e-commerce platforms, i.e., Amazon and Yelp. In specific, we divide product lifetime into three consecutive stages, namely early, majority and laggards. A user who has posted a review in the early stage is considered as an early reviewer. We quantitatively characterize early reviewers based on their rating behaviors, the helpfulness scores received from others and the correlation of their reviews with product popularity. We have Found that (1) an early reviewer tends to assign a higher average rating score; and (2) an early reviewer tends to post more helpful reviews. Our analysis of product reviews also indicates that early reviewers' ratings and their received Helpfulness scores are likely to influence product popularity. By viewing review posting process as a multiplayer competition game, we propose a novel margin based embedding model for early reviewer prediction. Extensive experiments on two different e-commerce datasets have shown that our proposed approach outperforms a number of competitive baselines.

Title: Title: A Subword-based Deep Learning Approach for Sentiment Analysis of Political Tweets

Abstract—The successful use of online material in political campaigns over the past two decades has motivated the inclusion of social media platforms—such as Twitter—as an integral part of the political apparatus. Political analysts are increasingly turning to Twitter as an indicator of public opinion. We are interested in learning how positive and negative opinions propagate through Twitter and how Important events influence public opinion. In this paper, we present a neural network-based approach to analyse the sentiment expressed on political tweets. First, our approach represents the text by dense vectors comprising subword information to better detect word similarities by exploiting both morphology and semantics. Then, a Convolutional Neural Network is trained to learn how to classify tweets depending on sentiment, based on an available labelled dataset. Finally, the model is applied to perform the sentiment analysis of a collection of tweets retrieved during the days prior to the latest UK General Election. Results are promising and show that the neural network approach represents an improvement over lexicon-based approaches for positive/negative sentence classification.

Title: Corporate Communication Network and Stock Price Movements: Insights from Data Mining

Abstract—Grounded on communication theories, we propose to use a data-mining algorithm to detect communication patterns within a company to determine if such

patterns may reveal the performance of the company. Specifically, we would like to find out whether or not there exist any association relationships between the frequency of e-mail exchange of the key employees in a company and the performance of the company as reflected in its stock prices. If such relationships do exist, we would also like to know whether or not the company's stock price could be accurately predicted based on the detected relationships. To detect the association relationships, a data-mining algorithm is proposed Here to mine e-mail communication records and historical stock prices so that based on the detected relationship, rules that can predict changes in stock prices can be constructed. Using the data-mining algorithm and a set of publicly available Enron e-mail corpus and Enron's stock prices recorded during the same period, we discovered the existence of interesting, statistically significant, association relationships in the data. In addition, we also discovered that these relationships can predict stock price Movements with an average accuracy of around 80%. The results confirm the belief that corporate communication has identifiable patterns and such patterns can reveal meaningful information of Corporate performance as reflected by such indicators as stock market performance. Given the increasing popularity of social networks, the mining of interesting communication patterns could provide insights into the development of many useful applications in many areas.

Title: Traffic Accident Hotspots Identification Based on Clustering Ensemble Model

Abstract—In order to eliminate hidden danger of accident and improve traffic safety, an accident hotspots identification method based on principle component clustering ensemble model was proposed. This method can be used to analyze and quantify safety levels of different roads, to extract principle components and to carry out clustering classification for Comprehensive evaluation function of principle components through Canopy-Kmeans ensemble clustering algorithm so as to extract accident hotspots. The hotspots identification experiment on Anhui section of G50 Hu-Yu highway showed that principle component-clustering analysis method can not only be used to carry out scientific accident statistics analysis and effectively identify accident hotspots, but also to reflect real traffic safety situation, which will help to provide scientific and reasonable basis for improvement of traffic safety decision making performance.

Title: DeepMovRS: A unified framework for deep learning based movie recommender systems

Abstract—A novel unified framework for deep learning-based movie recommender systems, DeepMovRS, is proposed. The proposed framework accepts various heterogeneous inputs from user and item communities' knowledge to explicit and implicit feedbacks. To unify deep architecture of the framework for retrieving and ranking items, it uses suitable machine learning tools to improve the quality of recommendations. The proposed framework is flexible and modular, and it can be

generalized and distributed easily, so it can be a rational choice for commercial movie recommender systems.

Title: Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics

Abstract—Chronic Kidney Disease is a serious lifelong condition that induced by either kidney pathology or reduced kidney functions. Early prediction and proper treatments can possibly stop, or slow the progression of this chronic disease to end-stage, where dialysis or kidney transplantation is the only way to save patient's life. In this study, we examine the ability of several machine-learning methods for early prediction of Chronic Kidney Disease. This matter has been studied widely; however, we are supporting our methodology by the use of predictive analytics, in which we examine the relationship in between data parameters as well as with the target class attribute. Predictive analytics enables us to introduce the optimal subset of parameters to feed machine learning to build a set of predictive models. This study starts with 24 parameters in addition to the class attribute, and ends up by 30% of them as ideal sub set to predict Chronic Kidney Disease. A total of 4 machine learning based classifiers have been evaluated within a supervised learning setting, achieving highest performance outcomes of AUC 0.995, sensitivity 0.9897, and specificity 1. The experimental procedure concludes that advances in machine learning, with assist of predictive analytics, represent a promising setting by which to recognize intelligent solutions, which in turn prove the ability of predication in the kidney disease domain and beyond.

Title: Spammer Detection and Fake User Identification on Social Networks

Abstract—Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter and Facebook have a tremendous impact and occasionally undesirable repercussions for the daily life. The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. Twitter, for example, has become one of the most extravagantly used platforms of all times and therefore allows an unreasonable amount of spamming. Fake users send undesired tweets to users to promote services or websites that not only affect the legitimate users but also disrupt the resource consumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased that results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users on Twitter has become a common area of research in contemporary online social Networks (OSNs). In this paper, we perform a review of techniques used for detecting spammers on Twitter. Moreover, a taxonomy of the Twitter spam detection approaches is presented that classifies the techniques based on their ability to detect: (i) fake content, (ii) spam based on URL, (iii) spam in trending topics, and (iv) fake users. The presented techniques are also compared based on various features, such as user features, content features,

graph features, structure features, and time features. We are hopeful that the presented study will be a useful resource for researchers to find the highlights of recent developments in Twitter spam detection on a single platform.

Title: Detecting Malicious Social Bots Based on Clickstream Sequences

Abstract—With the significant increase in the volume, velocity, and variety of user data (e.g., user-generated data) in online social networks, there have been attempted to design new ways of collecting and analyzing such big data. For example, social bots have been used to perform automated analytical services and provide users with improved quality of service. However, malicious social bots have also been used to disseminate false information (e.g., fake news), and this can result in real-world consequences. Therefore, detecting and removing malicious social bots in online social networks is crucial. The most existing detection methods of malicious social bots analyze the quantitative features of their behavior. These features are easily imitated by social bots; thereby resulting in low accuracy of the analysis. A novel method of detecting malicious social bots, including both features selection based on the transition probability of clickstream sequences and semi-supervised clustering, is presented in this paper. This method not only analyzes transition probability of user behavior clickstreams but also considers the time feature of behavior. Findings from our experiments on real online social network platforms demonstrate that the detection accuracy for different types of malicious social bots by the detection method of malicious social bots based on transition probability of user behavior clickstreams increases by an average of 12.8%, in comparison to the detection method based on quantitative analysis of user behavior.

