<table>
<thead>
<tr>
<th>Project Code</th>
<th>COMPUTER SCIENCE (CSE or IT) IEEE PAPERS – 2016 PROJECT TITLES WITH ABSTRACT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>DATA MINING</strong></td>
</tr>
<tr>
<td>DM16NXT01</td>
<td><strong>TITLE:</strong> Cost Minimization for Rule Caching in Software Defined Networking</td>
</tr>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> Software-defined networking (SDN) is an emerging network paradigm that simplifies network management by decoupling the control plane and data plane, such that switches become simple data forwarding devices and network management is controlled by logically centralized servers. In SDN-enabled networks, network flow is managed by a set of associated rules that are maintained by switches in their local Ternary Content Addressable Memories (TCAMs) which support high-speed parallel lookup on wildcard patterns. Since TCAM is an expensive hardware and extremely power-hungry, each switch has only limited TCAM space and it is inefficient and even infeasible to maintain all rules at local switches. On the other hand, if we eliminate TCAM occupation by forwarding all packets to the centralized controller for processing, it results in a long delay and heavy processing burden on the controller. In this paper, we strive for the fine balance between rule caching and remote packet processing by formulating a minimum weighted flow provisioning (MWFP) problem with an objective of minimizing the total cost of TCAM occupation and remote packet processing. We propose an efficient offline algorithm if the network traffic is given, otherwise, we propose two online algorithms with guaranteed competitive ratios.</td>
</tr>
<tr>
<td>DM16NXT02</td>
<td><strong>TITLE:</strong> On Binary Decomposition based Privacy-Preserving Aggregation Schemes in Real-time Monitoring Systems</td>
</tr>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> In real-time monitoring systems, fine-grained measurements would pose great privacy threats to the participants as real-time measurements could disclose accurate people-centric activities. Differential privacy has been proposed to formalize and guide the design of privacy-preserving schemes. Nonetheless, due to the correlations and high fluctuations in time-series data, it</td>
</tr>
</tbody>
</table>
is hard to achieve an effective privacy and utility tradeoff by differential privacy mechanisms. To address this issue, in this paper, we first proposed novel multi-dimensional decomposition based schemes to compress the noise and enhance the utility in differential privacy. The key idea is to decompose the measurements into multi-dimensional records and to achieve differential privacy in bounded dimensions so that the error caused by unbounded measurements can be significantly reduced. We then extended our developed scheme and developed a binary decomposition scheme for privacy-preserving time-series aggregation in real-time monitoring systems. Through a combination of extensive theoretical analysis and experiments, our data shows that our proposed schemes can effectively improve usability while achieving the same level of differential privacy than existing schemes.

| DM16NXT03 | TITLE: An OpenMP Extension that Supports Thread-Level Speculation  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT: OpenMP directives are the de-facto standard for shared-memory parallel programming. However, OpenMP does not guarantee the correctness of the parallel execution of a given loop if runtime data dependences arise. Consequently, many highly-parallel regions cannot be safely parallelized with OpenMP due to the possibility of a dependence violation. In this paper, we propose to augment OpenMP capabilities, by adding thread-level speculation (TLS) support. Our contribution is threefold. First, we have defined a new speculative clause for variables inside parallel loops. This clause ensures that all accesses to these variables will be carried out according to sequential semantics. Second, we have created a new, software-based TLS runtime library to ensure correctness in the parallel execution of OpenMP loops that include speculative variables. Third, we have developed a new GCC plugin, which seamlessly translates our OpenMP speculative clause into calls to our TLS runtime engine. The result is the ATLaS C Compiler framework, which takes advantage of TLS techniques to expand OpenMP functionalities, and guarantees the sequential semantics of any parallelized loop.</td>
</tr>
</tbody>
</table>
DM16NXT04  TITLE: Real-Time Semantic Search Using Approximate Methodology for Large-Scale Storage Systems

ABSTRACT: The challenges of handling the explosive growth in data volume and complexity cause the increasing needs for semantic queries. The semantic queries can be interpreted as the correlation-aware retrieval, while containing approximate results. Existing cloud storage systems mainly fail to offer an adequate capability for the semantic queries. Since the true value or worth of data heavily depends on how efficiently semantic search can be carried out on the data in (near-) real-time, large fractions of data end up with their values being lost or significantly reduced due to the data staleness. To address this problem, we propose a near-real-time and cost-effective semantic queries based methodology, called FAST. The idea behind FAST is to explore and exploit the semantic correlation within and among datasets via correlation-aware hashing and manageable flat-structured addressing to significantly reduce the processing latency, while incurring acceptably small loss of data-search accuracy. The near-real-time property of FAST enables rapid identification of correlated files and the significant narrowing of the scope of data to be processed. FAST supports several types of data analytics, which can be implemented in existing searchable storage systems. We conduct a real-world use case in which children reported missing in an extremely crowded environment (e.g., a highly popular scenic spot on a peak tourist day) are identified in a timely fashion by analyzing 60 million images using FAST. FAST is further improved by using semantic-aware namespace to provide dynamic and adaptive namespace management for ultra-large storage systems. Extensive experimental results demonstrate the efficiency and efficacy of FAST in the performance improvements.

DM16NXT05  TITLE: Incremental and Decremental Max-flow for Online Semi-supervised Learning

ABSTRACT: In classification, if a small number of instances is added or removed, incremental and decremental techniques can be applied to quickly update the model. However, the design of incremental and decremental algorithms involves
many considerations. In this paper, we focus on linear classifiers including logistic regression and linear SVM because of their simplicity over kernel or other methods. By applying a warm start strategy, we investigate issues such as using primal or dual formulation, choosing optimization methods, and creating practical implementations. Through theoretical analysis and practical experiments, we conclude that a warm start setting on a high-order optimization method for primal formulations is more suitable than others for incremental and decremental learning of linear classification.

<table>
<thead>
<tr>
<th>DM16NXT06</th>
<th>TITLE: Personalized Influential Topic Search via Social Network Summarization</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: Social networks are a vital mechanism to disseminate information to friends and colleagues. In this work, we investigate an important problem - the personalized influential topic search, or PIT-Search in a social network: Given a keyword query q issued by a user u in a social network, a PIT-Search is to find the top-k q-related topics that are most influential for the query user u. The influence of a topic to a query user depends on the social connection between the query user and the social users containing the topic in the social network. To measure the topics’ influence at the similar granularity scale, we need to extract the social summarization of the social network regarding topics. To make effective topic-aware social summarization, we propose two random-walk based approaches: random clustering and an L-length random walk. Based on the proposed approaches, we can find a small set of representative users with assigned influential scores to simulate the influence of the large number of topic users in the social network with regards to the topic. The selected representative users are denoted as the social summarization of topic-aware influence spread over the social network. And then, we verify the usefulness of the social summarization by applying it to the problem of personalized influential topic search. Finally, we evaluate the performance of our algorithms using real-world datasets, and show the approach is efficient and effective in practice.</td>
<td></td>
</tr>
</tbody>
</table>
DM16NXT07  TITLE: Survey on Aspect-Level Sentiment Analysis
ABSTRACT: The field of sentiment analysis, in which sentiment is gathered, analyzed, and aggregated from text, has seen a lot of attention in the last few years. The corresponding growth of the field has resulted in the emergence of various subareas, each addressing a different level of analysis or research question. This survey focuses on aspect-level sentiment analysis, where the goal is to find and aggregate sentiment on entities mentioned within documents or aspects of them. An in-depth overview of the current state-of-the-art is given, showing the tremendous progress that has already been made in finding both the target, which can be an entity as such, or some aspect of it, and the corresponding sentiment. Aspect-level sentiment analysis yields very fine-grained sentiment information which can be useful for applications in various domains. Current solutions are categorized based on whether they provide a method for aspect detection, sentiment analysis, or both. Furthermore, a breakdown based on the type of algorithm used is provided. For each discussed study, the reported performance is included. To facilitate the quantitative evaluation of the various proposed methods, a call is made for the standardization of the evaluation methodology that includes the use of shared data sets. Semantically rich concept-centric aspect-level sentiment analysis is discussed and identified as one of the most promising future research direction.

DM16NXT08  TITLE: Multilabel Classification via Co-evolutionary Multilabel Hypernetwork
ABSTRACT: Multilabel classification is prevalent in many real-world applications where data instances may be associated with multiple labels simultaneously. In multilabel classification, exploiting label correlations is an essential but nontrivial task. Most of the existing multilabel learning algorithms are either ineffective or computational demanding and less scalable in exploiting label correlations. In this paper, we propose a co-evolutionary multilabel hypernetwork (Co-MLHN) as an attempt to exploit label correlations in an effective and efficient way. To this end, we firstly convert the traditional hypernetwork into a multilabel hypernetwork (MLHN) where label correlations are explicitly represented. We
then propose a co-evolutionary learning algorithm to learn an integrated classification model for all labels. The proposed Co-MLHN exploits arbitrary order label correlations and has linear computational complexity with respect to the number of labels. Empirical studies on a broad range of multilabel data sets demonstrate that Co-MLHN achieves competitive results against state-of-the-art multilabel learning algorithms, in terms of both classification performance and scalability with respect to the number of labels.

<table>
<thead>
<tr>
<th>DM16NXT09</th>
<th>TITLE: Answering Pattern Queries Using Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: Answering queries using views has proven effective for querying relational and semistructured data. This paper investigates this issue for graph pattern queries based on graph simulation. We propose a notion of pattern containment to characterize graph pattern matching using graph pattern views. We show that a pattern query can be answered using a set of views if and only if it is contained in the views. Based on this characterization, we develop efficient algorithms to answer graph pattern queries. We also study problems for determining (minimal, minimum) containment of pattern queries. We establish their complexity (from cubic-time to NP-complete) and provide efficient checking algorithms (approximation when the problem is intractable). In addition, when a pattern query is not contained in the views, we study maximally contained rewriting to find approximate answers; we show that it is in cubic-time to compute such rewriting, and present a rewriting algorithm. We experimentally verify that these methods are able to efficiently answer pattern queries on large real-world graphs.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DM16NXT10</th>
<th>TITLE: Similarity Measure Selection for Clustering Time Series Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: In the past few years, clustering has become a popular task associated with time series. The choice of a suitable distance measure is crucial to the clustering process and, given the vast number of distance measures for time series available in the literature and their diverse characteristics, this selection is not straightforward. With the objective of simplifying this task, we propose a multi-label classification framework that provides the means to</td>
<td></td>
</tr>
</tbody>
</table>
automatically select the most suitable distance measures for clustering a time series database. This classifier is based on a novel collection of characteristics that describe the main features of the time series databases and provide the predictive information necessary to discriminate between a set of distance measures. In order to test the validity of this classifier, we conduct a complete set of experiments using both synthetic and real time series databases and a set of five common distance measures. The positive results obtained by the designed classification framework for various performance measures indicate that the proposed methodology is useful to simplify the process of distance selection in time series clustering tasks.

**DM16NXT11**

**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 counterparts recommendation techniques.
<table>
<thead>
<tr>
<th>DM16NXT12</th>
<th>TITLE: Automatically Mining Facets for Queries from Their Search Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> We address the problem of finding query facets which are multiple groups of words or phrases that explain and summarize the content covered by a query. We assume that the important aspects of a query are usually presented and repeated in the query’s top retrieved documents in the style of lists, and query facets can be mined out by aggregating these significant lists. We propose a systematic solution, which we refer to as QDMiner, to automatically mine query facets by extracting and grouping frequent lists from free text, HTML tags, and repeat regions within top search results. Experimental results show that a large number of lists do exist and useful query facets can be mined by QDMiner. We further analyze the problem of list duplication, and find better query facets can be mined by modeling fine-grained similarities between lists and penalizing the duplicated lists.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DM16NXT13</th>
<th>TITLE: Booster in High Dimensional Data Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Classification problems in high dimensional data with a small number of observations are becoming more common especially in microarray data. During the last two decades, lots of efficient classification models and feature selection (FS) algorithms have been proposed for higher prediction accuracies. However, the result of an FS algorithm based on the prediction accuracy will be unstable over the variations in the training set, especially in high dimensional data. This paper proposes a new evaluation measure Q-statistic that incorporates the stability of the selected feature subset in addition to the prediction accuracy. Then, we propose the Booster of an FS algorithm that boosts the value of the Q-statistic of the algorithm applied. Empirical studies based on synthetic data and 14 microarray data sets show that Booster boosts not only the value of the Q-statistic but also the prediction accuracy of the algorithm applied unless the data set is intrinsically difficult to predict with the given algorithm.</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td><strong>Building an Intrusion Detection System Using A Filter-Based Feature Selection Algorithm</strong></td>
</tr>
</tbody>
</table>
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. Experimental results on a large dataset constructed from the largest Chinese microblogging service Sina Weibo and the largest Chinese B2C e-commerce website JingDong have shown the effectiveness of our proposed framework.
DM16NXT16  TITLE: Cross-Domain Sentiment Classification Using Sentiment Sensitive Embeddings

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

DM16NXT17  TITLE: Crowdsourcing for Top-K Query Processing Over Uncertain Data

ABSTRACT: Both social media and sensing infrastructures are producing an unprecedented mass of data characterized by their uncertain nature, due to either the noise inherent in sensors or the imprecision of human contributions. Therefore query processing over uncertain data has become an active research
field. In the well-known class of applications commonly referred to as top-K queries, the objective is to find the best K objects matching the user's information need, formulated as a scoring function over the objects' attribute values. If both the data and the scoring function are deterministic, the best K objects can be univocally determined and totally ordered so as to produce a single ranked result set (as long as ties are broken by some deterministic rule). However, in application scenarios involving uncertain data and fuzzy information needs, this does not hold: when either the attribute values or the scoring function are nondeterministic, there may be no consensus on a single ordering, but rather a space of possible orderings. To determine the correct ordering, one needs to acquire additional information so as to reduce the amount of uncertainty associated with the queried data and consequently the number of orderings in such a space. An emerging trend in data processing is crowdsourcing, defined as the systematic engagement of humans in the resolution of tasks through online distributed work. Our approach combines human and automatic computation in order to solve complex problems: when data ambiguity can be resolved by human judgment, crowdsourcing becomes a viable tool for converging towards a unique or at least less uncertain query result.