Title: Active Online Learning for Social Media Analysis to Support Crisis Management

Abstract—People use social media (SM) to describe and discuss different situations they are involved in, like crises. It is therefore worthwhile to exploit SM contents to support crisis management, in particular by revealing useful and unknown information about the crises in real-time. Hence, we propose a novel active online multiple-prototype classifier, called AOMPC. It identifies relevant data related to a crisis. AOMPC is an online learning algorithm that operates on data streams and which is equipped with active learning mechanisms to actively query the label of ambiguous unlabeled data. The number of queries is controlled by a fixed budget strategy. Typically, AOMPC accommodates partly labeled data streams. AOMPC was evaluated using two types of data: (1) synthetic data and (2) SM data from Twitter related to two crises, Colorado Floods and Australia Bushfires. To provide a thorough evaluation, a whole set of known metrics was used to study the quality of the results. Moreover, a sensitivity analysis was conducted to show the effect of AOMPC's parameters on the accuracy of the results. A comparative study of AOMPC against other available online learning algorithms was performed. The experiments

showed very good behavior of AOMPC for dealing with evolving, partly-labeled data streams.

Title: Detecting Pickpocket Suspects from Large-Scale Public Transit Records

Abstract—Massive data collected by automated fare collection (AFC) systems provide opportunities for studying both personal traveling behaviors and collective mobility patterns in urban areas. Existing studies on AFC data have primarily focused on identifying passengers' movement patterns. However, we creatively leveraged such data for identifying pickpocket suspects. Stopping pickpockets in the public transit system has been crucial for improving passenger satisfaction and public safety. Nonetheless, in practice, it is challenging to discern thieves from regular passengers. In this paper, we developed a suspect detection and surveillance system, which can identify pickpocket suspects based on their daily transit records. Specifically, we first extracted a number of useful features from each passenger's daily activities in the transit system. Then, we took a two-step approach that exploits the strengths of unsupervised outlier detection and supervised classification models to identify thieves, who typically exhibit abnormal traveling behaviors. Experimental results demonstrated the effectiveness of our method. We also developed a prototype system for potential uses by security personnel.

Title: A Hybrid E-learning Recommendation Approach Based on Learners' Influence Propagation

Abstract—In e-learning recommender systems, interpersonal information between learners is very scarce, which makes it difficult to apply collaborative filtering (CF) techniques. In this study, we propose a hybrid filtering (HF) recommendation approach (SI-IFL) combining learner influence model (LIM), self-organization based (SOB) recommendation strategy and sequential pattern mining (SPM) together for recommending learning objects (LOs) to learners. The method works as follows: (1) LIM is applied to acquire the interpersonal information by computing the influence that a learner exerts on others. LIM consists of learner similarity, knowledge credibility, and learner aggregation. LIM is independent of ratings. (2) A SOB recommendation strategy is applied to recommend the optimal learner cliques for active learners by simulating the influence propagation among learners. Influence propagation means that a learner can move toward active learners, and such behaviors can stimulate the moving behaviors of his neighbors. This SOB recommendation approach achieves a stable structure based on distributed and bottom-up behaviors of individuals. (3) SPM is applied to decide the final learning objects (LOs) and navigational paths based on the recommended learner cliques. The experimental results demonstrate that SI-IFL can provide personalized and diversified recommendations, and it shows promising efficiency and adaptability in e-learning scenarios.

Serendipitous Recommendation in E-Commerce Using Innovator-Based Collaborative Filtering

Abstract—The rapid development of information technology has facilitated an elegant trading environment in the Internet. There are many trading platforms nowadays but there is no good platform designed for direct consumer-to-consumer (C2C) trading primarily for university students, to buy and sell their goods and services directly to other students within their university or city. Such a need arises in a social network where items should be traded or exchanged easily with a small community. The famous websites such as Amazon or eBay are too global in nature and does not support the direct trading of goods and services among the students in a small social network such as a campus environment.

Title: Accelerating Test Automation through a Domain Specific Language

Abstract—The proposed approach makes use of Accelerating Test Automation Platform (ATAP) which is aimed at making test automation accessible to non-programmers. ATAP allows the creation of an automation test script through a domain specific language based on English. The English-like test scripts are automatically converted to machine executable code using Selenium Web Driver. ATAP's English-like test script makes it easy for non-programmers to author. The functional flow of an ATAP script is easy to understand as well thus making maintenance simpler.

Title: Analyzing Sentiments in One Go: A Supervised Joint Topic Modeling Approach

Abstract—In this work, we focus on modeling user-generated review and overall rating pairs, and aim to identify semantic aspects and aspect-level sentiments from review data as well as to predict overall sentiments of reviews. We propose a novel probabilistic supervised joint aspect and sentiment model (SJASM) to deal with the problems in one go under a unified framework. SJASM represents each review document in the form of opinion pairs, and can simultaneously model aspect terms and corresponding opinion words of the review for hidden aspect and sentiment detection. It also leverages sentimental overall ratings, which often come with online reviews, as supervision data, and can infer the semantic aspects and aspect-level sentiments that are not only meaningful but also predictive of overall sentiments of reviews. Moreover, we also develop efficient inference method for parameter estimation of SJASM based on collapsed Gibbs sampling. We evaluate SJASM extensively on real-world review data, and experimental results

demonstrate that the proposed model outperforms seven well-established baseline methods for sentiment analysis tasks.

Title: Enhanced password processing scheme based on visual cryptography and OCR

Abstract—Traditional password conversion scheme for user authentication is to transform the passwords into hash values. These hash-based password schemes are comparatively simple and fast because those are based on text and famed cryptography. However, those can be exposed to cyber-attacks utilizing password by cracking tool or hash-cracking online sites. Attackers can thoroughly figure out an original password from hash value when that is relatively simple and plain. As a result, many hacking accidents have been happened predominantly in systems adopting those hash-based schemes. In this work, we suggest enhanced password processing scheme based on image using visual cryptography (VC). Different from the traditional scheme based on hash and text, our scheme transforms a user ID of text type to two images encrypted by VC. The user should make two images consisted of sub pixels by random function with SEED which includes personal information. The server only has user's ID and one of the images instead of password. When the user logs in and sends another image, the server can extract ID by utilizing OCR (Optical Character Recognition). As a result, it can authenticate user by comparing extracted ID with the saved one. Our proposal has lower computation, prevents cyber-attack aimed at hash-cracking, and supports authentication not to expose personal information such as ID to attackers.

Title: A Rating Approach based on Sentiment Analysis

Abstract—Sentiment Analysis is the study of analysis of opinions, expressions, likes and dislikes of customers towards various entities like products, services, organizations, individuals etc. With the exponent growth of social media and ecommerce websites like flipkart.com, amazon.com etc. where people can share their experiences about various products through web descriptions, comments or ratings. Product repute is based on its cumulative opinion of the online users. Sentiment analysis or computational analysis of opinion has attracted a great deal of attention due to various potential applications of sentiment analysis in e-commerce domain, online discussion forums and web description sites. Sentiment Analysis is challenging, as it doesn't work well with basic lexical-based classification. This is because the web descriptions are unstructured and are written in natural language.

Title: **A Novel Recommendation Model Regularized with User Trust and Item Ratings**

Abstract—We propose TrustSVD, a trust-based matrix factorization technique for recommendations. TrustSVD 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. TrustSVD 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 TrustSVD achieves better accuracy than other ten counterpart's recommendation techniques.