**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. 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.

DM16NXT19 | TITLE: Domain-Sensitive Recommendation with User-Item Subgroup Analysis

ABSTRACT: Collaborative Filtering (CF) is one of the most successful recommendation approaches to cope with information overload in the real world. However, typical CF methods equally treat every user and item, and cannot distinguish the variation of user's interests across different domains. This violates the reality that user's interests always center on some specific domains, and the users having similar tastes on one domain may have totally different tastes on another domain. Motivated by the observation, in this paper, we propose a novel Domain-sensitive Recommendation (DsRec) algorithm, to make the rating prediction by exploring the user-item subgroup analysis simultaneously, in which a user-item subgroup is deemed as a domain consisting of a subset of items with similar attributes and a subset of users who have interests in these items. The proposed framework of DsRec includes three components: a matrix factorization model for the observed rating reconstruction, a bi-clustering model for the user-item subgroup analysis, and two regularization terms to connect the above two components into a unified formulation. Extensive experiments on Movielen-100K and two real-world product review datasets show that our method achieves the better performance in terms of prediction accuracy criterion over the state-of-the-art methods.
| DM16NXT20 | **TITLE:** Efficient Algorithms for Mining Top-K High Utility Itemsets  
**ABSTRACT:** High utility itemsets (HUIs) mining is an emerging topic in data mining, which refers to discovering all itemsets having a utility meeting a user-specified minimum utility threshold min_util. However, setting min_util appropriately is a difficult problem for users. Generally speaking, finding an appropriate minimum utility threshold by trial and error is a tedious process for users. If min_util is set too low, too many HUIs will be generated, which may cause the mining process to be very inefficient. On the other hand, if min_util is set too high, it is likely that no HUIs will be found. In this paper, we address the above issues by proposing a new framework for top-K high utility itemset mining, where k is the desired number of HUIs to be mined. Two types of efficient algorithms named TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining such itemsets without the need to set min_util. We provide a structural comparison of the two algorithms with discussions on their advantages and limitations. Empirical evaluations on both real and synthetic datasets show that the performance of the proposed algorithms is close to that of the optimal case of state-of-the-art utility mining algorithms. |
| DM16NXT21 | **TITLE:** Use of Reality Mining Dataset for Human Behavior Analysis -- A Survey  
**ABSTRACT:** Recent advancements in mobile technologies have designed a new kind of device: the smart phone. Smart phones have become quintessential for a large fraction of people in developed as well as developing countries in their day to day life. The device integrates a number of sensor technologies for automatic observation which can provide data with high accuracy & precision, can be programmed to interact with the user, and can have communication with remote researchers. This allows relentless and cost-effective access to erstwhile inaccessible history of data on everyday social behavior, such as intimacy of people, phone calls, and movement patterns. In this paper, the concept and applications of reality mining are summarized and its significance towards... |
human behavior analysis is illustrated. The sensor data collected from smart phones is experimented with survey data by applying machine learning algorithms to predict human behavior in social networks. This survey conjointly highlights reality mining applications, future issues and challenges.

<table>
<thead>
<tr>
<th>Title</th>
<th>Abstract</th>
</tr>
</thead>
</table>
| **DM16NXT22** | **Title:** A framework of data mining for logging reservoir evaluation  
**Abstract:** Data mining provides an automatic and effective way for reservoir evaluation based on well logging interpretation. In this paper, we propose a framework of data mining for logging reservoir evaluation and verify its performance on two well logging dataset. Experimental results show that the conclusions of well-logging interpretation by the framework proposed in this paper are consistent with the test oil results. Compared with traditional evaluation methods, it is efficient and not so dependent on expertise. This framework provides a reference for the construction of big data platform for oil exploration and development. |
| **DM16NXT23** | **Title:** Research and Implementation of the Positioning System for Coal Mining Staff Based on Random Forests  
**Abstract:** Considering the problem that accuracy, stability, real-time performance and resources occupation of the positioning system in actual production are difficult to be met together, this paper proposes a positioning algorithm for coal mining staff based on random forests, and also implements the algorithm in the smart phone. By simulating the production environment, we get a quantity of test data and verify the proposed model in this paper. The results shows that the accuracy of the model reaches around 90%, it also has good real-time, high stability and occupies few resources, thus the system can fulfill the demands of actual production environment. |
| **DM16NXT24** | **Title:** Human heart disease prediction system using data mining techniques  
**Abstract:** Nowadays, health disease are increasing day by day due to life style, hereditary. Especially, heart disease has become more common these days, i.e. life of people is at risk. Each individual has different values for Blood pressure, |
cholesterol and pulse rate. But according to medically proven results the normal values of Blood pressure is 120/90, cholesterol is and pulse rate is 72. This paper gives the survey about different classification techniques used for predicting the risk level of each person based on age, gender, Blood pressure, cholesterol, pulse rate. The patient risk level is classified using data mining classification techniques such as Naïve Bayes, KNN, Decision Tree Algorithm, Neural Network. etc., Accuracy of the risk level is high when using more number of attributes.

DM16NXT25

TITLE: Framework for data model to personalized health systems

ABSTRACT: When large amounts of data is handled, it is important to obtain the desired compatibility between such data to perform activities of access and storage of information; data models are a tool that helps to determine the structure of the information, in order to improve communication and accuracy in applications that use and exchange data with each other for a common purpose. Nowadays, there is no framework for health supporting the data modeling design, i.e. the existing models are generic and therefore are not suitable to support personalized systems and they do not consider the quality of clinical and personal data, required in health care. Based on the CRISP-DM methodology, a framework is proposed to design a data model for personalized health systems. This framework ensures the security of personal and clinical data to relate it with health standards, particularly with the Personal Health (PHR) ISO/TR 14292 standard, which addresses the recommendations of the parameters that must be within a personalized health system. To perform accurate recommendations it is important to make a data mining process, where the data is related to guarantee an accurate and reliable personalization; these relations generated by the model should be taken into account to apply them a data mining technique.
BIG DATA

BD16NXT01

TITLE: Self-Healing in Mobile Networks with Big Data

ABSTRACT: Mobile networks have rapidly evolved in recent years due to the increase in multimedia traffic and offered services. This has led to a growth in the volume of control data and measurements that are used by self-healing systems. To maintain a certain quality of service, self-healing systems must complete their tasks in a reasonable time. The conjunction of a big volume of data and the limitation of time requires a big data approach to the problem of self-healing. This article reviews the data that self-healing uses as input and justifies its classification as big data. Big data techniques applied to mobile networks are examined, and some use cases along with their big data solutions are surveyed.

BD16NXT02

TITLE: A Parallel Patient Treatment Time Prediction Algorithm and Its Applications in Hospital Queuing-Recommendation in a Big Data Environment

ABSTRACT: Effective patient queue management to minimize patient wait delays and patient overcrowding is one of the major challenges faced by hospitals. Unnecessary and annoying waits for long periods result in substantial human resource and time wastage and increase the frustration endured by patients. For each patient in the queue, the total treatment time of all the patients before him is the time that he must wait. It would be convenient and preferable if the patients could receive the most efficient treatment plan and know the predicted waiting time through a mobile application that updates in real time. Therefore, we propose a Patient Treatment Time Prediction (PTTP) algorithm to predict the waiting time for each treatment task for a patient. We use realistic patient data from various hospitals to obtain a patient treatment time model for each task. Based on this large-scale, realistic dataset, the treatment time for each patient in the current queue of each task is predicted. Based on the predicted waiting time, a Hospital Queuing-Recommendation (HQR) system is developed. HQR calculates and predicts an efficient and convenient treatment plan recommended for the patient. Because of the large-scale, realistic dataset and the requirement for
real-time response, the PTTP algorithm and HQR system mandate efficiency and low-latency response. We use an Apache Spark-based cloud implementation at the National Supercomputing Center in Changsha to achieve the aforementioned goals. Extensive experimentation and simulation results demonstrate the effectiveness and applicability of our proposed model to recommend an effective treatment plan for patients to minimize their wait times in hospitals.

| BD16NXT03 | TITLE: Protection of Big Data Privacy |
| ABSTRACT: In recent years, big data have become a hot research topic. The increasing amount of big data also increases the chance of breaching the privacy of individuals. Since big data require high computational power and large storage, distributed systems are used. As multiple parties are involved in these systems, the risk of privacy violation is increased. There have been a number of privacy-preserving mechanisms developed for privacy protection at different stages (e.g., data generation, data storage, and data processing) of a big data life cycle. The goal of this paper is to provide a comprehensive overview of the privacy preservation mechanisms in big data and present the challenges for existing mechanisms. In particular, in this paper, we illustrate the infrastructure of big data and the state-of-the-art privacy-preserving mechanisms in each stage of the big data life cycle. Furthermore, we discuss the challenges and future research directions related to privacy preservation in big data. |

| BD16NXT04 | TITLE: Big Data Analytics in Mobile Cellular Networks |
| ABSTRACT: Mobile cellular networks have become both the generators and carriers of massive data. Big data analytics can improve the performance of mobile cellular networks and maximize the revenue of operators. In this paper, we introduce an unified data model based on random matrix theory and machine learning. Then, we present an architectural framework for applying big data analytics in mobile cellular networks. Moreover, we describe several illustrative examples, including big signaling data, big traffic data, big location data, big radio waveforms data, and big heterogeneous data in mobile cellular networks. Finally, we discuss a number of open research challenges of big data |
| BD16NXT05 | **TITLE:** Spark-based Large-scale Matrix Inversion for Big Data Processing  
**ABSTRACT:** Matrix inversion is a fundamental operation for solving linear equations for many computational applications, especially for various emerging big data applications. However, it is a challenging task to invert large-scale matrices of extremely high order (several thousands or millions), which are common in most web-scale systems such as social networks and recommendation systems. In this paper, we present a LU decomposition-based block-recursive algorithm for large-scale matrix inversion. We present its well-designed implementation with optimized data structure, reduction of space complexity and effective matrix multiplication on the Spark parallel computing platform. The experimental evaluation results show that the proposed algorithm is efficient to invert large-scale matrices on a cluster composed of commodity servers and is scalable for inverting even larger matrices. The proposed algorithm and implementation will become a solid foundation for building a high-performance linear algebra library on Spark for big data processing and applications. |
| BD16NXT06 | **TITLE:** A Tutorial on Secure Outsourcing of Large-scale Computations for Big Data  
**ABSTRACT:** Today's society is collecting a massive and exponentially growing amount of data that can potentially revolutionize scientific and engineering fields, and promote business innovations. With the advent of cloud computing, in order to analyze data in a cost-effective and practical way, users can outsource their computing tasks to the cloud, which offers access to vast computing resources on an on-demand and pay-per-use basis. However, since users' data contains sensitive information that needs to be kept secret for ethical, security, or legal reasons, many users are reluctant to adopt cloud computing. To this end, researchers have proposed techniques that enable users to offload computations to the cloud while protecting their data privacy. In this paper, we review the recent advances in the secure outsourcing of large-scale |
computations for a big data analysis. We first introduce two most fundamental and common computational problems, i.e., linear algebra and optimization, and then provide an extensive review of the data privacy preserving techniques. After that, we explain how researchers have exploited the data privacy preserving techniques to construct secure outsourcing algorithms for large-scale computations.

### BD16NXT07

**TITLE:** A Cloud Service Architecture for Analyzing Big Monitoring Data

**ABSTRACT:** Cloud monitoring is of a source of big data that are constantly produced from traces of infrastructures, platforms, and applications. Analysis of monitoring data delivers insights of the system's workload and usage pattern and ensures workloads are operating at optimum levels. The analysis process involves data query and extraction, data analysis, and result visualization. Since the volume of monitoring data is big, these operations require a scalable and reliable architecture to extract, aggregate, and analyze data in an arbitrary range of granularity. Ultimately, the results of analysis become the knowledge of the system and should be shared and communicated. This paper presents our cloud service architecture that explores a search cluster for data indexing and query. We develop REST APIs that the data can be accessed by different analysis modules. This architecture enables extensions to integrate with software frameworks of both batch processing (such as Hadoop) and stream processing (such as Spark) of big data. The analysis results are structured in Semantic Media Wiki pages in the context of the monitoring data source and the analysis process. This cloud architecture is empirically assessed to evaluate its responsiveness when processing a large set of data records under node failures.

### BD16NXT08

**TITLE:** Data and Energy Integrated Communication Networks for Wireless Big Data

**ABSTRACT:** This paper describes a new type of communication network called data and energy integrated communication networks (DEINs), which integrates the traditionally separate two processes, i.e., wireless information transfer (WIT) and wireless energy transfer (WET), fulfilling co-transmission of data and energy.
In particular, the energy transmission using radio frequency is for the purpose of energy harvesting (EH) rather than information decoding. One driving force of the advent of DEINs is wireless big data, which comes from wireless sensors that produce a large amount of small piece of data. These sensors are typically powered by battery that drains sooner or later and will have to be taken out and then replaced or recharged. EH has emerged as a technology to wirelessly charge batteries in a contactless way. Recent research work has attempted to combine WET with WIT, typically under the label of simultaneous wireless information and power transfer. Such work in the literature largely focuses on the communication side of the whole wireless networks with particular emphasis on power allocation. The DEIN communication network proposed in this paper regards the convergence of WIT and WET as a full system that considers not only the physical layer but also the higher layers, such as media access control and information routing. After describing the DEIN concept and its high-level architecture/protocol stack, this paper presents two use cases focusing on the lower layer and the higher layer of a DEIN network, respectively. The lower layer use case is about a fair resource allocation algorithm, whereas the high-layer section introduces an efficient data forwarding scheme in combination with EH. The two case studies aim to give a better explanation of the DEIN concept. Some future research directions and challenges are also pointed out.

**BD16NXT09**

**TITLE:** Wide Area Analytics for Geographically Distributed Datacenters  
**ABSTRACT:** Big data analytics, the process of organizing and analyzing data to get useful information, is one of the primary uses of cloud services today. Traditionally, collections of data are stored and processed in a single datacenter. As the volume of data grows at a tremendous rate, it is less efficient for only one datacenter to handle such large volumes of data from a performance point of view. Large cloud service providers are deploying datacenters geographically around the world for better performance and availability. A widely used approach for analytics of geo-distributed data is the centralized approach, which aggregates all the raw data from local datacenters to a central datacenter.
However, it has been observed that this approach consumes a significant amount of bandwidth, leading to worse performance. A number of mechanisms have been proposed to achieve optimal performance when data analytics are performed over geo-distributed datacenters. In this paper, we present a survey on the representative mechanisms proposed in the literature for wide area analytics. We discuss basic ideas, present proposed architectures and mechanisms, and discuss several examples to illustrate existing work. We point out the limitations of these mechanisms, give comparisons, and conclude with our thoughts on future research directions.

**TITLE:** Privacy Preserving Deep Computation Model on Cloud for Big Data Feature Learning

**ABSTRACT:** To improve the efficiency of big data feature learning, the paper proposes a privacy preserving deep computation model by offloading the expensive operations to the cloud. Privacy concerns become evident because there are a large number of private data by various applications in the smart city, such as sensitive data of governments or proprietary information of enterprises. To protect the private data, the proposed model uses the BGV encryption scheme to encrypt the private data and employs cloud servers to perform the high-order back-propagation algorithm on the encrypted data efficiently for deep computation model training. Furthermore, the proposed scheme approximates the Sigmoid function as a polynomial function to support the secure computation of the activation function with the BGV encryption. In our scheme, only the encryption operations and the decryption operations are performed by the client while all the computation tasks are performed on the cloud. Experimental results show that our scheme is improved by approximately 2.5 times in the training efficiency compared to the conventional deep computation model without disclosing the private data using the cloud computing including ten nodes. More importantly, our scheme is highly scalable by employing more cloud servers, which is particularly suitable for big data.
<table>
<thead>
<tr>
<th>TITLE: Assessing Big Data SQL Frameworks for Analyzing Event Logs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Performing Process Mining by analyzing event logs generated by various systems is a very computation and I/O intensive task. Distributed computing and Big Data processing frameworks make it possible to distribute all kinds of computation tasks to multiple computers instead of performing the whole task in a single computer. This paper assesses whether contemporary structured query language (SQL) supporting Big Data processing frameworks are mature enough to be efficiently used to distribute computation of two central Process Mining tasks to two dissimilar clusters of computers providing BPM as a service in the cloud. Tests are performed by using a novel automatic testing framework detailed in this paper and its supporting materials. As a result, an assessment is made on how well selected Big Data processing frameworks manage to process and to parallelize the analysis work required by Process Mining tasks.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TITLE: A Comparative Study of Various Clustering Techniques on Big Data Sets Using Apache Mahout</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Clustering algorithms have materialized as an unconventional tool to precisely examine the immense volume of data produced by present applications. In specific, their main objective is to classify data into clusters such that objects are grouped in the same cluster when they are similar rendering to particular metrics and dissimilar to objects of other groups. From the machine learning perspective clustering can be viewed as unsupervised learning of concepts. Hadoop is a distributed file system and an open-source implementation of Map Reduce dealing with big data. Apache Mahout clustering algorithms are implemented on top of Hadoop using Map Reduce paradigm. In this paper three clustering algorithms are described: K-means, Fuzzy K-Means (FKM) and Canopy clustering implemented by using Apache Mahout as well as providing a comparison. In addition, we underlined the clustering algorithms that are the preeminent performing for big data.</td>
</tr>
<tr>
<td>Title</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Data Analysis for Chronic Disease - Diabetes Using Map Reduce Technique</strong></td>
</tr>
<tr>
<td><strong>Novel Scheduling Algorithms for Efficient Deployment of MapReduce Applications in Heterogeneous Computing Environments</strong></td>
</tr>
<tr>
<td>Title</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td><strong>BD16NXT15</strong>&lt;br&gt;Title: Evaluate H2Hadoop and Amazon EMR Performances by Processing MR Jobs in Text Data Sets&lt;br&gt;<strong>Abstract:</strong> Text data is defined as sequences of characters that may become big data that has no specific format and only can be processed using the original Hadoop. Amazon Web Services AWS provides virtual Cloud Computing services such as storing data using S3 service and processing big data using EMR service. Amazon Elastic Map Reduce EMR uses the original Hadoop as a processing environment to its Cloud Computing services. Also, H2Hadoop is a developed version of Hadoop that provides big data processing service that uses the metadata of related jobs to improve Hadoop performance. In this paper, we process a find sequence job in text data using Amazon EMR and H2Hadoop, and we came up with a comparison between them that shows H2Hadoop performance is more efficient than Amazon EMR in some cases under different considerations.</td>
</tr>
</tbody>
</table>
build the execution profile of individual Hive queries by extracting information from HIVE and Hadoop logs. The profile consists of detailed information about Map Reduce jobs, tasks and attempts belonging to a query. It is stored as a JSON document in MongoDB and can be retrieved to generate reports in charts or tables. We have run several experiments on AWS with TPC-H data sets and queries to demonstrate that our profiling tool is able to assist developers in comparing HIVE queries written in different formats, running on different data sets and configured with different parameters. It is also able to compare tasks/attempts within the same job to diagnose performance issues.

| BD16NXT17 | TITLE: An Efficient Key Partitioning Scheme for Heterogeneous MapReduce Clusters  
ABSTRACT: Hadoop is a standard implementation of MapReduce framework for running data-intensive applications on the clusters of commodity servers. By thoroughly studying the framework we find out that the shuffle phase, all-to-all input data fetching phase in reduce task significantly affect the application performance. There is a problem of variance in both the intermediate key's frequencies and their distribution among data nodes throughout the cluster in Hadoop's MapReduce system. This variance in system causes network overhead which leads to unfairness on the reduce input among different data nodes in the cluster. Because of the above problem, applications experience performance degradation due to shuffle phase of MapReduce applications. We develop a new novel algorithm; unlike previous systems our algorithm considers a node's capabilities as heuristics to decide a better available trade-off for the locality and fairness in the system. By comparing with the default Hadoop's partitioning algorithm and Leen algorithm, on the average our approach achieve performance gain of 29% and 17%, respectively. |

| BD16NXT18 | TITLE: An Improved HDFS for Small File  
ABSTRACT: Hadoop is an open source distributed computing platform, and HDFS is Hadoop distributed file system. The HDFS has a powerful data storage capacity. Therefore, it is suitable for cloud storage system. However, HDFS was |
originally developed for the streaming access on large software, it has low storage efficiency for massive small files. To solve this problem, the HDFS file storage process is improved. The files are judged before uploading to HDFS clusters. If the file is a small file, it is merged and the index information of the small file is stored in the index file with the form of key-value pairs. The simulation shows that the improved HDFS has lower NameNode memory consumption than original HDFS and Hadoop Archives (HAR files). Thus, it can improve the access efficiency.

<table>
<thead>
<tr>
<th>BD16NXT19</th>
<th>TITLE: FiDoop-DP: Data Partitioning in Frequent Itemset Mining on Hadoop Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Traditional parallel algorithms for mining frequent itemsets aim to balance load by equally partitioning data among a group of computing nodes. We start this study by discovering a serious performance problem of the existing parallel Frequent Itemset Mining algorithms. Given a large dataset, data partitioning strategies in the existing solutions suffer high communication and mining overhead induced by redundant transactions transmitted among computing nodes. We address this problem by developing a data partitioning approach called FiDoop-DP using the MapReduce programming model. The overarching goal of FiDoop-DP is to boost the performance of parallel Frequent Itemset Mining on Hadoop clusters. At the heart of FiDoop-DP is the Voronoi diagram-based data partitioning technique, which exploits correlations among transactions. Incorporating the similarity metric and the Locality-Sensitive Hashing technique, FiDoop-DP places highly similar transactions into a data partition to improve locality without creating an excessive number of redundant transactions. We implement FiDoop-DP on a 24-node Hadoop cluster, driven by a wide range of datasets created by IBM Quest Market-Basket Synthetic Data Generator. Experimental results reveal that FiDoop-DP is conducive to reducing network and computing loads by the virtue of eliminating redundant transactions on Hadoop nodes. FiDoop-DP significantly improves the performance of the existing parallel frequent-pattern scheme by up to 31% with</td>
<td></td>
</tr>
</tbody>
</table>
| BD16NXT20 | **TITLE:** Defining Human Behaviors Using Big Data Analytics in Social Internet of Things  

**ABSTRACT:** As we delve into the Internet of Things (IoT), we are witnessing the intensive interaction and heterogeneous communication among different devices over the Internet. Consequently, these devices generate a massive volume of Big Data. The potential of these data has been analyzed by the complex network theory, describing a specialized branch, known as 'Human Dynamics.' In this extension, the goal is to describe human behavior in the social area at real-time. These objectives are starting to be practicable through the quantity of data provided by smart phones, social network, and smart cities. These make the environment more intelligent and offer an intelligent space to sense our activities or actions, and the evolution of the ecosystem. To address the aforementioned needs, this paper presents the concept of 'defining human behavior' using Big Data in SIOT by proposing system architecture that processes and analyzes big data in real-time. The proposed architecture consists of three operational domains, i.e., object, SIOT server, application domain. Data from object domain is aggregated at SIOT server domain, where the data is efficiently store and process and intelligently respond to the outer stimuli. The proposed system architecture focuses on the analysis the ecosystem provided by Smart Cities, wearable devices (e.g., body area network) and Big Data to determine the human behaviors as well as human dynamics. Furthermore, the feasibility and efficiency of the proposed system are implemented on Hadoop single node setup on UBUNTU 14.04 LTS coreTMi5 machine with 3.2 GHz processor and 4 GB memory. |
<table>
<thead>
<tr>
<th>TITLE: A Novel Approach to Predictive Graphs Using Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> In enterprise models, relationships between data entities can be expressed by graphically connecting edges and vertices based on the application domain. Such graphs are optimal for understanding and processing simpler relationship-based models. When relationships get huge and complex in terms of large real-world problems such as friend or follower networks in social media or numerous client-corporate relationships for large multinational corporations, then converting and processing these graphs in near real-time is not trivial. This is because the number of relationships (edges) can grow super-exponentially in the number of vertices. We propose a novel method to create and update scalable relationship-data graphs for visualization and for prediction. Given a Big Table (BT) of related data expressed as structured (or semi-structured) data-tuples with associated business keys, the table can be quite easily transformed into a relationship data graph using current Big Data (BD) technologies. Predictive business models (e.g., from machine learning) can then be applied to the graph by our methods given here, utilizing a combination of data-parallel and graph-parallel computations. Applying a new data update to an existing data point in the graph such as a vertex (or edge), it is possible to predict the corresponding change in the model output variable on any vertex (or edge) in the graph. By combining predictive analytics models with data updates at either graph vertices or edges, we can utilize our method to propagate data updates in near real-time to the predictive graph.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TITLE: Research of association rule algorithm based on data mining</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Association rule data mining is an important part in the field of data mining data mining, its algorithm performance directly affects the efficiency of data mining and the integrity, effectiveness of ultimate data mining results. Based on the existing association rule mining algorithms, this paper studies and analyzes their efficiency and effectiveness, and according to the efficiency defects of Apriorialgorithm, proposes an improved algorithm. This algorithm can reduce the data base I/Operation time, improve mining efficiency.</td>
</tr>
<tr>
<td>TITLE: Trust Issues for Big Data about High-Value Manufactured Parts</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>ABSTRACT:</strong> In the world of high-value manufacturing, an imperative for productivity and quality are driving the manufacturing community towards integration of all information captured about a product along the manufacturing value chain, from its design and manufacture through usage, maintenance, and decommissioning. Much of this information is already captured, but it is scattered across many organizations, with significant variation among levels of security consciousness and information technology sophistication. Any weak link in this chain of organizations constitutes a threat that can have a major negative impact for organizations all along the chain. In this paper, we explain the reasons for the move towards integration of information about high-value manufactured products. We introduce the concept of a digital thread, which is the entire set of information about the life history of a manufactured object. We outline several key threats to digital threads that have not been fully addressed in previous work on securing provenance information, and propose digital-threads-as-a-service (DTaaS) as a potential way to mitigate several of the open issues.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TITLE: Securing Big Data Environments from Attacks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> In this paper we propose techniques for securing big data environments such as public cloud with tenants using their virtual machines for different services such as utility and healthcare. Our model makes use of state based monitoring of the data sources for service specific detection of the attacks and offline traffic analysis of multiple data sources to detect attacks such as botnets.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TITLE: Investigation of Reconfigurable FPGA Design for Processing Big Data Streams</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Big Data situation has placed a tremendous pressure on the existing computational models. The challenges of Big Data call for a new approach to solve both software and hardware problems. Streaming applications is a form of on-demand software distribution. In streaming scenarios, only essential portions of an application's code need to be installed on the system, while the receiver performs the main operations. The necessary code and files are delivered over</td>
</tr>
</tbody>
</table>
the network as, and when, they are required. The hardware architecture plays an important role in improving the efficiency of a streaming system. The variance of hardware performance on different HW architectures is quite interesting. Previous work confirms that the CPUs, GPUs, and FPGAs are performing differently on specific applications. The previous efforts of hardware benchmarking show that GPUs outperformed the other platforms in terms of execution time. CPUs outperformed in overall execution combined with transfer time. FPGAs outperformed for fixed algorithms using streaming [1]. Hence, this paper evaluates the performance of streaming applications on a pipelined FPGA design. In the context of real-time processing, it elects one of the Big Data streaming problems that gets a candidate for majority element on-the-fly, that is Moore's Voting Algorithm. The performance analysis of Moore's algorithm on FPGA highlights a noticeable improvement by using a pipelining architecture.

<table>
<thead>
<tr>
<th>Project Code</th>
<th>IMAGE PROCESSING</th>
</tr>
</thead>
<tbody>
<tr>
<td>IM16NXT01</td>
<td>TITLE: Adaptive Part-Level Model Knowledge Transfer for Gender</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT: In this letter, we propose an adaptive part-level model knowledge transfer approach for gender classification of facial images based on Fisher vector (FV). Specifically, we first decompose the whole face image into several parts and compute the dense FVs on each face part. An adaptive transfer learning model is then proposed to reduce the discrepancies between the training data and the testing data for enhancing classification performance. Compared to the existing gender classification methods, the proposed approach is more adaptive to the testing data, which is quite beneficial to the performance improvement. Extensive experiments on several public domain face data sets clearly demonstrate the effectiveness of the proposed approach.</td>
</tr>
<tr>
<td>IM16NXT02</td>
<td>TITLE: Patch-Based Video Denoising With Optical Flow Estimation</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT: A novel image sequence denoising algorithm is presented. The proposed approach takes advantage of the selfsimilarity and redundancy of adjacent frames. The algorithm is inspired by fusion algorithms, and as the number of frames increases, it tends to a pure temporal average. The use of</td>
</tr>
</tbody>
</table>
Motion compensation by regularized optical flow methods permits robust patch comparison in a spatiotemporal volume. The use of principal component analysis ensures the correct preservation of fine texture and details. An extensive comparison with the state-of-the-art methods illustrates the superior performance of the proposed approach, with improved texture and detail reconstruction.

| IM16NXT03 | TITLE: 2D Orthogonal Locality Preserving Projection for Image Denoising  
**ABSTRACT:** Sparse representations using transform-domain techniques are widely used for better interpretation of the raw data. Orthogonal locality preserving projection (OLPP) is a linear technique that tries to preserve local structure of data in the transform domain as well. Vectorized nature of OLPP requires high-dimensional data to be converted to vector format, hence may lose spatial neighborhood information of raw data. On the other hand, processing 2D data directly, not only preserves spatial information, but also improves the computational efficiency considerably. The 2D OLPP is expected to learn the transformation from 2D data itself. This paper derives mathematical foundation for 2D OLPP. The proposed technique is used for image denoising task. Recent state-of-the-art approaches for image denoising work on two major hypotheses, i.e., non-local self-similarity and sparse linear approximations of the data. Locality preserving nature of the proposed approach automatically takes care of self-similarity present in the image while inferring sparse basis. A global basis is adequate for the entire image. The proposed approach outperforms several state-of-the-art image denoising approaches for gray-scale, color, and texture images. |

| IM16NXT04 | TITLE: Microwave Unmixing With Video Segmentation for Inferring Broadleaf and Needleleaf Brightness Temperatures and Abundances From Mixed Forest Observations  
**ABSTRACT:** Passive microwave sensors have better capability of penetrating forest layers to obtain more information from forest canopy and ground surface. |
For forest management, it is useful to study passive microwave signals from forests. Passive microwave sensors can detect signals from needleleaf, broadleaf, and mixed forests. The observed brightness temperature of a mixed forest can be approximated by a linear combination of the needleleaf and broadleaf brightness temperatures weighted by their respective abundances. For a mixed forest observed by an N-band microwave radiometer with horizontal and vertical polarizations, there are 2N observed brightness temperatures. It is desirable to infer 4N + 2 unknowns: 2N broadleaf brightness temperatures, 2N needleleaf brightness temperatures, 1 broadleaf abundance, and 1 needleleaf abundance. This is a challenging underdetermined problem. In this paper, we devise a novel method that combines microwave unmixing with video segmentation for inferring broadleaf and needleleaf brightness temperatures and abundances from mixed forests. We propose an improved Otsu method for video segmentation to infer broadleaf and needleleaf abundances. The brightness temperatures of needleleaf and broadleaf trees can then be solved by the nonnegative least squares solution. For our mixed forest unmixing problem, it turns out that the ordinary least squares solution yields the desired positive brightness temperatures. The experimental results demonstrate that the proposed method is able to unmix broadleaf and needleleaf brightness temperatures and abundances well. The absolute differences between the reconstructed and observed brightness temperatures of the mixed forest are well within 1 K.