Title: **Connecting Social Media to E-Commerce: Cold-Start Product Recommendation Using Microblogging Information**

Abstract—In recent years, the boundaries between e-commerce and social networking have become increasingly blurred. Many e-commerce Web sites support the mechanism of social login where users can sign on the Web sites 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 cold-start product recommendation, which aims to recommend products from e-commerce Web sites 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 Web sites (users who have social networking accounts and have made purchases on e-commerce Web sites) 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 e-commerce Web sites 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.

Title: Cross-Platform Identification of Anonymous Identical Users in Multiple Social Media Networks

Abstract—The last few years have witnessed the emergence and evolution of a vibrant research stream on a large variety of online social media network (SMN) platforms. Recognizing anonymous, yet identical users among multiple SMNs is still an intractable problem. Clearly, cross-platform exploration may help solve many problems in social computing in both theory and applications. Since public profiles can be duplicated and easily impersonated by users with different purposes, most current user identification resolutions, which mainly focus on text mining of users' public profiles, are fragile. Some studies have attempted to match users based on the location and timing of user content as well as writing style. However, the locations are sparse in the majority of SMNs, and writing style is difficult to discern from the short sentences of leading SMNs such as Sina Microblog and Twitter. Moreover, since online SMNs are quite symmetric, existing user identification schemes based on network structure are not effective. The real-world friend cycle is highly individual and virtually no two users share a congruent friend cycle. Therefore, it is more accurate to use a friendship structure to analyze cross-platform SMNs. Since identical users tend to set up partial similar friendship structures in different SMNs, we proposed the Friend Relationship-Based User Identification (FRUI) algorithm. FRUI calculates a match degree for all candidate User Matched Pairs (UMPs), and only UMPs with top ranks are considered as identical users. We also developed two propositions to improve the efficiency of the algorithm. Results of extensive experiments demonstrate that FRUI performs much better than current network structure-based algorithms.

Title: 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, we propose a new representation learning method to tackle this problem. Our method named semantic-enhanced marginalized denoising auto-encoder (smSDA) is developed via semantic extension of the popular deep learning model stacked denoising autoencoder (SDA). 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. Our proposed method is able to exploit the hidden feature structure of bullying information and learn a robust and discriminative representation of text. Comprehensive experiments on two public cyberbullying corpora (Twitter and MySpace) are conducted, and the results

show that our proposed approaches outperform other baseline text representation learning methods.

Title: Disease Prediction by Machine Learning over BigData from Healthcare Communities(mongo db)

Abstract—With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In this paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral infarction. We propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.

Title: NetSpam: A Network-Based Spam Detection Framework for Reviews in Online Social Media

Abstract—Nowadays, a big part of people rely on available content in social media in their decisions (e.g., reviews and feedback on a topic or product). The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research, and although a considerable number of studies have been done recently toward this end, but so far the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type. In this paper, we propose a novel framework, named NetSpam, which utilizes spam features for modeling review data sets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features helps us to obtain better results in terms of different metrics experimented on real-world review data sets from Yelp and Amazon Web sites. The results show that NetSpam outperforms the existing methods and among four categories of features, including review-behavioral, user-behavioral, review-linguistic, and user-linguistic, the first type of features performs better than the other categories.

Title: SmartCrawler: A Two-Stage Crawler for Efficiently Harvesting Deep-Web Interfaces

Abstract—As deep web grows at a very fast pace, there has been increased interest in techniques that help efficiently locate deep-web interfaces. However, due to the large volume of web resources and the dynamic nature of deep web, achieving wide coverage and high efficiency is a challenging issue. We propose a two-stage framework, namely SmartCrawler, for efficient harvesting deep web interfaces. In the first stage, SmartCrawler performs site-based searching for center pages with the help of search engines, avoiding visiting a large number of pages. To achieve more accurate results for a focused crawl, SmartCrawler ranks websites to prioritize highly relevant ones for a given topic. In the second stage, SmartCrawler achieves fast in-site searching by excavating most relevant links with an adaptive link-ranking. To eliminate bias on visiting some highly relevant links in hidden web directories, we design a link tree data structure to achieve wider coverage for a website. Our experimental results on a set of representative domains show the agility and accuracy of our proposed crawler framework, which efficiently retrieves deep-web interfaces from large-scale sites and achieves higher harvest rates than other crawlers.

Title: A Shoulder Surfing Resistant Graphical Authentication System

Abstract—Authentication based on passwords is used largely in applications for computer security and privacy. However, human actions such as choosing bad passwords and inputting passwords in an insecure way are regarded as “the weakest link” in the authentication chain. Rather than arbitrary alphanumeric strings, users tend to choose passwords either short or meaningful for easy memorization. With web applications and mobile apps piling up, people can access these applications anytime and anywhere with various devices. This evolution brings great convenience but also increases the probability of exposing passwords to shoulder surfing attacks. Attackers can observe directly or use external recording devices to collect users' credentials. To overcome this problem, we proposed a novel authentication system PassMatrix, based on graphical passwords to resist shoulder surfing attacks. With a one-time valid login indicator and circulative horizontal and vertical bars covering the entire scope of pass-images, PassMatrix offers no hint for attackers to figure out or narrow down the password even they conduct multiple camera-based attacks. We also implemented a PassMatrix prototype on Android and carried out real user experiments to evaluate its memorability and usability. From the experimental result, the proposed system achieves better resistance to shoulder surfing attacks while maintaining usability.

Title: Improving Automated Bug Triaging with Specialized Topic Model

Abstract—Software companies spend over 45 percent of cost in dealing with software bugs. An inevitable step of fixing bugs is bug triage, which aims to

correctly assign a developer to a new bug. To decrease the time cost in manual work, text classification techniques are applied to conduct automatic bug triage. In this paper, we address the problem of data reduction for bug triage, i.e., how to reduce the scale and improve the quality of bug data. We combine instance selection with feature selection to simultaneously reduce data scale on the bug dimension and the word dimension. To determine the order of applying instance selection and feature selection, we extract attributes from historical bug data sets and build a predictive model for a new bug data set.

Title: Search Rank Fraud and Malware Detection in Google Play

Abstract—Fraudulent behaviors in Google Play, the most popular Android app market, fuel search rank abuse and malware proliferation. To identify malware, previous work has focused on app executable and permission analysis. In this paper, we introduce FairPlay, a novel system that discovers and leverages traces left behind by fraudsters, to detect both malware and apps subjected to search rank fraud. FairPlay correlates review activities and uniquely combines detected review relations with linguistic and behavioral signals gleaned from Google Play app data (87K apps, 2.9M reviews, and 2.4M reviewers, collected over half a year), in order to identify suspicious apps. FairPlay achieves over 95% accuracy in classifying gold standard datasets of malware, fraudulent and legitimate apps. We show that 75% of the identified malware apps engage in search rank fraud. FairPlay discovers hundreds of fraudulent apps that currently evade Google Bouncer's detection technology. FairPlay also helped the discovery of more than 1,000 reviews, reported for 193 apps that reveal a new type of "coercive" review campaign users are harassed into writing positive reviews, and install and review other apps.