**IM16NXT05**

**TITLE: Spectral–Spatial Adaptive Sparse Representation for Hyperspectral Image Denoising**

**ABSTRACT:** In this paper, a novel spectral-spatial adaptive sparse representation (SSASR) method is proposed for hyperspectral image (HSI) denoising. The proposed SSASR method aims at improving noise-free estimation for noisy HSI by making full use of highly correlated spectral information and highly similar spatial information via sparse representation, which consists of the following three steps. First, according to spectral correlation across bands, the HSI is
partitioned into several nonoverlapping band subsets. Each band subset contains multiple continuous bands with highly similar spectral characteristics. Then, within each band subset, shape-adaptive local regions consisting of spatially similar pixels are searched in spatial domain. This way, spectral-spatial similar pixels can be grouped. Finally, the highly correlated and similar spectral-spatial information in each group is effectively used via the joint sparse coding, in order to generate better noise-free estimation. The proposed SSASR method is evaluated by different objective metrics in both real and simulated experiments. The numerical and visual comparison results demonstrate the effectiveness and superiority of the proposed method.

**IM16NXT06**  
**TITLE:** A Decomposition Framework for Image Denoising Algorithms  
**ABSTRACT:** In this paper, we consider an image decomposition model that provides a novel framework for image denoising. The model computes the components of the image to be processed in a moving frame that encodes its local geometry (directions of gradients and level lines). Then, the strategy we develop is to denoise the components of the image in the moving frame in order to preserve its local geometry, which would have been more affected if processing the image directly. Experiments on a whole image database tested with several denoising methods show that this framework can provide better results than denoising the image directly, both in terms of Peak signal-to-noise ratio and Structural similarity index metrics.

**IM16NXT07**  
**TITLE:** Scalable Feature Matching by Dual Cascaded Scalar Quantization for Image Retrieval  
**ABSTRACT:** In this paper, we investigate the problem of scalable visual feature matching in large-scale image search and propose a novel cascaded scalar quantization scheme in dual resolution. We formulate the visual feature matching as a range-based neighbor search problem and approach it by identifying hyper-cubes with a dual-resolution scalar quantization strategy. Specifically, for each dimension of the PCA-transformed feature, scalar quantization is performed at both coarse and fine resolutions. The scalar
quantization results at the coarse resolution are cascaded over multiple dimensions to index an image database. The scalar quantization results over multiple dimensions at the fine resolution are concatenated into a binary super-vector and stored into the index list for efficient verification. The proposed cascaded scalar quantization (CSQ) method is free of the costly visual codebook training and thus is independent of any image descriptor training set. The index structure of the CSQ is flexible enough to accommodate new image features and scalable to index large-scale image database. We evaluate our approach on the public benchmark datasets for large-scale image retrieval. Experimental results demonstrate the competitive retrieval performance of the proposed method compared with several recent retrieval algorithms on feature quantization.

**IM16NXT08**

**TITLE:** Rotation Invariant Texture Description Using Symmetric Dense Microblock Difference

**ABSTRACT:** This letter is devoted to the problem of rotation invariant texture classification. Novel rotation invariant feature, symmetric dense microblock difference (SDMD), is proposed which captures the information at different orientations and scales. N-fold symmetry is introduced in the feature design configuration, while retaining the random structure that provides discriminative power. The symmetry is utilized to achieve a rotation invariance. The SDMD is extracted using an image pyramid and encoded by the Fisher vector approach resulting in a descriptor which captures variations at different resolutions without increasing the dimensionality. The proposed image representation is combined with the linear SVM classifier. Extensive experiments are conducted on four texture data sets [Brodatz, UMD, UIUC, and Flickr material data set (FMD)] using standard protocols. The results demonstrate that our approach outperforms the state of the art in texture classification.

**IM16NXT09**

**TITLE:** PiCode: a New Picture-Embedding 2D Barcode

**ABSTRACT:** Nowadays, 2D barcodes have been widely used as an interface to connect potential customers and advertisement contents. However, the appearance of a conventional 2D barcode pattern is often too obtrusive for
integrating into an aesthetically designed advertisement. Besides, no human readable information is provided before the barcode is successfully decoded. This paper proposes a new picture-embedding 2D barcode, called PiCode, which mitigates these two limitations by equipping a scannable 2D barcode with a picturesque appearance. PiCode is designed with careful considerations on both the perceptual quality of the embedded image and the decoding robustness of the encoded message. Comparisons with the existing beautified 2D barcodes show that PiCode achieves one of the best perceptual qualities for the embedded image, and maintains a better tradeoff between image quality and decoding robustness in various application conditions. PiCode has been implemented in the MATLAB on a PC and some key building blocks have also been ported to Android and iOS platforms. Its practicality for real-world applications has been successfully demonstrated.

IM16NXT10

TITLE: OCR Based Feature Extraction and Template Matching Algorithms for Qatari Number Plate

ABSTRACT: There are several algorithms and methods that could be applied to perform the character recognition stage of an automatic number plate recognition system; however, the constraints of having a high recognition rate and real-time processing should be taken into consideration. In this paper four algorithms applied to Qatari number plates are presented and compared. The proposed algorithms are based on feature extraction (vector crossing, zoning, combined zoning and vector crossing) and template matching techniques. All four proposed algorithms have been implemented and tested using MATLAB. A total of 2790 Qatari binary character images were used to test the algorithms. Template matching based algorithm showed the highest recognition rate of 99.5% with an average time of 1.95 ms per character.
IM16NXT11  
**TITLE:** A Hands-on Application-Based Tool for STEM Students to Understand Differentiation  

**ABSTRACT:** The main goal of this project is to illustrate to college students in science, technology, engineering, and mathematics (STEM) fields some fundamental concepts in calculus. A high-level technical computing language - MATLAB, is the core platform used in the construction of this project. A graphical user interface (GUI) is designed to interactively explain the intuition behind a key mathematical concept: differentiation. The GUI demonstrates how a derivative operation (as a form of kernel) can be applied on one-dimensional (1D) and two-dimensional (2D) images (as a form of vector). The user can interactively select from a set of predetermined operations and images in order to show how the selected kernel operates on the corresponding image. Such interactive tools in calculus courses are of great importance and need, especially for STEM students who seek an intuitive and visual understanding of mathematical notions that are often presented to them as abstract concepts. In addition to students, instructors can greatly benefit from using such tools to elucidate the use of fundamental concepts in mathematics in a real world context.

IM16NXT12  
**TITLE:** Noise Power Spectrum Measurements in Digital Imaging With Gain Nonuniformity Correction  

**ABSTRACT:** The noise power spectrum (NPS) of an image sensor provides the spectral noise properties needed to evaluate sensor performance. Hence, measuring an accurate NPS is important. However, the fixed pattern noise from the sensor's nonuniform gain inflates the NPS, which is measured from images acquired by the sensor. Detrending the low-frequency fixed pattern is traditionally used to accurately measure NPS. However, detrending methods cannot remove high-frequency fixed patterns. In order to efficiently correct the fixed pattern noise, a gain-correction technique based on the gain map can be used. The gain map is generated using the average of uniformly illuminated images without any objects. Increasing the number of images $n$ for averaging can
reduce the remaining photon noise in the gain map and yield accurate NPS values. However, for practical finite n, the photon noise also significantly inflates NPS. In this paper, a nonuniform-gain image formation model is proposed and the performance of the gain correction is theoretically analyzed in terms of the signal-to-noise ratio (SNR). It is shown that the SNR is $O(\sqrt{n})$. An NPS measurement algorithm based on the gain map is then proposed for any given n. Under a weak nonuniform gain assumption, another measurement algorithm based on the image difference is also proposed. For real radiography image detectors, the proposed algorithms are compared with traditional detrending and subtraction methods, and it is shown that as few as two images ($n = 1$) can provide an accurate NPS because of the compensation constant ($1 + 1/n$).

**IM16NXT13**

**TITLE:** A DCT-based Total JND Profile for Spatio-Temporal and Foveated Masking Effects

**ABSTRACT:** In image and video processing fields, DCT-based just noticeable difference (JND) profiles have effectively been utilized to remove perceptual redundancies in pictures for compression. In this paper, we solve two problems that are often intrinsic to the conventional DCT-based JND profiles: (i) no foveated masking (FM) JND model has been incorporated in modeling the DCT-based JND profiles; and (ii) the conventional temporal masking (TM) JND models assume that all moving objects in frames can be well tracked by the eyes and that they are projected on the fovea regions of the eyes, which is not a realistic assumption and may result in poor estimation of JND values for untracked moving objects (or image regions). To solve these two problems, we first propose a generalized JND model for joint effects between TM and FM effects. With this model, called the temporal-foveated masking (TFM) JND model, JND thresholds for any tracked/untracked and moving/still image regions can be elaborately estimated. Finally, the TFM-JND model is incorporated into a total DCT-based JND profile with a spatial contrast sensitivity function, luminance masking, and contrast masking JND models. In addition, we propose a JND adjustment method for our total JND profile to avoid overestimation of JND.
values for image blocks of fixed sizes with various image characteristics. To validate the effectiveness of the total JND profile, an experiment involving a subjective distortion visibility assessment has been conducted. The experiment results show that the proposed total DCT-based JND profile yields significant performance improvement with much higher capability of distortion concealment (average 5.6 dB lower PSNR) compared to state-of-the-art JND profiles.

<table>
<thead>
<tr>
<th>IM16NXT14</th>
<th>TITLE: Dimension Reduction With Extreme Learning Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Data may often contain noise or irrelevant information, which negatively affect the generalization capability of machine learning algorithms. The objective of dimension reduction algorithms, such as principal component analysis (PCA), non-negative matrix factorization (NMF), random projection (RP), and auto-encoder (AE), is to reduce the noise or irrelevant information of the data. The features of PCA (eigenvectors) and linear AE are not able to represent data as parts (e.g. nose in a face image). On the other hand, NMF and non-linear AE are maimed by slow learning speed and RP only represents a subspace of original data. This paper introduces a dimension reduction framework which to some extend represents data as parts, has fast learning speed, and learns the between-class scatter subspace. To this end, this paper investigates a linear and non-linear dimension reduction framework referred to as extreme learning machine AE (ELM-AE) and sparse ELM-AE (SELM-AE). In contrast to tied weight AE, the hidden neurons in ELM-AE and SELM-AE need not be tuned, and their parameters (e.g, input weights in additive neurons) are initialized using orthogonal and sparse random weights, respectively. Experimental results on USPS handwritten digit recognition data set, CIFAR-10 object recognition, and NORB object recognition data set show the efficacy of linear and non-linear ELM-AE and SELM-AE in terms of discriminative capability, sparsity, training time, and normalized mean square error.</td>
<td></td>
</tr>
<tr>
<td>IM16NXT15</td>
<td>TITLE: Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>ABSTRACT:</strong></td>
<td>In free-viewpoint video, there is a recent trend to represent scene objects as solids rather than using multiple depth maps. Point clouds have been used in computer graphics for a long time, and with the recent possibility of real-time capturing and rendering, point clouds have been favored over meshes in order to save computation. Each point in the cloud is associated with its 3D position and its color. We devise a method to compress the colors in point clouds, which is based on a hierarchical transform and arithmetic coding. The transform is a hierarchical sub-band transform that resembles an adaptive variation of a Haar wavelet. The arithmetic encoding of the coefficients assumes Laplace distributions, one per sub-band. The Laplace parameter for each distribution is transmitted to the decoder using a custom method. The geometry of the point cloud is encoded using the well-established octtree scanning. Results show that the proposed solution performs comparably with the current state-of-the-art, in many occasions outperforming it, while being much more computationally efficient. We believe this paper represents the state of the art in intra-frame compression of point clouds for real-time 3D video.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IM16NXT16</th>
<th>TITLE: Discriminant Incoherent Component Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong></td>
<td>Face images convey rich information which can be perceived as a superposition of low-complexity components associated with attributes, such as facial identity, expressions, and activation of facial action units (AUs). For instance, low-rank components characterizing neutral facial images are associated with identity, while sparse components capturing non-rigid deformations occurring in certain face regions reveal expressions and AU activations. In this paper, the discriminant incoherent component analysis (DICA) is proposed in order to extract low-complexity components, corresponding to facial attributes, which are mutually incoherent among different classes (e.g., identity, expression, and AU activation) from training data, even in the presence</td>
</tr>
</tbody>
</table>
of gross sparse errors. To this end, a suitable optimization problem, involving the minimization of nuclear-and $\ell_1$-norm, is solved. Having found an ensemble of class-specific incoherent components by the DICA, an unseen (test) image is expressed as a group-sparse linear combination of these components, where the non-zero coefficients reveal the class(es) of the respective facial attribute(s) that it belongs to. The performance of the DICA is experimentally assessed on both synthetic and real-world data. Emphasis is placed on face analysis tasks, namely, joint face and expression recognition, face recognition under varying percentages of training data corruption, subject-independent expression recognition, and AU detection by conducting experiments on four data sets.

**IM16NXT17**

**TITLE:** Comprehensive and Practical Vision System for Self-Driving Vehicle Lane-Level Localization

**ABSTRACT:** Vehicle lane-level localization is a fundamental technology in autonomous driving. To achieve accurate and consistent performance, a common approach is to use the LIDAR technology. However, it is expensive and computational demanding, and thus not a practical solution in many situations. This paper proposes a stereovision system, which is of low cost, yet also able to achieve high accuracy and consistency. It integrates a new lane line detection algorithm with other lane marking detectors to effectively identify the correct lane line markings. It also fits multiple road models to improve accuracy. An effective stereo 3D reconstruction method is proposed to estimate vehicle localization. The estimation consistency is further guaranteed by a new particle filter framework, which takes vehicle dynamics into account. Experiment results based on image sequences taken under different visual conditions showed that the proposed system can identify the lane line markings with 98.6% accuracy. The maximum estimation error of the vehicle distance to lane lines is 16 cm in daytime and 26 cm at night, and the maximum estimation error of its moving direction with respect to the road tangent is 0.06 rad in daytime and 0.12 rad at night. Due to its high accuracy and consistency, the proposed system can be implemented in autonomous driving vehicles as a practical solution to vehicle
<table>
<thead>
<tr>
<th>IM16NXT18</th>
<th>TITLE: Robust Blur Kernel Estimation for License Plate Images From Fast Moving Vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> As the unique identification of a vehicle, license plate is a key clue to uncover over-speed vehicles or the ones involved in hit-and-run accidents. However, the snapshot of over-speed vehicle captured by surveillance camera is frequently blurred due to fast motion, which is even unrecognizable by human. Those observed plate images are usually in low resolution and suffer severe loss of edge information, which cast great challenge to existing blind deblurring methods. For license plate image blurring caused by fast motion, the blur kernel can be viewed as linear uniform convolution and parametrically modeled with angle and length. In this paper, we propose a novel scheme based on sparse representation to identify the blur kernel. By analyzing the sparse representation coefficients of the recovered image, we determine the angle of the kernel based on the observation that the recovered image has the most sparse representation when the kernel angle corresponds to the genuine motion angle. Then, we estimate the length of the motion kernel with Radon transform in Fourier domain. Our scheme can well handle large motion blur even when the license plate is unrecognizable by human. We evaluate our approach on real-world images and compare with several popular state-of-the-art blind image deblurring algorithms. Experimental results demonstrate the superiority of our proposed approach in terms of effectiveness and robustness.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IM16NXT19</th>
<th>TITLE: Adaptive Pairing Reversible Watermarking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> This letter revisits the pairwise reversible watermarking scheme of Ou et al., 2013. An adaptive pixel pairing that considers only pixels with similar prediction errors is introduced. This adaptive approach provides an increased number of pixel pairs where both pixels are embedded and decreases the number of shifted pixels. The adaptive pairwise reversible watermarking outperforms the state-of-the-art low embedding bit-rate schemes proposed so</td>
<td></td>
</tr>
</tbody>
</table>
**IM16NXT20**  
**TITLE:** BIT: Biologically Inspired Tracker  

**ABSTRACT:** Visual tracking is challenging due to image variations caused by various factors, such as object deformation, scale change, illumination change, and occlusion. Given the superior tracking performance of human visual system (HVS), an ideal design of biologically inspired model is expected to improve computer visual tracking. This is, however, a difficult task due to the incomplete understanding of neurons' working mechanism in the HVS. This paper aims to address this challenge based on the analysis of visual cognitive mechanism of the ventral stream in the visual cortex, which simulates shallow neurons (S1 units and C1 units) to extract low-level biologically inspired features for the target appearance and imitates an advanced learning mechanism (S2 units and C2 units) to combine generative and discriminative models for target location. In addition, fast Gabor approximation and fast Fourier transform are adopted for real-time learning and detection in this framework. Extensive experiments on large-scale benchmark data sets show that the proposed biologically inspired tracker performs favorably against the state-of-the-art methods in terms of efficiency, accuracy, and robustness. The acceleration technique in particular ensures that biologically inspired tracker maintains a speed of approximately 45 frames/s.