CLASS Cloud Log Assuring Soundness and Secrecy Scheme for Cloud Forensics

Abstract—User activity logs can be a valuable source of information in cloud forensic investigations; hence, ensuring the reliability and integrity of such logs is crucial. Most existing solutions for secure logging are designed for conventional systems rather than the complexity of a cloud environment. In this paper, we propose the Cloud Log Assuring Soundness and Secrecy (CLASS) process as an alternative scheme for the securing of logs in a cloud environment. In CLASS, logs are encrypted using the individual user's public key so that only the user is able to decrypt the content. In order to prevent unauthorized modification of the log, we generate proof of past log (PPL) using Rabin's fingerprint and Bloom filter. Such an approach reduces verification time significantly. Findings from our experiments deploying CLASS in Open Stack demonstrate the utility of CLASS in a real-world context.

Fast Phrase Search for Encrypted Cloud Storage

Abstract—Cloud computing has generated much interest in the research community in recent years for its many advantages, but has also raise security and privacy concerns. The storage and access of confidential documents have been identified as one of the central problems in the area. In particular, many researchers

investigated solutions to search over encrypted documents stored on remote cloud servers. While many schemes have been proposed to perform conjunctive keyword search, less attention has been noted on more specialized searching techniques. In this paper, we present a phrase search technique based on Bloom filters that is significantly faster than existing solutions, with similar or better storage and communication cost. Our technique uses a series of n-gram filters to support the functionality. The scheme exhibits a trade-off between storage and false positive rate, and is adaptable to defend against inclusion-relation attacks. A design approach based on an application's target false positive rate is also described.

Title: Secure Data Group Sharing and Conditional Dissemination with Multi-Owner in Cloud Computing

Abstract—With the rapid development of cloud services, huge volume of data is shared via cloud computing. Although cryptographic techniques have been utilized to provide data confidentiality in cloud computing, current mechanisms cannot enforce privacy concerns over ciphertext associated with multiple owners, which makes co-owners unable to appropriately control whether data disseminators can actually disseminate their data. In this paper, we propose a secure data group sharing and conditional dissemination scheme with multi-owner in cloud computing, in which data owner can share private data with a group of users via the cloud in a secure way, and data disseminator can disseminate the data to a new group of users if the attributes satisfy the access policies in the ciphertext. We further present a multiparty access control mechanism over the disseminated ciphertext, in which the data co-owners can append new access policies to the ciphertext due to their privacy preferences. Moreover, three policy aggregation strategies, including full permit, owner priority and majority permit, are provided to solve the privacy conflicts problem caused by different access policies.

Title: Enabling Identity-Based Integrity Auditing and Data Sharing With Sensitive Information Hiding for Secure Cloud Storage

Abstract—With cloud storage services, users can remotely store their data to the cloud and realize the data sharing with others. Remote data integrity auditing is proposed to guarantee the integrity of the data stored in the cloud. In some common cloud storage systems such as the electronic health records system, the cloud file might contain some sensitive information. The sensitive information should not be exposed to others when the cloud file is shared. Encrypting the whole shared file can realize the sensitive information hiding, but will make this shared file unable to be used by others. How to realize data sharing with sensitive information hiding in remote data integrity auditing still has not been explored up to now. In order to address this problem, we propose a remote data integrity auditing scheme that realizes data sharing with sensitive information hiding in this paper. In this scheme, a sanitizer is used to sanitize the data blocks corresponding to the sensitive information of the file and transforms these data blocks' signatures into valid ones

for the sanitized file. These signatures are used to verify the integrity of the sanitized file in the phase of integrity auditing. As a result, our scheme makes the file stored in the cloud able to be shared and used by others on the condition that the sensitive information is hidden, while the remote data integrity auditing is still able to be efficiently executed. Meanwhile, the proposed scheme is based on identity-based cryptography, which simplifies the complicated certificate management. The security analysis and the performance evaluation show that the proposed scheme is secure and efficient.

Efficient and Expressive Keyword Search Over Encrypted Data in Cloud

Abstract—In this project Searchable encryption allows a cloud server to conduct keyword search over encrypted data on behalf of the data users without learning the underlying plaintexts. However, most existing searchable encryption schemes only support single or conjunctive keyword search, while a few other schemes that are able to perform expressive keyword search are computationally inefficient since they are built from bilinear pairings over the composite-order groups. In this paper, we propose an expressive public-key searchable encryption scheme in the prime-order groups, which allows keyword search policies (i.e., predicates, access structures) to be expressed in conjunctive, disjunctive or any monotonic Boolean formulas and achieves significant performance improvement over existing schemes. We formally define its security, and prove that it is selectively secure in the standard model. Also, we implement the proposed scheme using a rapid prototyping tool called Charm and conduct several experiments to evaluate it performance. The results demonstrate that our scheme is much more efficient than the ones built over the composite-order groups. Keyword research is one of the most important, valuable, and high return activities in the search marketing field. Ranking for the right keywords can make or break your website. By researching your market's keyword demand, you can not only learn which terms and phrases to target with SEO, but also learn more about your customers as a whole.

Audit-Free Cloud Storage via Deniable Attribute-based Encryption

Abstract—Cloud storage services have become increasingly popular. Because of the importance of privacy, many cloud storage encryption schemes have been proposed to protect data from those who do not have access. All such schemes assumed that cloud storage providers are safe and cannot be hacked; however, in practice, some authorities (i.e., coercers) may force cloud storage providers to reveal user secrets or confidential data on the cloud, thus altogether circumventing storage encryption schemes. In this paper, we present our design for a new cloud storage encryption scheme that enables cloud storage providers to create convincing fake user secrets to protect user privacy. Since coercers cannot tell if obtained secrets are true or not, the cloud storage providers ensure that user privacy is still securely protected. Most of the proposed schemes assume cloud storage service providers or trusted

third parties handling key management are trusted and cannot be hacked; however, in practice, some entities may intercept communications between users and cloud storage providers and then compel storage providers to release user secrets by using government power or other means. In this case, encrypted data are assumed to be known and storage providers are requested to release user secrets.

SEPDP Secure and Efficient Privacy Preserving Provable Data Possession in Cloud Storage

Abstract—Cloud computing is an emergent paradigm to provide reliable and resilient infrastructure enabling the users (data owners) to store their data and the data consumers (users) can access the data from cloud servers. This paradigm reduces storage and maintenance cost of the data owner. At the same time, the data owner loses the physical control and possession of data which leads to many security risks. Therefore, auditing service to check data integrity in the cloud is essential. This issue has become a challenge as the possession of data needs to be verified while maintaining the privacy. To address these issues this work proposes a secure and efficient privacy preserving provable data possession (SEPDP). Further, we extend SEPDP to support multiple owners, data dynamics and batch verification. The most attractive feature of this scheme is that the auditor can verify the possession of data with low computational overhead.

Towards Green Cloud Computing Demand Allocation and Pricing Policies for Cloud Service Brokerage

Abstract—Functioning as an intermediary between tenants and cloud providers, cloud service brokerages (CSBs) can bring about great benefits to the cloud market. As energy costs of cloud computing have been increasing rapidly, there is a need for cloud providers To optimize energy efficiency while maintain high service level performance to tenants, not only for their own benefit but also for social welfares. Thus, for green cloud companies, two questions have arisen 1) under what pricing policies from the cloud providers to the CSB, a profit-driven CSB is willing to minimize the total energy cost while satisfy tenant demands and 2) how should a CSB distribute tenants demands to achieve this objective? To address question 1), we find a pricing policy for cloud providers such that maximizing CSBs profit is equivalent to minimizing cloud providers energy cost. To address question 2), we first devise a greedy solution, and then propose an approximation algorithm and a decomposition-based solution with a constant approximation ratio. Both simulation and real-world Amazon EC2 experimental results demonstrate the effectiveness of our pricing policy to incentivize CSBs to save energy and the superior performance of our algorithms in energy efficiency and resource utilization over the previous algorithms.

SeSPHR A Methodology for Secure Sharing of Personal Health Records in the Cloud

Abstract—The widespread acceptance of cloud based services in healthcare sector has resulted in cost effective and convenient exchange of Personal Health Records (PHRs) among several participating entities of the e-Health systems. Nevertheless, storing confidential health information to cloud servers is susceptible to revelation or theft and calls for the development of methodologies that ensure the privacy of the PHRs. Therefore, we propose a methodology called SeSPHR for secure sharing of the PHRs in the cloud. The SeSPHR scheme ensures patient-centric control on the PHRs and preserves the confidentiality of the PHRs. The patients store the encrypted PHRs on the un-trusted cloud servers and selectively grant access to different types of users on different portions of the PHRs. A semi-trusted proxy called Setup and Re-encryption Server (SRS) is introduced to set up the public/private key pairs and to produce the re-encryption keys. Moreover, the methodology is secure against insider threats and also enforces a forward and backward access control. Furthermore, we formally analyze and verify the working of SeSPHR methodology through High Level Petri Nets (HLPN). Performance evaluation with regard to time consumption indicates that the SeSPHR methodology has potential to be employed for securely sharing the PHRs in the cloud.

Privacy Protection and Intrusion Avoidance for Cloudlet-based Medical Data Sharing

Abstract—Wireless medical sensor networks is a key enabling technology in e-healthcare that allows the data of a patient's vital body parameters to be collected by a wearable or implantable biosensors. The major issue is the security and privacy protection of the collected data because of the resource constraints in the medical sensor network devices. There is a high demand for both security and privacy in practicality. Here we propose a lightweight and secure medical sensor networks. The technologies used in this system are hash-chain based key updating mechanism and proxy protected signature technique. The important feature of hash-chain based key updating mechanism is that for each transmission of data the key is updated. These technologies are helpful to achieve efficient secure transmission and fine-grained data access control. This system also provides the backward secrecy and privacy. This system requires symmetric key encryption/decryption and hash operations .These techniques are suitable for low power sensor nodes. This is the best secure data transmission and access control system for medical sensor networks.

A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud

Abstract—Benefited from cloud computing, users can achieve an effective and economical approach for data sharing among group members in the cloud with the characters of low maintenance and little management cost. Meanwhile, we must provide security guarantees for the sharing data files since they are outsourced.

Unfortunately, because of the frequent change of the membership, sharing data while providing privacy-preserving is still a challenging issue, especially for an untrusted cloud due to the collusion attack. Moreover, for existing schemes, the security of key distribution is based on the secure communication channel, however, to have such channel is a strong assumption and is difficult for practice. In this paper, we propose a secure data sharing scheme for dynamic members. Firstly, we propose a secure way for key distribution without any secure communication channels, and the users can securely obtain their private keys from group manager. Secondly, our scheme can achieve fine-grained access control, any user in the group can use the source in the cloud and revoked users cannot access the cloud again after they are revoked. Thirdly, we can protect the scheme from collusion attack, which means that revoked users cannot get the original data file even if they conspire with the untrusted cloud. In our approach, by leveraging polynomial function, we can achieve a secure user revocation scheme. Finally, our scheme can achieve fine efficiency, which means previous users need not to update their private keys for the situation either a new user joins in the group or a user is revoked from the group.

SecRBAC Secure Data in Clouds

Abstract—Most current security solutions are based on perimeter security. However, Cloud computing breaks the organization perimeters. When data resides in the Cloud, they reside outside the organizational bounds. This leads users to a loss of control over their data and raises reasonable security concerns that slow down the adoption of Cloud computing. Is the Cloud service provider accessing the data? Is it legitimately applying the access control policy defined by the user? This paper presents a data-centric access control solution with enriched role-based expressiveness in which security is focused on protecting user data regardless the Cloud service provider that holds it. Novel identity-based and proxy re-encryption techniques are used to protect the authorization model. Data is encrypted and authorization rules are cryptographically protected to preserve user data against the service provider access or misbehavior. The authorization model provides high expressiveness with role hierarchy and resource hierarchy support. The solution takes advantage of the logic formalism provided by Semantic Web technologies, which enables advanced rule management like semantic conflict detection. A proof of concept implementation has been developed and a working prototypical deployment of the proposal has been integrated within Google services.

Secure Data Sharing in Cloud Computing Using Revocable-Storage Identity-Based Encryption.

Abstract—Cloud computing provides a flexible and convenient way for data sharing, which brings various benefits for both the society and individuals. But there exists a natural resistance for users to directly outsource the shared data to the cloud server since the data often contain valuable information. Thus, it is necessary to place cryptographically enhanced access control on the shared data. Identity-based

encryption is a promising cryptographically primitive to build a practical data sharing system. However, access control is not static. That is, when some user's authorization is expired, there should be a mechanism that can remove him/her from the system. Consequently, the revoked user cannot access both the previously and subsequently shared data. To this end, we propose a notion called revocable-storage identity-based encryption (RS-IBE), which can provide the forward/backward security of ciphertext by introducing the functionalities of user revocation and ciphertext update simultaneously. Furthermore, we present a concrete construction of RS-IBE, and prove its security in the defined security model. The performance comparisons indicate that the proposed RS-IBE scheme has advantages in terms of functionality and efficiency, and thus is feasible for a practical and cost-effective data-sharing system. Finally, we provide implementation results of the proposed scheme to demonstrate its practicability.

A Cloud Environment for Backup and Data Storage

Abstract—Currently derived from advances and technological developments can have Input-Output devices ever better able to store more information. The use of the disks of the nodes of a cluster as global storage system is an inexpensive solution for a cloud environment. The need for the available of information from anywhere is increasing; this represents a problem for many users who use applications such as databases, media, personal file, documents, etc. The I/O data demands of these applications get higher as they get larger. In order to improve performance of these applications can use parallel file systems. PVFS2 is a free parallel file system developed by a multi-institution team of parallel I/O, networking and storage experts. In this paper we present the design of an implementation for cloud environment for able to store and back up data through using remote servers that can be accessed through the Internet. The implementation aims to increase the availability of data and reduce in loss of information.

Secure Auditing and Deduplicating Data in Cloud

Abstract—As the cloud computing technology develops during the last decade, outsourcing data to cloud service for storage becomes an attractive trend, which benefits in sparing efforts on heavy data maintenance and management. Nevertheless, since the outsourced cloud storage is not fully trustworthy, it raises security concerns on how to realize data deduplication in cloud while achieving integrity auditing. In this work, we study the problem of integrity auditing and secure deduplication on cloud data. Specifically, aiming at achieving both data integrity and deduplication in cloud, we propose two secure systems, namely SecCloud and SecCloud+. SecCloud introduces an auditing entity with a maintenance of a MapReduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is designed motivated by the fact that

customers always want to encrypt their data before uploading, and enables integrity auditing and secure deduplication on encrypted data.