**IM16NXT21**  
**TITLE:** Joint Low-Rank and Sparse Principal Feature Coding for Enhanced Robust Representation and Visual Classification  

**ABSTRACT:** Recovering low-rank and sparse subspaces jointly for enhanced robust representation and classification is discussed. Technically, we first propose a transductive low-rank and sparse principal feature coding (LSPFC) formulation that decomposes given data into a component part that encodes low-rank sparse principal features and a noise-fitting error part. To well handle the outside data, we then present an inductive LSPFC (I-LSPFC). I-LSPFC incorporates embedded low-rank and sparse principal features by a projection.
into one problem for direct minimization, so that the projection can effectively map both inside and outside data into the underlying subspaces to learn more powerful and informative features for representation. To ensure that the learned features by I-LSPFC are optimal for classification, we further combine the classification error with the feature coding error to form a unified model, discriminative LSPFC (D-LSPFC), to boost performance. The model of D-LSPFC seamlessly integrates feature coding and discriminative classification, so the representation and classification powers can be enhanced. The proposed approaches are more general, and several recent existing low-rank or sparse coding algorithms can be embedded into our problems as special cases. Visual and numerical results demonstrate the effectiveness of our methods for representation and classification.

<table>
<thead>
<tr>
<th>IM16NXT22</th>
<th>TITLE: Image inpainting through neural networks hallucinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: We consider in this paper the problem of image inpainting, where the objective is to reconstruct large continuous regions of missing or deteriorated parts of an image. Traditional in-painting algorithms are unfortunately not well adapted to handle such corruptions as they rely on image processing techniques that cannot properly infer missing information when the corrupted holes are too large. To tackle this problem, we propose a novel approach where we rely on the hallucinations of pre-trained neural networks to fill large holes in images. To generate globally coherent images, we further impose smoothness and consistency regularization, thereby constraining the neural network hallucinations. Through illustrative experiments, we show that pre-trained neural networks contain crucial prior information that can effectively guide the reconstruction process of complex inpainting problems.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IM16NXT23</th>
<th>TITLE: Classification of image degradation using Riesz transform</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: This paper presents new method for classification of type of image degradation based on the Riesz transform and BRISQUE no-reference quality measure. Riesz transform has great properties and it can be used in many</td>
<td></td>
</tr>
</tbody>
</table>
applications. Some of its benefits are: the ability to construct family of steerable wavelets with arbitrary order and any number of dimensions and it can bring the algorithm of filter banks with perfect reconstruction and also go to dimensions higher than two. Statistical properties of MSCN coefficients used by BRISQUE change in presence of distortion and by quantifying this changes with features calculated by using GGD and AGGD model the class of distortion can be determined. We calculated 18 statistical features out of spatial coefficients defined by BRISQUE measure and 19 parameters out of Riesz coefficients to get 37 features in total and then used features as input in SVM regressor in order to identify the type of image degradation. Then, we compared new method with BRISQUE method by using McNemar's statistical test to show statistical significance of our method.

<table>
<thead>
<tr>
<th>IM16NXT24</th>
<th>TITLE: Convolutional sparse representations as an image model for impulse noise restoration</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: Standard sparse representations, applied independently to a set of overlapping image blocks, are a very effective approach to a wide variety of image reconstruction problems. Convolutional sparse representations, which provide a single-valued representation optimised over an entire image, provide an alternative form of sparse representation that has recently started to attract interest for image reconstruction problems. The present paper provides some insight into the suitability of the convolutional form for this type of application by comparing its performance as an image model with that of the standard model in an impulse noise restoration problem.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IM16NXT25</th>
<th>TITLE: Optic disc localization in fundus images</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: Fundus images are suitable for automatic evaluation in the current clinical practice. The detection and localization of important areas such as the vascular bed, yellow spot (fovea) and optical disc are necessary for following pathological changes detection. After image preprocessing we use very rapid method based on gradients called Fast radial symmetry transform - FRST to find...</td>
<td></td>
</tr>
</tbody>
</table>
approximately circular (elliptical) shape of optic disc. We compare the results achieved by this method with results of two other detecting methods for circular to elliptical objects, Hough transform and Histogram method using template.

CLOUD COMPUTING

CC16NXT01  TITLE: Dynamic and Public Auditing with Fair Arbitration for Cloud Data
ABSTRACT: Cloud users no longer physically possess their data, so how to ensure the integrity of their outsourced data becomes a challenging task. Recently proposed schemes such as “provable data possession” and “proofs of retrievability” are designed to address this problem, but they are designed to audit static archive data and therefore lack of data dynamics support. Moreover, threat models in these schemes usually assume an honest data owner and focus on detecting a dishonest cloud service provider despite the fact that clients may also misbehave. This paper proposes a public auditing scheme with data dynamics support and fairness arbitration of potential disputes. In particular, we design an index switcher to eliminate the limitation of index usage in tag computation in current schemes and achieve efficient handling of data dynamics. To address the fairness problem so that no party can misbehave without being detected, we further extend existing threat models and adopt signature exchange idea to design fair arbitration protocols, so that any possible dispute can be fairly settled. The security analysis shows our scheme is provably secure, and the performance evaluation demonstrates the overhead of data dynamics and dispute arbitration are reasonable.

CC16NXT02  TITLE: Enabling Cloud Storage Auditing with Verifiable Outsourcing of Key Updates
ABSTRACT: Key-exposure resistance has always been an important issue for in-depth cyber defence in many security applications. Recently, how to deal with the key exposure problem in the settings of cloud storage auditing has been proposed and studied. To address the challenge, existing solutions all require the client to update his secret keys in every time period, which may inevitably bring in new local burdens to the client, especially those with limited computation
resources, such as mobile phones. In this paper, we focus on how to make the key updates as transparent as possible for the client and propose a new paradigm called cloud storage auditing with verifiable outsourcing of key updates. In this paradigm, key updates can be safely outsourced to some authorized party, and thus the key-update burden on the client will be kept minimal. In particular, we leverage the third party auditor (TPA) in many existing public auditing designs, let it play the role of authorized party in our case, and make it in charge of both the storage auditing and the secure key updates for key-exposure resistance. In our design, TPA only needs to hold an encrypted version of the client's secret key while doing all these burdensome tasks on behalf of the client. The client only needs to download the encrypted secret key from the TPA when uploading new files to cloud. Besides, our design also equips the client with capability to further verify the validity of the encrypted secret keys provided by the TPA. All these salient features are carefully designed to make the whole auditing procedure with key exposure resistance as transparent as possible for the client. We formalize the definition and the security model of this paradigm. The security proof and the performance simulation show that our detailed design instantiations are secure and efficient.

**CC16NXT03**

**TITLE:** Providing User Security Guarantees in Public Infrastructure Clouds

**ABSTRACT:** The infrastructure cloud (IaaS) service model offers improved resource flexibility and availability, where tenants – insulated from the minutiae of hardware maintenance – rent computing resources to deploy and operate complex systems. Large-scale services running on IaaS platforms demonstrate the viability of this model; nevertheless, many organizations operating on sensitive data avoid migrating operations to IaaS platforms due to security concerns. In this paper, we describe a framework for data and operation security in IaaS, consisting of protocols for a trusted launch of virtual machines and domain-based storage protection. We continue with an extensive theoretical analysis with proofs about protocol resistance against attacks in the defined threat model. The protocols allow trust to be established by remotely attesting
host platform configuration prior to launching guest virtual machines and ensure confidentiality of data in remote storage, with encryption keys maintained outside of the IaaS domain. Presented experimental results demonstrate the validity and efficiency of the proposed protocols. The framework prototype was implemented on a test bed operating a public electronic health record system, showing that the proposed protocols can be integrated into existing cloud environments.

<table>
<thead>
<tr>
<th>CC16NXT04</th>
<th>TITLE: Attribute-Based Data Sharing Scheme Revisited in Cloud Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> Cipher-text-policy attribute-based encryption (CP-ABE) is a very promising encryption technique for secure data sharing in the context of cloud computing. Data owner is allowed to fully control the access policy associated with his data which to be shared. However, CP-ABE is limited to a potential security risk that is known as key escrow problem, whereby the secret keys of users have to be issued by a trusted key authority. Besides, most of the existing CP-ABE schemes cannot support attribute with arbitrary state. In this paper, we revisit attribute-based data sharing scheme in order to solve the key escrow issue but also improve the expressiveness of attribute, so that the resulting scheme is more friendly to cloud computing applications. We propose an improved two-party key issuing protocol that can guarantee that neither key authority nor cloud service provider can compromise the whole secret key of a user individually. Moreover, we introduce the concept of attribute with weight, being provided to enhance the expression of attribute, which can not only extend the expression from binary to arbitrary state, but also lighten the complexity of access policy. Therefore, both storage cost and encryption complexity for a ciphertext are relieved. The performance analysis and the security proof show that the proposed scheme is able to achieve efficient and secure data sharing in cloud computing.</td>
</tr>
</tbody>
</table>
| CC16NXT05 | TITLE: An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing  
ABSTRACT: Ciphertext-policy attribute-based encryption (CP-ABE) has been a preferred encryption technology to solve the challenging problem of secure data sharing in cloud computing. The shared data files generally have the characteristic of multilevel hierarchy, particularly in the area of healthcare and the military. However, the hierarchy structure of shared files has not been explored in CP-ABE. In this paper, an efficient file hierarchy attribute-based encryption scheme is proposed in cloud computing. The layered access structures are integrated into a single access structure, and then, the hierarchical files are encrypted with the integrated access structure. The ciphertext components related to attributes could be shared by the files. Therefore, both ciphertext storage and time cost of encryption are saved. Moreover, the proposed scheme is proved to be secure under the standard assumption. Experimental simulation shows that the proposed scheme is highly efficient in terms of encryption and decryption. With the number of the files increasing, the advantages of our scheme become more and more conspicuous. |
| CC16NXT06 | TITLE: Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity Checking in Public Cloud  
ABSTRACT: More and more clients would like to store their data to public cloud servers (PCSs) along with the rapid development of cloud computing. New security problems have to be solved in order to help more clients process their data in public cloud. When the client is restricted to access PCS, he will delegate its proxy to process his data and upload them. On the other hand, remote data integrity checking is also an important security problem in public cloud storage. It makes the clients check whether their outsourced data are kept intact without downloading the whole data. From the security problems, we propose a novel proxy-oriented data uploading and remote data integrity checking model in identity-based public key cryptography: identity-based proxy-oriented data uploading and remote data integrity checking in public cloud (ID-PUIC). We give |
the formal definition, system model, and security model. Then, a concrete ID-PUIC protocol is designed using the bilinear pairings. The proposed ID-PUIC protocol is provably secure based on the hardness of computational Diffie-Hellman problem. Our ID-PUIC protocol is also efficient and flexible. Based on the original client's authorization, the proposed ID-PUIC protocol can realize private remote data integrity checking, delegated remote data integrity checking, and public remote data integrity checking.

<table>
<thead>
<tr>
<th>CC16NXT07</th>
<th>TITLE: Outsourcing Eigen-Decomposition and Singular Value Decomposition of Large Matrix to a Public Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT:</td>
<td>Cloud computing enables customers with limited computational resources to outsource their huge computation workloads to the cloud with massive computational power. However, in order to utilize this computing paradigm, it presents various challenges that need to be addressed, especially security. As eigen-decomposition (ED) and singular value decomposition (SVD) of a matrix are widely applied in engineering tasks, we are motivated to design secure, correct, and efficient protocols for outsourcing the ED and SVD of a matrix to a malicious cloud in this paper. In order to achieve security, we employ efficient privacy-preserving transformations to protect both the input and output privacy. In order to check the correctness of the result returned from the cloud, an efficient verification algorithm is employed. A computational complexity analysis shows that our protocols are highly efficient. We also introduce an outsourcing principle component analysis as an application of our two proposed protocols.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CC16NXT08</th>
<th>TITLE: Performance Limitations of A Text Search Application Running in Cloud Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT:</td>
<td>This article analyzes the performance of MySQL in clouds based on commodity hardware in order to identify the bottlenecks in the execution of series of scripts developed on the SQL standard. The developed scripts were designed in order to perform text search in a considerable amount of records. Two types of platforms were employed: a physical machine that serves as host...</td>
</tr>
</tbody>
</table>
and an instance within a cloud infrastructure. The results show that the intensive use of a relational database presents a greater loss of performance in a cloud instance due limitations in the primary storage system that was employed in the cloud infrastructure.

| CC16NXT09 | TITLE: A Dynamic Load Balancing Method of Cloud-Center Based on SDN  
ABSTRACT: In order to achieve dynamic load balancing based on data flow level, in this paper, we apply SDN technology to the cloud data center, and propose a dynamic load balancing method of cloud center based on SDN. The approach of using the SDN technology in the current task scheduling flexibility, accomplish real-time monitoring of the service node flow and load condition by the OpenFlow protocol. When the load of system is imbalanced, the controller can allocate globally network resources. What's more, by using dynamic correction, the load of the system is not obvious tilt in the long run. The results of simulation show that this approach can realize and ensure that the load will not tilt over a long period of time, and improve the system throughput. |

| CC16NXT10 | TITLE: Attribute-Based Access Control for Multi-Authority Systems with Constant Size Ciphertext in Cloud Computing  
ABSTRACT: In most existing CP-ABE schemes, there is only one authority in the system and all the public keys and private keys are issued by this authority, which incurs ciphertext size and computation costs in the encryption and decryption operations that depend at least linearly on the number of attributes involved in the access policy. We propose an efficient multi-authority CP-ABE scheme in which the authorities need not interact to generate public information during the system initialization phase. Our scheme has constant ciphertext length and a constant number of pairing computations. Our scheme can be proven CPA-secure in random oracle model under the decision q-BDHE assumption. When user's attributes revocation occurs, the scheme transfers most re-encryption work to the cloud service provider, reducing the data owner's computational cost on the premise of security. Finally the analysis and simulation result show that the schemes proposed in this thesis ensure the |
privacy and secure access of sensitive data stored in the cloud server, and be able to cope with the dynamic changes of users' access privileges in large-scale systems. Besides, the multi-authority ABE eliminates the key escrow problem, achieves the length of ciphertext optimization and enhances the efficiency of the encryption and decryption operations.

<table>
<thead>
<tr>
<th>Title</th>
<th>Dynamic Certification of Cloud Services: Trust, but Verify!</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong></td>
<td>Although intended to ensure cloud service providers' security, reliability, and legal compliance, current cloud service certifications are quickly outdated. Dynamic certification, on the other hand, provides automated monitoring and auditing to verify cloud service providers' ongoing adherence to certification requirements.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title</th>
<th>Auditing a Cloud Provider’s Compliance With Data Backup Requirements: A Game Theoretical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong></td>
<td>The new developments in cloud computing have introduced significant security challenges to guarantee the confidentiality, integrity, and availability of outsourced data. A service level agreement (SLA) is usually signed between the cloud provider (CP) and the customer. For redundancy purposes, it is important to verify the CP’s compliance with data backup requirements in the SLA. There exist a number of security mechanisms to check the integrity and availability of outsourced data. This task can be performed by the customer or be delegated to an independent entity that we will refer to as the verifier. However, checking the availability of data introduces extra costs, which can discourage the customer of performing data verification too often. The interaction between the verifier and the CP can be captured using game theory in order to find an optimal data verification strategy. In this paper, we formulate this problem as a two player non-cooperative game. We consider the case in which each type of data is replicated a number of times, which can depend on a set of parameters including, among others, its size and sensitivity. We analyze the strategies of the CP and the verifier at the Nash equilibrium and derive the expected behavior of both the players. Finally, we validate our model.</td>
</tr>
</tbody>
</table>
numerically on a case study and explain how we evaluate the parameters in the model.

| CC16NXT13 | TITLE: An Efficient Privacy-Preserving Ranked Keyword Search Method  
ABSTRACT: Cloud data owners prefer to outsource documents in an encrypted form for the purpose of privacy preserving. Therefore it is essential to develop efficient and reliable ciphertext search techniques. One challenge is that the relationship between documents will be normally concealed in the process of encryption, which will lead to significant search accuracy performance degradation. Also the volume of data in data centers has experienced a dramatic growth. This will make it even more challenging to design ciphertext search schemes that can provide efficient and reliable online information retrieval on large volume of encrypted data. In this paper, a hierarchical clustering method is proposed to support more search semantics and also to meet the demand for fast ciphertext search within a big data environment. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed in this paper. Experiments have been conducted using the collection set built from the IEEE Xplore. The results show that with a sharp increase of documents in the dataset the search time of the proposed method increases linearly whereas the search time of the traditional method increases exponentially. Furthermore, the proposed method has an advantage over the traditional method in the rank privacy and relevance of retrieved documents. |

| CC16NXT14 | TITLE: Towards Building Forensics Enabled Cloud Through Secure Logging-as-a-Service  
ABSTRACT: Collection and analysis of various logs (e.g., process logs, network logs) are fundamental activities in computer forensics. Ensuring the security of |
the activity logs is therefore crucial to ensure reliable forensics investigations. However, because of the black-box nature of clouds and the volatility and co-mingling of cloud data, providing the cloud logs to investigators while preserving users' privacy and the integrity of logs is challenging. The current secure logging schemes, which consider the logger as trusted cannot be applied in clouds since there is a chance that cloud providers (logger) collude with malicious users or investigators to alter the logs. In this paper, we analyze the threats on cloud users' activity logs considering the collusion between cloud users, providers, and investigators. Based on the threat model, we propose Secure-Logging-as-a-Service (SecLaaS), which preserves various logs generated for the activity of virtual machines running in clouds and ensures the confidentiality and integrity of such logs. Investigators or the court authority can only access these logs by the RESTful APIs provided by SecLaaS, which ensures confidentiality of logs. The integrity of the logs is ensured by hash-chain scheme and proofs of past logs published periodically by the cloud providers. In prior research, we used two accumulator schemes Bloom filter and RSA accumulator to build the proofs of past logs. In this paper, we propose a new accumulator scheme - Bloom-Tree, which performs better than the other two accumulators in terms of time and space requirement.

CC16NXT15

A Genetic Algorithm for Virtual Machine Migration in Heterogeneous Mobile Cloud Computing

ABSTRACT: Mobile Cloud Computing (MCC) improves the performance of a mobile application by executing it at a resourceful cloud server that can minimize execution time compared to a resource-constrained mobile device. Virtual Machine (VM) migration in MCC brings cloud resources closer to a user so as to further minimize the response time of an offloaded application. Such resource migration is very effective for interactive and real-time applications. However, the key challenge is to find an optimal cloud server for migration that offers the maximum reduction in computation time. In this paper, we propose a Genetic Algorithm (GA) based VM migration model, namely GAVMM, for heterogeneous
MCC system. In GAVMM, we take user mobility and load of the cloud servers into consideration to optimize the effectiveness of VM migration. The goal of GAVMM is to select the optimal cloud server for a mobile VM and to minimize the total number of VM migrations, resulting in a reduced task execution time. Additionally, we present a thorough numerical evaluation to investigate the effectiveness of our proposed model compared to the state-of-the-art VM migration policies.

**CC16NXT16**

**TITLE:** Resource Scheduling Under Concave Pricing for Cloud Computing  
**ABSTRACT:** With the booming growth of cloud computing industry, computational resources are readily and elastically available to the customers. In order to attract customers with various demands, most Infrastructure-as-a-service (IaaS) cloud service providers offer several pricing strategies such as pay as you go, pay less per unit when you use more (so called volume discount), and pay even less when you reserve. The diverse pricing schemes among different IaaS service providers or even in the same provider form a complex economic landscape that nurtures the market of cloud brokers. By strategically scheduling multiple customers' resource requests, a cloud broker can fully take advantage of the discounts offered by cloud service providers. In this paper, we focus on how a broker may help a group of customers to fully utilize the volume discount pricing strategy offered by cloud service providers through cost-efficient online resource scheduling. We present a randomized online stack-centric scheduling algorithm (ROSA) and theoretically prove the lower bound of its competitive ratio. Our simulation shows that ROSA achieves a competitive ratio close to the theoretical lower bound under a special case cost function. Trace driven simulation using Google cluster data demonstrates that ROSA is superior to the conventional online scheduling algorithms in terms of cost saving.

**CC16NXT17**

**TITLE:** Dynamic Bin Packing for On-Demand Cloud Resource Allocation  
**ABSTRACT:** Dynamic Bin Packing (DBP) is a variant of classical bin packing, which assumes that items may arrive and depart at arbitrary times. Existing works on DBP generally aim to minimize the maximum number of bins ever used in the
packing. In this paper, we consider a new version of the DBP problem, namely, the MinTotal DBP problem which targets at minimizing the total cost of the bins used overtime. It is motivated by the request dispatching problem arising from cloud gaming systems. We analyze the competitive ratios of the modified versions of the commonly used First Fit, Best Fit, and Any Fit packing (the family of packing algorithms that open a new bin only when no currently open bin can accommodate the item to be packed) algorithms for the Min Total DBP problem. We show that the competitive ratio of Any Fit packing cannot be better than $\mu + 1$, where $\mu$ is the ratio of the maximum item duration to the minimum item duration. The competitive ratio of Best Fit packing is not bounded for any given $\mu$. For First Fit packing, if all the item sizes are smaller than $\frac{1}{\beta}$ of the bin capacity ($\beta > 1$ is a constant), the competitive ratio has an upper bound of $\frac{\beta}{\beta - 1} \cdot \mu + 3\beta/\beta - 1 + 1$. For the general case, the competitive ratio of First Fit packing has an upper bound of $2\mu + 7$. We also propose a Hybrid First Fit packing algorithm that can achieve a competitive ratio no larger than $\frac{5}{4} \mu + \frac{19}{4}$ when $\mu$ is not known and can achieve a competitive ratio no larger than $\mu + 5$ when $\mu$ is known.

**CC16NXT18**

**TITLE:** A Scalable Data Chunk Similarity based Compression Approach for Efficient Big Sensing Data Processing on Cloud

**ABSTRACT:** Big sensing data is prevalent in both industry and scientific research applications where the data is generated with high volume and velocity. Cloud computing provides a promising platform for big sensing data processing and storage as it provides a flexible stack of massive computing, storage, and software services in a scalable manner. Current big sensing data processing on Cloud have adopted some data compression techniques. However, due to the high volume and velocity of big sensing data, traditional data compression techniques lack sufficient efficiency and scalability for data processing. Based on specific on-Cloud data compression requirements, we propose a novel scalable data compression approach based on calculating similarity among the partitioned data chunks. Instead of compressing basic data units, the
compression will be conducted over partitioned data chunks. To restore original data sets, some restoration functions and predictions will be designed. Map Reduce is used for algorithm implementation to achieve extra scalability on Cloud. With real world meteorological big sensing data experiments on U-Cloud platform, we demonstrate that the proposed scalable compression approach based on data chunk similarity can significantly improve data compression efficiency with affordable data accuracy loss.

CC16NXT19  

**TITLE:** A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data

**ABSTRACT:** Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF x IDF model are combined in the index construction and query generation. We construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly.
<table>
<thead>
<tr>
<th>Title</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CC16NXT20</strong>&lt;br&gt;TITLE: A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud</td>
<td><strong>ABSTRACT:</strong> 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. First, 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. Second, 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. Third, 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.</td>
</tr>
<tr>
<td><strong>CC16NXT21</strong>&lt;br&gt;TITLE: CDA Generation and Integration for Health Information Exchange Based on Cloud Computing System</td>
<td><strong>ABSTRACT:</strong> Successful deployment of Electronic Health Record helps improve patient safety and quality of care, but it has the prerequisite of interoperability between Health Information Exchange at different hospitals. The Clinical</td>
</tr>
</tbody>
</table>
Document Architecture (CDA) developed by HL7 is a core document standard to ensure such interoperability, and propagation of this document format is critical for interoperability. Unfortunately, hospitals are reluctant to adopt interoperable HIS due to its deployment cost except for in a handful countries. A problem arises even when more hospitals start using the CDA document format because the data scattered in different documents are hard to manage. In this paper, we describe our CDA document generation and integration Open API service based on cloud computing, through which hospitals are enabled to conveniently generate CDA documents without having to purchase proprietary software. Our CDA document integration system integrates multiple CDA documents per patient into a single CDA document and physicians and patients can browse the clinical data in chronological order. Our system of CDA document generation and integration is based on cloud computing and the service is offered in Open API. Developers using different cloud platforms.

**CC16NXT22**

**TITLE:** Circuit Ciphertext-Policy Attribute-Based Hybrid Encryption with Verifiable Delegation in Cloud Computing

**ABSTRACT:** In the cloud, for achieving access control and keeping data confidential, the data owners could adopt attribute-based encryption to encrypt the stored data. Users with limited computing power are however more likely to delegate the mask of the decryption task to the cloud servers to reduce the computing cost. As a result, attribute-based encryption with delegation emerges. Still, there are caveats and questions remaining in the previous relevant works. For instance, during the delegation, the cloud servers could tamper or replace the delegated cipher text and respond a forged computing result with malicious intent. They may also cheat the eligible users by responding them that they are ineligible for the purpose of cost saving. Furthermore, during the encryption, the access policies may not be flexible enough as well. Since policy for general circuits enables to achieve the strongest form of access control, a construction for realizing circuit cipher text-policy attribute-based hybrid encryption with verifiable delegation has been considered in our work. In such a system,
combined with verifiable computation and encrypt-then-mac mechanism, the data confidentiality, the fine-grained access control and the correctness of the delegated computing results are well guaranteed at the same time. Besides, our scheme achieves security against chosen-plaintext attacks under the k-multi linear Decisional Diffie-Hellman assumption. Moreover, an extensive simulation campaign confirms the feasibility and efficiency of the proposed solution.

CC16NXT23

**TITLE:** CloudArmor: Supporting Reputation-Based Trust Management for Cloud Services

**ABSTRACT:** Trust management is one of the most challenging issues for the adoption and growth of cloud computing. The highly dynamic, distributed, and non-transparent nature of cloud services introduces several challenging issues such as privacy, security, and availability. Preserving consumers' privacy is not an easy task due to the sensitive information involved in the interactions between consumers and the trust management service. Protecting cloud services against their malicious users (e.g., such users might give misleading feedback to disadvantage a particular cloud service) is a difficult problem. Guaranteeing the availability of the trust management service is another significant challenge because of the dynamic nature of cloud environments. In this article, we describe the design and implementation of Cloud Armor, a reputation-based trust management framework that provides a set of functionalities to deliver trust as a service (TaaS), which includes i) a novel protocol to prove the credibility of trust feedbacks and preserve users' privacy, ii) an adaptive and robust credibility model for measuring the credibility of trust feedbacks to protect cloud services from malicious users and to compare the trustworthiness of cloud services, and iii) an availability model to manage the availability of the decentralized implementation of the trust management service. The feasibility and benefits of our approach have been validated by a prototype and experimental studies using a collection of real-world trust feedbacks on cloud services.
TITLE: Conditional Identity-Based Broadcast Proxy Re-Encryption and Its Application to Cloud Email

ABSTRACT: Recently, a number of extended Proxy Re-Encryptions (PRE), e.g. Conditional (CPRE), identity-based PRE (IPRE) and broadcast PRE (BPRE), have been proposed for flexible applications. By incorporating CPRE, IPRE and BPRE, this paper proposes a versatile primitive referred to as conditional identity-based broadcast PRE (CIBPRE) and formalizes its semantic security. CIBPRE allows a sender to encrypt a message to multiple receivers by specifying these receivers' identities, and the sender can delegate a re-encryption key to a proxy so that he can convert the initial cipher text into a new one to a new set of intended receivers. Moreover, the re-encryption key can be associated with a condition such that only the matching cipher texts can be re-encrypted, which allows the original sender to enforce access control over his remote cipher texts in a fine-grained manner. We propose an efficient CIBPRE scheme with provable security. In the instantiated scheme, the initial cipher text, the re-encrypted cipher text and the re-encryption key are all in constant size, and the parameters to generate a re-encryption key are independent of the original receivers of any initial cipher text. Finally, we show an application of our CIBPRE to secure cloud email system advantageous over existing secure email systems based on Pretty Good Privacy protocol or identity-based encryption.