Two-Factor Data Security Protection Mechanism for Cloud Storage System

Abstract—In this paper, we propose a two-factor data security protection mechanism with factor revocability for cloud storage system. Our system allows a sender to send an encrypted message to a receiver through a cloud storage server. The sender only needs to know the identity of the receiver but no other information (such as its public key or its certificate). The receiver needs to possess two things in order to decrypt the ciphertext. The first thing is his/her secret key stored in the computer. The second thing is a unique personal security device which connects to the computer. It is impossible to decrypt the ciphertext without either piece. More importantly, once the security device is stolen or lost, this device is revoked. It cannot be used to decrypt any ciphertext. This can be done by the cloud server which will immediately execute some algorithms to change the existing ciphertext to be un-decryptable by this device. This process is completely transparent to the sender. Furthermore, the cloud server cannot decrypt any ciphertext at any time. The security and efficiency analysis show that our system is not only secure but also practical.

Cloud-Assisted Mobile-Access of Health Data With Privacy and Audit ability

Abstract—Motivated by the privacy issues, curbing the adoption of electronic healthcare systems and the wild success of cloud service models, we propose to build privacy into mobile healthcare systems with the help of the private cloud. Our system offers salient features including efficient key management, privacy-preserving data storage, and retrieval, especially for retrieval at emergencies, and audit ability for misusing health data. Specifically, we propose to integrate key management from pseudorandom number generator for unlink ability, a secure indexing method for privacy preserving keyword search which hides both search and access patterns based on redundancy, and integrate the concept of attribute based encryption with threshold signing for providing role-based access control with audit ability to prevent potential misbehavior, in both normal and emergency cases.

TTSA An Effective Scheduling Approach for Delay Bounded Tasks in Hybrid Clouds

Abstract—The economy of scale provided by cloud attracts a growing number of organizations and industrial companies to deploy their applications in cloud data centers (CDCs) and to provide services to users around the world. The uncertainty of arriving tasks makes it a big challenge for private CDC to cost-effectively schedule delay bounded tasks without exceeding their delay bounds. Unlike previous studies, this paper takes into account the cost minimization problem for private CDC in hybrid clouds, where the energy price of private CDC and execution price of public clouds both show the temporal diversity. Then, this paper proposes a temporal task scheduling algorithm (TTSA) to effectively dispatch all arriving

tasks to private CDC and public clouds. In each iteration of TTSA, the cost minimization problem is modeled as a mixed integer linear program and solved by a hybrid simulated-annealing particle-swarm-optimization. The experimental results demonstrate that compared with the existing methods, the optimal or suboptimal scheduling strategy produced by TTSA can efficiently increase the throughput and reduce the cost of private CDC while meeting the delay bounds of all the tasks.

Cloud Computing Security From Single to Multi-Clouds.

Abstract—The use of cloud computing has increased rapidly in many organizations. Cloud computing provides many benefits in terms of low cost and accessibility of data. Ensuring the security of cloud computing is a major factor in the cloud computing environment, as users often store sensitive information with cloud storage providers but these providers may be untrusted. Dealing with "single cloud" providers is predicted to become less popular with customers due to risks of service availability failure and the possibility of malicious insiders in the single cloud. A movement towards "multi-clouds", or in other words, "interclouds" or "cloud-of-clouds" has emerged recently.This paper surveys recent research related to single and multi-cloud security and addresses possible solutions. It is found that the research into the use of multi-cloud providers to maintain security has received less attention from the research community than has the use of single clouds. This work aims to promote the use of multi-clouds due to its ability to reduce security risks that affect the cloud computing user.

Circuit Cipher text-policy Attribute-based Hybrid Encryption with Verifiable Delegation in Cloud Computing

Abstract—Data sharing is an important functionality in cloud storage. In this article, we show how to securely, efficiently, and flexibly share data with others in cloud storage. We describe new public-key cryptosystems which produce constant-size ciphertexts such that efficient delegation of decryption rights for any set of ciphertexts are possible. The novelty is that one can aggregate any set of secret keys and make them as compact as a single key, but encompassing the power of all the keys being aggregated. In other words, the secret key holder can release a constant-size aggregate key for flexible choices of ciphertext set in cloud storage, but the other encrypted files outside the set remain confidential. This compact aggregate key can be conveniently sent to others or be stored in a smart card with very limited secure storage. We provide formal security analysis of our schemes in the standard model. We also describe other application of our schemes. In particular, our schemes give the first public-key patient-controlled encryption for flexible hierarchy, which was yet to be known.

TEES An Efficient Search Scheme over Encrypted Data on Mobile Cloud

- **Abstract— Document storage in the cloud infrastructure is rapidly gaining popularity throughout the world. However, it poses risk to consumers unless the data is encrypted for security. Encrypted data should be effectively searchable and retrievable without any privacy leaks, particularly for the mobile client.**
- **Although recent research has solved many security issues, the architecture cannot be applied on mobile devices directly under the mobile cloud environment. This is due to the challenges imposed by wireless networks, such as latency sensitivity, poor connectivity, and low transmission rates. This leads to a long search time and extra network traffic costs when using traditional search schemes.**

This study addresses these issues by proposing an efficient Encrypted Data Search (TEES) scheme as a mobile cloud service. This innovative scheme uses a lightweight trapdoor (encrypted keyword) compression method, which optimizes the data communication process by reducing the trapdoor's size for traffic efficiency.

SUPERMAN Security Using Pre-Existing Routing for Mobile Ad hoc Networks

Abstract—The flexibility and mobility of Mobile Ad hoc Networks (MANETs) have made them increasing popular in a wide range of use cases. To protect these networks, security protocols have been developed to protect routing and application data. However, these protocols only protect routes or communication, not both. Both secure routing and communication security protocols must be implemented to provide full protection. The use of communication security protocols originally developed for wireline and WiFi networks can also place a heavy burden on the limited network resources of a MANET. To address these issues, a novel secure framework (SUPERMAN) is proposed. The framework is designed to allow existing network and routing protocols to perform their functions, whilst providing node authentication, access control, and communication security mechanisms. This paper presents a novel security framework for MANETs, SUPERMAN. Simulation results comparing SUPERMAN with IPsec, SAODV and SOLSR are provided to demonstrate the proposed frameworks suitability for wireless communication security.

Routing in Accumulative Multi-Hop Networks

Abstract—This paper investigates the problem of finding optimal paths in single-source single-destination accumulative multihop networks. We consider a single source that communicates to a single destination assisted by several relays through multiple hops. At each hop, only one node transmits, while all the other nodes receive the transmitted signal, and store it after processing/decoding and mixing it with the signals received in previous hops. That is, we consider that terminals make use of advanced energy accumulation transmission/reception techniques, such as

maximal ratio combining reception of repetition codes, or information accumulation with rateless codes. Accumulative techniques increase communication reliability, reduce energy consumption, and decrease latency. We investigate the properties that a routing metric must satisfy in these accumulative networks to guarantee that optimal paths can be computed with Dijkstra's algorithm. We model the problem of routing in accumulative multi-hop networks, as the problem of routing in a hypergraph. We show that optimality properties in a traditional multi-hop network (monotonicity and isotonicity) are no longer useful and derive a new set of sufficient conditions for optimality. We illustrate these results by studying the minimum energy routing problem in static accumulative multi-hop networks for different forwarding strategies at relays.

Network Capability in Localizing Node Failures via End-to-End Path Measurements

Abstract—We investigate the capability of localizing node failures in communication networks from binary states (normal/failed) of end-to-end paths. Given a set of nodes of interest, uniquely localizing failures within this set requires that different observable path states associate with different node failure events. However, this condition is difficult to test on large networks due to the need to enumerate all possible node failures. Our first contribution is a set of sufficient/necessary conditions for identifying a bounded number of failures within an arbitrary node set that can be tested in polynomial time. In addition to network topology and locations of monitors, our conditions also incorporate constraints imposed by the probing mechanism used. We consider three probing mechanisms that differ according to whether measurement paths are (i) arbitrarily controllable; (ii) controllable but cycle-free; or (iii) uncontrollable (determined by the default routing protocol). Our second contribution is to quantify the capability of failure localization through 1) the maximum number of failures (anywhere in the network) such that failures within a given node set can be uniquely localized and 2) the largest node set within which failures can be uniquely localized under a given bound on the total number of failures. Both measures in 1) and 2) can be converted into the functions of a per-node property, which can be computed efficiently based on the above sufficient/necessary conditions.

We demonstrate how measures 1) and 2) proposed for quantifying failure localization capability can be used to evaluate the impact of various parameters, including topology, number of monitors, and probing mechanisms.

Mitigating Cross-Site Scripting Attacks with a Content Security Policy

Abstract—A content security policy (CSP) can help Web application developers and server administrator's better control website content and avoid vulnerabilities to cross site scripting (XSS). In experiments with a prototype website, the authors' CSP implementation successfully mitigated all XSS attack types in four popular browsers. Among the many attacks on Web applications, cross site scripting (XSS) is one of the most common. An XSS attack involves injecting malicious script into a trusted website that executes on a visitor's browser without the visitor's knowledge and thereby enables the attacker to access sensitive user data, such as session tokens and

cookies stored on the browser.1 With this data, attackers can execute several malicious acts, including identity theft, key logging, phishing, user impersonation, and webcam activation. Content Security Policy (CSP) is an added layer of security that helps to detect and mitigate certain types of attacks, including Cross Site Scripting (XSS) and data injection attacks. These attacks are used for everything from data theft to site defacement or distribution of malware. CSP is designed to be fully backward compatible; browsers that don't support it still work with servers that implement it, and vice-versa. Browsers that don't support CSP simply ignore it, functioning as usual, defaulting to the standard same-origin policy for web content. If the site doesn't offer the CSP header, browsers likewise use the standard same-origin policy.Enabling CSP is as easy as configuring your web server to return the Content-Security-Policy HTTP header. (Prior to Firefox 23, the X-Content-Security-Policy header was used). See Using Content Security Policy for details on how to configure and enable CSP.

A Hop-by-Hop Routing Mechanism for Green Internet

Abstract—In this paper we study energy conservation in the Internet. We observe that different traffic volumes on a link can result in different energy consumption; this is mainly due to such technologies as trunking (IEEE 802.1AX), adaptive link rates, etc. We design a green Internet routing scheme, where the routing can lead traffic in a way that is green. We differ from previous studies where they switch network components, such as line cards and routers, into sleep mode. We do not prune the Internet topology. We first develop a power model, and validate it using real commercial routers. Instead of developing a centralized optimization algorithm, which requires additional protocols such as MPLS to materialize in the Internet, we choose a hop-by-hop approach. It is thus much easier to integrate our scheme into the current Internet. We progressively develop three algorithms, which are loop-free, substantially reduce energy consumption, and jointly consider green and QoS requirements such as path stretch. We further analyze the power saving ratio, the routing dynamics, and the relationship between hop-by-hop green routing and QoS requirements. We comprehensively evaluate our algorithms through simulations on synthetic, measured, and real topologies, with synthetic and real traffic traces. We show that the power saving in the line cards can be as much as 50 percent.

Secure and Efficient Data Communication Protocol for Wireless Body Area Networks

Abstract—Wireless medical sensor networks is a key enabling technology in e-healthcare that allows the data of a patient's vital body parameters to be collected by a wearable or implantable biosensors. The major issue is the security and privacy protection of the collected data because of the resource constraints in the medical sensor network devices. There is a high demand for both security and privacy in practicality. Here we propose a lightweight and secure medical sensor networks. The technologies used in this system are hash-chain based key updating mechanism and proxy protected signature technique. The important feature of hash-chain based key updating mechanism is that for each transmission of data the key is updated. These technologies are helpful to achieve efficient secure transmission and fine-grained data access control. This system also provides the backward secrecy and privacy. This system requires symmetric key encryption/decryption and hash operations .These techniques are suitable for low power sensor nodes. This is the best secure data transmission and access control system for medical sensor networks.

FastGeo Efficient Geometric Range Queries on Encrypted Spatial Data

Abstract—Spatial data have wide applications, e.g., location-based services, and geometric range queries (i.e., finding points inside geometric areas, e.g., circles or polygons) are one of the fundamental search functions over spatial data. The rising demand of outsourcing data is moving large-scale datasets, including large-scale spatial datasets, to public clouds. Meanwhile, due to the concern of insider attackers and hackers on public clouds, the privacy of spatial datasets should be cautiously preserved while querying them at the server side, especially for location-based and medical usage. In this paper, we formalize the concept of Geometrically Searchable Encryption, and propose an efficient scheme, named FastGeo, to protect the privacy of clients' spatial datasets stored and queried at a public server. With FastGeo, which is a novel two-level search for encrypted spatial data, an honest-but-curious server can efficiently perform geometric range queries, and correctly return data points that are inside a geometric range to a client without learning sensitive data points or this private query. FastGeo supports arbitrary geometric areas, achieves sub linear search time, and enables dynamic Updates over encrypted spatial datasets. Our scheme is provably secure, and our experimental results on real-world spatial datasets in cloud platform demonstrate that FastGeo can boost search time over 100 times.

A Stochastic Model to Investigate Data Center Performance and QoS in IaaS Cloud Computing Systems

Abstract—Cloud data center management is a key problem due to the numerous and heterogeneous strategies that can be applied, ranging from the VM placement to the federation with other clouds. Performance evaluation of Cloud Computing

infrastructures is required to predict and quantify the cost-benefit of a strategy portfolio and the corresponding Quality of Service (QoS) experienced by users. Such analyses are not feasible by simulation or on-the-field experimentation, due to the great number of parameters that have to be investigated. In this paper, we present an analytical model, based on Stochastic Reward Nets (SRNs), that is both scalable to model systems composed of thousands of resources and flexible to represent different policies and cloud-specific strategies. Several performance metrics are defined and evaluated to analyze the behavior of a Cloud data center utilization, availability, waiting time, and responsiveness. A resiliency analysis is also provided to take into account load bursts. Finally, a general approach is presented that, starting from the concept of system capacity, can help system managers to opportunely set the data center parameters under different working conditions.

Data Lineage in Malicious Environments

Abstract—Intentional or unintentional leakage of confidential data is undoubtedly one of the most severe security threats that organizations face in the digital era. The threat now extends to our personal lives a plethora of personal information is available to social networks and smartphone providers and is indirectly transferred to untrustworthy third party and fourth party applications. In this work, we present a generic data lineage framework LIME for data flow across multiple entities that take two characteristic, principal roles (i.e., owner and consumer). We define the exact security guarantees required by such a data lineage mechanism toward identification of a guilty entity, and identify the simplifying non-repudiation and honesty assumptions. We then develop and analyze a novel accountable data transfer protocol between two entities within a malicious environment by building upon oblivious transfer, robust watermarking, and signature primitives. Finally, we perform an experimental evaluation to demonstrate the practicality of our protocol and apply our framework to the important data leakage scenarios of data outsourcing and social networks. In general, we consider LIME , our lineage framework for data transfer, to be an key step towards achieving accountability by design.