TITLE: Federated Campus Cloud Colombian Initiative

ABSTRACT: Desktop cloud paradigm arises from combining cloud computing with volunteer computing systems in order to harvest the idle computational resources of volunteers' computers. Students usually underuse university computer rooms. As a result, a desktop cloud can be seen as a form of high performance computing (HPC) at a low cost. When the capacity of a desktop cloud is insufficient to execute a HPC project, a new opportunity for collaborative work among universities appears, resulting in a federation of desktop cloud systems to create a significant amount of virtual resources from multiple
providers on non-dedicated infrastructure. Even though cloud federation generates research activity today, neither interoperability among several implementations of cloud computing nor the federation of desktop clouds are resolved issues. Therefore, our initiative is related to gathering the existing and idle computer resources provided by the universities that take part to form a cloud federation on non-dedicated infrastructure.

### MOBILE COMPUTING

<table>
<thead>
<tr>
<th>MC16NXT01</th>
<th>TITLE: Grasping Popular Applications in Cellular Networks with Big Data Analytics Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> Internet access through cellular networks is rapidly growing, driven by the great success of the mobile apps paradigm and the overwhelming popularity of social-related multimedia services such as YouTube, Facebook, or even WhatsApp. Understanding the functioning, performance and traffic generated by these applications is paramount for ISPs, especially for cellular operators, who must manage the huge surge of volume and number of users with the constraints and challenges of cellular networks. In this paper we study important networking aspects of three popular applications in cellular networks: YouTube, Facebook and WhatsApp. Our evaluations span the Content Delivery Networks (CDNs) hosting these services, their traffic characteristics, and their performance. The analysis is performed on top of real cellular network traffic monitored at the nationwide cellular network of a major European ISP. Due to privacy issues and given the huge amount of data generated by these applications as well as the large number of monitored customers, the analysis has been done in an online fashion, using a customized Big Data Analytics (BDA) platform called DB Stream. We overview DB Stream and discuss other potential solutions currently available for traffic monitoring and analysis of big networking data. To the best of our knowledge, this is the first paper providing a complete analysis of popular services in cellular networks, using BDA platforms.</td>
</tr>
</tbody>
</table>
| MC16NXT02 | TITLE: Computing with Nearby Mobile Devices: a Work Sharing Algorithm for Mobile Edge-Clouds  
ABSTRACT: As mobile devices evolve to be powerful and pervasive computing tools, their usage also continues to increase rapidly. However, mobile device users frequently experience problems when running intensive applications on the device itself, or offloading to remote clouds, due to resource shortage and connectivity issues. Ironically, most users’ environments are saturated with devices with significant computational resources. This paper argues that nearby mobile devices can efficiently be utilized as a crowd-powered resource cloud to complement the remote clouds. Node heterogeneity, unknown worker capability, and dynamism are identified as essential challenges to be addressed when scheduling work among nearby mobile devices. We present a work sharing model, called Honeybee, using an adaptation of the well-known work stealing method to load balance independent jobs among heterogeneous mobile nodes, able to accommodate nodes randomly leaving and joining the system. The overall strategy of Honeybee is to focus on short-term goals, taking advantage of opportunities as they arise, based on the concepts of proactive workers and opportunistic delegator. We evaluate our model using a prototype framework built using Android and implement two applications. We report speedups of up to 4 with seven devices and energy savings up to 71% with eight devices. |
| MC16NXT03 | TITLE: Fake Mask: A Novel Privacy Preserving Approach for Smartphones  
ABSTRACT: Users can enjoy personalized services provided by various context-aware applications that collect users' contexts through sensor-equipped smartphones. Meanwhile, serious privacy concerns arise due to the lack of privacy preservation mechanisms. Currently, most mechanisms apply passive defense policies in which the released contexts from a privacy preservation system are always real, leading to a great probability with which an adversary infers the hidden sensitive contexts about the users. In this paper, we apply a deception policy for privacy preservation and present a novel technique, Fake Mask, in which fake contexts may be released to provably preserve users' privacy. The |
output sequence of contexts by Fake Mask can be accessed by the untrusted context-aware applications or be used to answer queries from those applications. Since the output contexts may be different from the original contexts, an adversary has greater difficulty in inferring the real contexts. Therefore, Fake Mask limits what adversaries can learn from the output sequence of contexts about the user being in sensitive contexts, even if the adversaries are powerful enough to have the knowledge about the system and the temporal correlations among the contexts. The essence of Fake Mask is a privacy checking algorithm which decides whether to release a fake context for the current context of the user. We present a novel privacy checking algorithm and an efficient one to accelerate the privacy checking process. Extensive evaluation experiments on real smartphone context traces of users demonstrate the improved performance of Fake Mask over other approaches.

**MC16NXT04**  
**TITLE:** Understanding Smartphone Sensor and App Data for Enhancing the Security of Secret Questions  
**ABSTRACT:** Many web applications provide secondary authentication methods, i.e., secret questions (or password recovery questions), to reset the account password when a user’s login fails. However, the answers to many such secret questions can be easily guessed by an acquaintance or exposed to a stranger that has access to public online tools (e.g., online social networks); moreover, a user may forget her/his answers long after creating the secret questions. Today’s prevalence of smartphones has granted us new opportunities to observe and understand how the personal data collected by smartphones sensors and apps can help create personalized secret questions without violating the users’ privacy concerns. In this paper, we present a Secret-Question based Authentication system, called “Secret-QA”, that creates a set of secret questions on basic of people’s smartphone usage. We develop a prototype on Android smartphones, and evaluate the security of the secret questions by asking the acquaintance/stranger who participate in our user study to guess the answers with and without the help of online tools; meanwhile, we observe the questions’
reliability by asking participants to answer their own questions. Our experimental results reveal that the secret questions related to motion sensors, calendar, app installment, and part of legacy app usage history (e.g., phone calls) have the best memo ability for users as well as the highest robustness to attacks.

**MC16NXT05**  
**TITLE:** FLANDROID: Energy-Efﬁcient Recommendations of Reliable Context Providers for Android Applications  
**ABSTRACT:** Mobile applications are becoming more and more popular with the prevalence of mobile operating systems and mobile Internet. Many of them consume services provided by the underlying infrastructure and platforms as a part of their application environmental contexts. However, application failures or downgrade in performance may be the results due to inadequate provisions of these environmental issues in the implementations of the mobile applications. In this paper, we propose a framework to enable mobile applications to consume services offered by a reliable context provider with high probability in run time. We report a case study on a suite of five real-world mobile applications with 74 real faults on real mobile phones involving 50 users. The results of the case study show that our framework can significantly improve the reliability of mobile applications with respect to the failures due to buggy-context-provider faults with low slowdown and energy overheads.

**MC16NXT06**  
**TITLE:** Towards a Fully Cloudified Mobile Network Infrastructure  
**ABSTRACT:** Cloud computing enables the on-demand delivery of resources for a multitude of services and gives the opportunity for small agile companies to compete with large industries. In the telco world, cloud computing is currently mostly used by Mobile Network Operators (MNO) for hosting non-critical support services and selling cloud services such as applications and data storage. MNOs are investigating the use of cloud computing to deliver key telecommunication services in the access and core networks. Without this, MNOs lose the opportunities of both combining this with over-the-top (OTT) and value-added services to their fundamental service offerings and leveraging cost-effective commodity hardware. Being able to leverage cloud computing
technology effectively for the telco world is the focus of Mobile Cloud Networking (MCN). This paper presents the key results of MCN integrated project that includes its architecture advancements, prototype implementation and evaluation. Results show the efficiency and the simplicity that a MNO can deploy and manage the complete service lifecycle of fully cloudified, composed services that combine OTT/IT- and mobile-network-based services running on commodity hardware. The extensive performance evaluation of MCN using two key Proof-of-Concept scenarios that compose together many services to deliver novel converged elastic, on-demand mobile-based but innovative OTT services proves the feasibility of such fully virtualized deployments. Results show that it is beneficial to extend cloud computing to telco usage and run fully cloudified mobile-network-based systems with clear advantages and new service opportunities for MNOs and end users.

MC16NXT07

TITLE: Rapid Localization and Extraction of Street Light Poles in Mobile LiDAR Point Clouds: A Supervoxel-Based Approach

ABSTRACT: This paper presents a supervoxel-based approach for automated localization and extraction of street light poles in point clouds acquired by a mobile LiDAR system. The method consists of five steps: preprocessing, localization, segmentation, feature extraction, and classification. First, the raw point clouds are divided into segments along the trajectory, the ground points are removed, and the remaining points are segmented into supervoxels. Then, a robust localization method is proposed to accurately identify the pole-like objects. Next, a localization-guided segmentation method is proposed to obtain pole-like objects. Subsequently, the pole features are classified using the support vector machine and random forests. The proposed approach was evaluated on three datasets with 1,055 street light poles and 701 million points. Experimental results show that our localization method achieved an average recall value of 98.8%. A comparative study proved that our method is more robust and efficient than other existing methods for localization and extraction of street light poles.
| MC16NXT08 | TITLE: Rethinking Permission Enforcement Mechanism on Mobile Systems  
ABSTRACT: To protect sensitive resources from unauthorized use, modern mobile systems, such as Android and iOS, design a permission-based access control model. However, current model could not enforce fine-grained control over the dynamic permission use contexts, causing two severe security problems. First, any code package in an application could use the granted permissions, inducing attackers to embed malicious payloads into benign apps. Second, the permissions granted to a benign application may be utilized by an attacker through vulnerable application interactions. Although ad hoc solutions have been proposed, none could systematically solve these two issues within a unified framework. This paper presents the first such framework to provide context-sensitive permission enforcement that regulates permission use policies according to system-wide application contexts, which cover both intra-application context and inter-application context. We build a prototype system on Android, named Fine Droid, to track such context during the application execution. To flexibly regulate the context-sensitive permission rules, Fine Droid features a policy framework that could express generic application contexts. We demonstrate the benefits of Fine Droid by instantiating several security extensions based on the policy framework, for three potential users: end users, administrators, and developers. Furthermore, Fine Droid is showed to introduce a minor overhead. |
|---|---|
| MC16NXT09 | TITLE: Heating Dispersal for Self-Healing NAND Flash Memory  
ABSTRACT: Substantially reduced lifetimes are becoming a critical issue in NAND flash memory with the advent of multi-level cell and triple-level cell flash memory. Researchers discovered that heating can cause worn-out NAND flash cells to become reusable and greatly extend the lifetime of flash memory cells. However, the heating process consumes a substantial amount of power, and some fundamental changes are required for existing NAND flash management techniques. In particular, all existing wear-leveling techniques are based on the principle of evenly distributing writes and erases. For self-healing NAND flash, |
this may cause NAND flash cells to be worn out in a short period of time. Moreover, frequently healing these cells may drain the energy quickly in battery driven mobile devices, which is defined as the concentrated heating problem. In this paper, we propose a novel wear-leveling scheme called DHeating (Dispersed Heating) to address the problem. In DHeating, rather than evenly distributing writes and erases over a time period, write and erase operations are scheduled on a small number of flash memory cells at a time, so that these cells can be worn out and healed much earlier than other cells. In this way, we can avoid quick energy depletion caused by concentrated heating. In addition, the heating process takes several seconds and has become the new performance bottleneck. In order to address this issue, we propose a lazy heating repair scheme. The lazy heating repair scheme can ease the long time delays caused by the heating via delaying the heating operation and using the system idle time to repair. Furthermore, the flash memory’s reliability becomes worse with the flash memory cells reaching the excepted worn out time. We propose an early heating strategy to solve the reliability problem. With the extended lifetime provided by self-healing, we can trade some lifetimes for reliability. The idea is to start the healing process earlier than the expected worn-out time. We evaluate our scheme based on an embedded platform. The experimental results show that the proposed scheme can effectively prolong the consecutive heating time interval, alleviate the long time delays caused by the heating, and enhance the reliability for self-healing flash memory.

MC16NXT10

**TITLE: Energy-Efficient Target Tracking by Mobile Sensors with Limited Sensing Range**

**ABSTRACT:** The recent advancements of technology in robotics and wireless communication have enabled the low-cost and large-scale deployment of mobile sensor nodes for target tracking which is a critical application scenario of wireless sensor networks. Due to the constraints of limited sensing range, it is of great importance to design node coordination mechanism for reliable tracking so that at least the target can always be detected with a high probability while the
total network energy cost can be reduced for longer network lifetime. In this paper, we deal with this problem considering both the unreliable wireless channel and the network energy constraint. We transfer the original problem into a dynamic coverage problem, and decompose it into two sub problems. By exploiting the online estimate of target location, we first decide the locations where the mobile nodes should move into so that the reliable tracking can be guaranteed. Then, we assign different nodes to each location in order that the total energy cost in terms of moving distance can be minimized. Extensive simulations under various system settings are employed to evaluate the effectiveness of our solution.

| MC16NXT11 | TITLE: Fusion of Modified Bat Algorithm Soft Computing and Dynamic Model Hard Computing to Online Self-Adaptive Fuzzy Control of Autonomous Mobile Robots  
ABSTRACT: This paper presents a fusion methodology of modified bat algorithm (BA) soft computing and dynamic model hard computing to online self-adaptive fuzzy control of autonomous mobile robots in a field-programmable gate array (FPGA) chip. This fusion approach, called fusion of BA fuzzy control (FBAFC), gains the benefits of BA, online fuzzy control, dynamic model, Taguchi method, and FPGA technique. The FPGA realization of the proposed FBAFC fusion method is more effective in practice for challenging real-world embedded applications by using system-on-a-programmable-chip (SOPC) technology. Experimental results and comparative works are conducted to illustrate the merits of the proposed SOPC-based FBAFC optimal controller over other existing methods for mobile robots. |

| MC16NXT12 | TITLE: Characterizing and Modeling User Behavior in a Large-scale Mobile Live Streaming System  
ABSTRACT: In mobile live streaming systems, user are fairly limited in interaction with the streaming objects due to the constraints coming from mobile devices and event-driven nature of live content. The constraints could lead to unique user behavior characteristics, which have yet to be explored. This |
paper investigates over 9 million access logs collected from the PPTV live streaming system, with an emphasis on the discrepancies that might exist when users access the live streaming catalog from mobile and non-mobile terminals. We observe a much higher likelihood of abandoning sessions by mobile users, and examine the structure of abandoned sessions from the perspectives of time of day, channel content and mobile device types. Surprisingly, we find relatively low abandonment rates during peak-load time periods and a notable impact of mobile device type (i.e. Android or iOS) on the abandonment behavior. To further capture the intrinsic characteristics of user behavior, we develop a series of models for session duration, user activity and time-dynamics of user arrivals/departures. More importantly, we relate the model parameters to physical and real-life meanings. The observations and models shed light on video delivery system, telco-CDNs and mobile applications.