Packet-Hiding Methods for Preventing Selective Jamming Attacks

Abstract—The open nature of the wireless medium leaves it vulnerable to intentional interference attacks, typically referred to as jamming. This intentional interference with wireless transmissions can be used as a launch pad for mounting Denial-of-Service attacks on wireless networks. Typically, jamming has been addressed under an external threat model. However, adversaries with internal knowledge of protocol specifications and network secrets can launch low-effort jamming attacks that are difficult to detect and counter. In this work, we address the problem of selective jamming attacks in wireless networks. In these attacks, the adversary is active only for a short period of time, selectively targeting messages of high importance. We illustrate the advantages of selective jamming in terms of network performance degradation and adversary effort by presenting two case studies; a selective attack on TCP and one on routing.We show that selective jamming attacks can be launched by

performing real-time packet classification at the physical layer. To mitigate these attacks, we develop three schemes that prevent real-time packet classification by combining cryptographic primitives with physical-layer attributes. We analyze the security of our methods and evaluate their computational and communication overhead.

Title: AccountTrade: Accountability against Dishonest Big Data Buyers and Sellers

Abstract—In this paper, a set of accountable protocols denoted as AccountTrade is proposed for big data trading among dishonest consumers. For achieving a secure big data trading environment, AccountTrade achieves book-keeping ability and accountability against dishonest consumers throughout the trading (i.e., buying and selling) of datasets. We investigate the consumers' responsibilities in the dataset trading, then we design AccountTrade to achieve accountability against dishonest consumers that are likely to deviate from the responsibilities. Specifically, a uniqueness index is defined and proposed, which is a new rigorous measurement of the data uniqueness for this purpose. Furthermore, several accountable trading protocols are presented to enable data brokers to blame the misbehaving entities when misbehavior is detected. The accountability of AccountTrade is formally defined, proved, and evaluated by an automatic verification tool as well as extensive simulation with real-world datasets. Our evaluation shows that AccountTrade incurs at most 10-kB storage overhead per file, and it is capable of 8-1000 concurrent data upload requests per server.

Title: Credit Card Fraud Detection Using AdaBoost and Majority Voting

Abstract—Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are first used. Then, hybrid methods which use AdaBoost and majority voting methods are applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then, a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

Title: Applying Data Mining techniques in Cyber Crimes

Abstract-Globally the internet is been accessed by enormous people within their restricted domains. When the client and server exchange messages among each other, there is an activity that can be observed in log files. Log files give a detailed description of the activities that occur in a network that shows the IP address, login and logout durations, the user's behavior etc. There are several types of attacks occurring from

the internet. Our focus of research in this paper is Denial of Service (DoS) attacks with the help of pattern recognition techniques in data mining. Through which the Denial of Service attack is identified. Denial of service is a very dangerous attack that jeopardizes the IT resources of an organization by overloading with imitation messages or multiple requests from unauthorized users.

Title: Predicting Transitional Interval of Kidney Disease Stages 3 to 5 Using Data Mining Method

Abstract- The number of kidney disease patients, one of the worldwide public health problems, has been increased yearly. Due to the high possibility of death within a short period of time, a patient must be hospitalized and appropriately cured since the first day of being diagnosed as stage 3. This is due to the fact that the patient's stage progression depends pretty much on medical history and treatment. Moreover, kidney dialysis for the stage-5 patient and end stage can be very costly, where few can afford this treatment, especially in Thailand. The challenging issue is to disclose patterns of transitional interval, with the possibility to delay the stage development. Therefore, the main objective of this study is to create a classification model for predicting transitional interval of Kidney disease stages 3 to 5. The existing medical records of Hemodialysis patient from Phan Hospital, Chiang Rai, Thailand, have been exploited as the case study. Decision tree, Knearest neighbor, Nalve Bayes and Artificial neural networks were used for eliciting the knowledge and creating classification model with the selected set of attributes. Based on the experiment results, the proposed classification framework is promising as a decision support tool. This can also be useful for Thai armed forces, especially since a large sum of budget and resources have been allocated to staffs and relatives with kidney disease.

Title: Framework For Data Model to Personalized Health Systems.

Abstract-You may feel there's nothing you can do about stress. The bills won't stop coming, there will never be more hours in the day, and your work and family responsibilities will always be demanding. But you have a lot more control than you might think. Stress management is all about taking charge: of your lifestyle, thoughts, emotions, and the way you deal with problems. No matter how stressful your life seems, there are steps you can take to relieve the pressure and regain control. It's easy to identify sources of stress following a major life event such as changing jobs, moving home, or losing a loved one, but pinpointing the sources of everyday stress can be more complicated. It's all too easy to overlook your own thoughts, feelings, and behaviors that contribute to your stress levels. Sure, you may know that you're constantly worried about work deadlines, but maybe it's your procrastination, rather than the actual job demands, that is causing the stress.

Title: Automatic Generation of Social Event Storyboard from Image Click-through Data

Abstract- Recent studies have shown that a noticeable percentage of web search traffic is about social events. While traditional websites can only show human-edited events, in this paper we present a novel system to automatically detect events from search log data and generate storyboards where the events are arranged chronologically. We chose image search log as the resource for event mining, as search logs can directly reflect people's interests. To discover events from log data, we present a Smooth Nonnegative Matrix Factorization framework (SNMF) which combines the information of query semantics, temporal correlations, search logs and time continuity. Moreover, we consider the time factor an important element since different events will develop in different time tendencies. In addition, to provide a media-rich and visually appealing storyboard, each event is associated with a set of representative photos arranged along a timeline. These relevant photos are automatically selected from image search results by analyzing image content features. We use celebrities as our test domain, which takes a large percentage of image search traffics. Experiments consisting of web search traffic on 200 celebrities, for a period of six months, show very encouraging results compared with handcrafted editorial storyboards.

IGeekS Technologies

Bridging Technology.

No:19, MN Complex, 2nd Cross,
Sampige Main Road, Malleswaram,
Bangalore Karnataka (560003) India.
Above HOP Salon,
Opp. Joyalukkas, Malleswaram, Land mark : Near to Mantri Mall, Malleswaram Bangalore.
Email: nanduigeeks2010@gmail.com ,
nandu@igeekstechnologies.com
Office Phone:
9590544567 / 7019280372
Contact Person:
Mr. Nandu Y,
Director-Projects,
Mobile: 9590544567,7019280372
E-mail: nandu@igeekstechnologies.com
nanduigeeks2010@gmail.com

RAJAJINAGAR:

#531, 63rd Cross,
12th Main, after sevabhai hospital,
5th Block, Rajajinagar,
Bangalore-10.
Landmark: Near Bashyam circle.

JAYANAGAR:

#65, 'Bhagyadeep', 8th 'B' Main, 27th Cross,
Jayanagar 3rd Block (Next to Pizza Hut),Bangalore 560011.

More than 13 years' experience in IEEE Final Year Project Center, IGEEKS Technologies Supports you in Java, IOT, Python, Bigdata Hadoop, Machine Learning, Data Mining, Networking, Embedded, VLSI, MATLAB, Power Electronics, Power System Technologies.

For Titles and Abstracts visit our website www.makefinalyearproject.com