MC16NXT13

TITLE: Assessing the Implications of Cellular Network Performance on Mobile Content Access

ABSTRACT: Mobile applications such as VoIP, (live) gaming, or video streaming have diverse QoS requirements ranging from low delay to high throughput. The optimization of the network quality experienced by end-users requires detailed knowledge of the expected network performance. Also, the achieved service quality is affected by a number of factors, including network operator and available technologies. However, most studies measuring the cellular network do not consider the performance implications of network configuration and management. To this end, this paper reports about an extensive data set of cellular network measurements, focused on analyzing root causes of mobile network performance variability. Measurements conducted on a 4G cellular network in Germany show that management and configuration decisions have a substantial impact on the performance. Specifically, it is observed that the association of mobile devices to a point of presence (POP) within the operator's network can influence the end-to-end performance by a large extent. Given the collected data, a model predicting the POP assignment and its resulting RTT
leveraging Markov chain and machine learning approaches is developed. RTT increases of 58% to 73% compared to the optimum performance are observed in more than 57% of the measurements. Measurements of the response and page load times of popular websites lead to similar results, namely, a median increase of 40% between the worst and the best performing POP.

| MC16NXT14 | **TITLE:** Magnetic Tensor Sensor and Way-finding Method based on Geomagnetic Field Effects with Applications for Visually Impaired Users  
**ABSTRACT:** This paper presents a method utilizing geomagnetic field effects commonly found in nature to help the visually impaired persons (VIPs) navigate safely and efficiently; both indoor and outdoor applications are considered. Magnetic information indicating special locations can be incorporated as waypoints on a map to provide a basis to help the user follow a map that amalgamates the waypoints into spatial information. Along with a magnetic tensor sensor (MTS), a navigation system for helping VIPs more effectively comprehend their surroundings is presented. With the waypoint-enhanced map and an improved dynamic time warping algorithm, this system estimates the user’s locations from real-time measured magnetic data. Methods using image data to enhance waypoints at dangerous locations are discussed. The MTS-enhanced method can be integrated into existing personal mobile devices (with built-in sound, image, video and vibration alert capabilities) to take advantages of the rapidly developing internet, global positioning systems (GPS) and computing technologies to overcome several shortcomings of blind-assistive devices. A prototype MTS-enhanced system for indoor/outdoor navigation has been developed and demonstrated experimentally. Although the MTS and algorithm are presented in the context of way-finding for a VIP, the findings presented here provide a basis for a wide range of applications where geomagnetic field effects offer an advantage. |

| MC16NXT15 | **TITLE:** Declarative Framework for Specification, Simulation and Analysis of Distributed Applications  
**ABSTRACT:** Researchers have recently shown that declarative database query
languages, such as Data log, could naturally be used to specify and implement network protocols and services. In this paper, we present a declarative framework for the specification, execution, simulation, and analysis of distributed applications. Distributed applications, including routing protocols, can be specified using a Declarative Networking language, called D2C, whose semantics capture the notion of a Distributed State Machine (DSM), i.e., a network of computational nodes that communicate with each other through the exchange of data. The D2C specification can be directly executed using the DSM computational infrastructure of our framework. The same specification can be simulated and formally verified. The simulation component integrates the DSM tool within a network simulation environment and allows developers to simulate network dynamics and collect data about the execution in order to evaluate application responses to network changes. The formal analysis component of our framework, instead, complements the empirical testing by supporting the verification of different classes of properties of distributed algorithms, including convergence of network routing protocols. To demonstrate the generality of our framework, we show how it can be used to analyze two classes of network routing protocols, a path vector and a Mobile Ad-Hoc Network (MANET) routing protocol, and execute a distributed algorithm for pattern formation in multi-robot systems.

**MC16NXT16**

**TITLE:** Complexity Reduction by Modified Scale-Space Construction in SIFT Generation Optimized for a Mobile GPU

**ABSTRACT:** Scale-invariant feature transform (SIFT) is one of the most widely used local features for computer vision in mobile devices. A mobile GPU is often used to run computer-vision applications using SIFT features, but the performance in such a case is not powerful enough to generate SIFT features in real time. This paper proposes an efficient scheme to optimize the SIFT algorithm for a mobile GPU. It analyzes the conventional scale-space construction step in the SIFT generation, finding that reducing the size of the Gaussian filter and the scale-space image leads to a significant speed-up with only a slight degradation.
of the quality of the features. Based on this observation, the SIFT algorithm is modified and implemented for real-time execution. Additional optimization techniques are employed for a further speed-up by efficiently utilizing both the CPU and the GPU in a mobile processor. The proposed SIFT generation scheme achieves a processing speed of 28.30 fps for an image with a resolution of 1280x720 running on a Galaxy S5 LTE-A device, thereby gaining a speed-up by factors of 114.78 and 4.53 over CPU- and GPU-only implementations, respectively.

MC16NXT17  

**TITLE:** Experimental Evaluation of Impulsive Ultrasonic Intra-Body Communications for Implantable Biomedical Devices  

**ABSTRACT:** Biomedical systems of miniaturized implantable sensors and actuators interconnected in an intra-body area network could enable revolutionary clinical applications. Given the well understood limitations of radio frequency (RF) propagation in the human body, in our previous work we investigated the use of ultrasonic waves as an alternative physical carrier of information, and proposed Ultrasonic Wideband (UsWB), an ultrasonic multipath-resilient integrated physical and medium access control (MAC) layer protocol. In this paper, we discuss the design and implementation of a software-defined test bed architecture for ultrasonic intra-body area networks, and propose the first experimental demonstration of the feasibility of ultrasonic communications in tissue mimicking materials. We first discuss in detail our FPGA-based prototype implementation of UsWB. We then demonstrate how the prototype can flexibly trade performance off for power consumption, and achieve, for bit error rates (BER) no higher than $10^{-6}$, either (i) high-data rate transmissions up to 700kbit/s at a transmit power of -14dBm (40 W), or (ii) low-data rate and lower-power transmissions down to -21dBm (8W) at 70kbit/s. We demonstrate that the UsWB MAC protocol allows multiple transmitter-receiver pairs to coexist and dynamically adapt the transmission rate according to channel and interference conditions to maximize throughput while satisfying predefined reliability constraints. We also show how UsWB can be used to
| MC16NXT18 | **TITLE:** On Energy-Efficient Offloading in Mobile Cloud for Real Time Video Applications  
**ABSTRACT:** Batteries of modern mobile devices remain severely limited in capacity, which makes energy consumption a key concern for mobile applications, particularly for the computation intensive video applications. Mobile devices can save energy by offloading computation tasks to the cloud; yet the energy gain must exceed the additional communication cost for cloud migration to be beneficial. The situation is further complicated with real-time video applications that have stringent delay and bandwidth constraints. In this paper, we closely examine the performance and energy efficiency of representative mobile cloud applications under dynamic wireless network channels and state of the-art mobile platforms. We identify the unique challenges and opportunities for offloading real time video applications, and develop a generic model for energy-efficient computation offloading accordingly in this context. We propose a scheduling algorithm that makes adaptive offloading decisions in fine granularity in dynamic wireless network conditions, and verify its effectiveness through trace-driven simulations. We further present case studies with advanced mobile platforms and practical applications to demonstrate the superiority of our solution and the substantial gain of our approach over baseline approaches. |
|-----------------|---------------------------------|

| MC16NXT19 | **TITLE:** The Design of a Wearable RFID System for Real-time Activity Recognition using Radio Patterns  
**ABSTRACT:** Elderly care is one of the many applications supported by real-time activity recognition systems. Traditional approaches use cameras, body sensor networks, or radio patterns from various sources for activity recognition. However, these approaches are limited due to ease-of-use, coverage, or privacy preserving issues. In this paper, we present a novel wearable Radio Frequency |
Identification (RFID) system aims at providing an easy-to-use solution with high detection coverage. Our system uses passive tags which are maintenance-free and can be embedded into the clothes to reduce the wearing and maintenance efforts. A small RFID reader is also worn on the user’s body to extend the detection coverage as the user moves. We exploit RFID radio patterns and extract both spatial and temporal features to characterize various activities. We also address the issues of false negative of tag readings and tag/antenna calibration, and design a fast online recognition system. Antenna and tag selection is done automatically to explore the minimum number of devices required to achieve target accuracy. We develop a prototype system which consists of a wearable RFID system and a smart phone to demonstrate the working principles, and conduct experimental studies with four subjects over two weeks. The results show that our system achieves a high recognition accuracy of 93.6% with a latency of 5 seconds. Additionally, we show that the system only requires two antennas and four tagged body parts to achieve a high recognition accuracy of 85%.

MC16NXT20

**TITLE:** Drive Now, Text Later: Nonintrusive Texting-while-Driving Detection using Smart phones

**ABSTRACT:** Texting-while-driving (T and D) is one of the top dangerous behaviors for drivers. Many interesting systems and mobile phone applications have been designed to help to detect or combat T and D. However, for a T and D detection system to be practical, a key property is its capability to distinguish driver’s mobile phone from passengers’. Existing solutions to this problem generally rely on user’s manual input, or utilize specific localization devices to determine whether a mobile phone is at driver’s location. In this paper, we propose a method which is able to detect T and D automatically without using any extra devices. The idea is very simple: when a user is composing messages, the smart phone embedded sensors (i.e. gyroscopes, accelerometers, and GPS) collect the associated information such as touch strokes, holding orientation and vehicle speed. This information will then be analyzed to see whether there exists
some specific T and D patterns. Extensive experiments have been conducted by different persons and in different driving scenarios. The results show that our approach can achieve a good detection accuracy with low false positive rate. Besides being infrastructure-free and with high accuracy, the method does not access the content of messages and therefore is privacy-preserving.

| MC16NXT21 | TITLE: NoPSM: A Concurrent MAC Protocol over Low-data-rate Low-power Wireless Channel without PRR-SINR Model |

**ABSTRACT:** Concurrent MAC protocols can improve channel usage of wireless sensor networks (WSNs), and provide a high-performance infrastructure for data intensive applications. Most of existing concurrent MAC protocols are based on proactively constructed physical interference models, i.e. PRR-SINR models (PSM). However, it incurs relatively high bandwidth and energy overheads to construct PSM for WSNs. In this paper, we propose NoPSM, which does not take PSM as base to determine transmission concurrency. Instead, the base of NoPSM is reactively constructed interference relationships by passively analyzing overlapping relationships among time logs of block data transmissions and corresponding reception status of each packet in blocks. In this way, NoPSM has two salient features. Firstly, NoPSM is able to construct interference relationships among nodes quickly and accurately along with block data transmissions without needs of network downtime. Secondly, based on the constructed interference relationships, NoPSM can make decisions of transmission concurrency with a comprehensive criterion, which not only estimates quality of any active links after initiating a new link, but also estimates throughput improvement gained from concurrent transmissions. NoPSM has been implemented in Tinyos-2.1 and extensively evaluated in TOSSIM. Experimental results show that NoPSM improves system throughput by up to 60% compared with a traditional CSMA protocol, which cannot exploit potential transmission concurrency. Moreover, NoPSM can gain up to 55% throughput improvement as compared to an existing reactive concurrent MAC.
| MC16NXT22 | TITLE: MobiCoRE: Mobile Device based Cloudlet Resource Enhancement for Optimal Task Response  
ABSTRACT: Cloudlets are small self maintained clouds, with hotspot like deployment, to enhance the computational capabilities of the mobile devices. The limited resources of cloudlets can become heavily loaded during peak utilization. Consequently, per user available computational capacity decreases and at times mobile devices find no execution time benefit for using the cloudlet. Researchers have proposed augmenting the cloudlet resources using mobile devices; however, the proposed approaches do not consider the offered service to load ratio while using mobile device resources. In this paper, we propose easy to implement Mobile Device based Cloudlet Resource Enhancement (MobiCoRE) while ensuring that: (i) mobile device always have time benefit for its tasks submitted to the cloudlet and (ii) cloudlet induced mobile device load is a fraction of its own service requirement from the cloudlet. We map MobiCoRE on M/M/c/K queue and model the system using birth death markov chain. Given the arrival rate of , c cpu cores in cloudlet, maximum tasks in the cloudlet to be K and $P_0 = f(c; K)$ be probability of having no user in cloudlet, we derive the condition $1 - P_0 = K - cK - c!K - 1000$ for optimal average service time 1 of cloudlet such that the mobile applications have maximum benefit for using cloudlet services. We show that the optimal average service time is independent of the applications service requirement. Evaluation shows that MobiCoRE can accommodate up to 50% extra users when operating at optimal service time and sharing mobile resources for remaining task, compared to completing the entire user applications in cloudlet. Similarly, up to 47% time benefit can be achieved for mobile devices by sharing only 16% computational resources with the cloudlet. |
| MC16NXT23 | TITLE: Mobile-Edge Computing: Partial Computation Offloading Using Dynamic Voltage Scaling  
ABSTRACT: The incorporation of dynamic voltage scaling (DVS) technology into computation offloading offers more flexibilities for mobile edge computing. In |
this paper, we investigate partial computation offloading by jointly optimizing the computational speed of smart mobile device (SMD), transmit power of SMD, and offloading ratio with two system design objectives: energy consumption of SMD minimization (ECM) and latency of application execution minimization (LM). Considering the case that the SMD is served by a single cloud server, we formulate both the ECM problem and the LM problem as no convex problems. To tackle the ECM problem, we recast it as a convex one with the variable substitution technique and obtain its optimal solution. To address the no convex and no smooth LM problem, we propose a locally optimal algorithm with the univariate search technique. Furthermore, we extend the scenario to a multiple cloud servers system, where the SMD could offload its computation to a set of cloud servers. In this scenario, we obtain the optimal computation distribution among cloud servers in closed form for the ECM and LM problem. Finally, extensive simulations demonstrate that our proposed algorithms can significantly reduce the energy consumption and shorten the latency with respect to the existing offloading schemes.

MC16NXT24

TITLE: Understanding Requirements for Mobile Collaborative Applications in Domains of Use

ABSTRACT: Several initiatives have implemented collaborative applications for mobile settings as diverse as hospital work, wildlife, transportation, and museums. The changing nature of mobile technology has resulted in a wide variety of applications. We explored models, architectures, and applications developed in the past 13 years to categorize the types of existing software and extract a set of common core requirements that support mobile collaboration independently of the current technology. This paper provides an analysis of the domain of mobile collaborative systems including a proposal division into several domains of use, and a study of the types of systems that exist in each of them. In this way, developers can analyze their scenario of development to get an idea of the most important requirements that should be considered for development.
**MC16NXT25**

**TITLE:** Iunius: A Cross Layer Peer-to-peer System with Device-to-device Communications

**ABSTRACT:** Device-to-device (D2D) communications utilizing licensed spectrum have been considered as a promising technology to improve cellular network spectral efficiency and offload local traffic from cellular base stations (BSs). In this paper, we develop Iunius: a peer-to-peer (P2P) system based on harvesting data in a community utilizing multi-hop D2D communications. The Iunius system optimizes D2D communications for P2P local file sharing, improves user experience, and offloads traffic from the BSs. The Iunius system features cross-layer integration of: 1) a wireless P2P protocol based on the Bit torrent protocol in the application layer; 2) a simple centralized routing mechanism for multi-hop D2D communications; 3) an interference cancellation technique for conventional cellular (CC) uplink communications; and 4) a radio resource management scheme to mitigate the interference between CC and D2D communications that share the cellular uplink radio resources while maximizing the throughput of D2D communications. Simulation results show that the proposed Iunius system can increase the cellular spectral efficiency, reduce the traffic load of BSs, and improve the data rate and energy saving for mobile users.

**NETWORK SECURITY**

**NS16NXT01**

**TITLE:** Contributory Broadcast Encryption with Efficient Encryption and Short Ciphertexts

**ABSTRACT:** Broadcast encryption (BE) schemes allow a sender to securely broadcast to any subset of members but require a trusted party to distribute decryption keys. Group key agreement (GKA) protocols enable a group of members to negotiate a common encryption key via open networks so that only the group members can decrypt the cipher texts encrypted under the shared encryption key, but a sender cannot exclude any particular member from decrypting the cipher texts. In this paper, we bridge these two notions with a hybrid primitive referred to as contributory broadcast encryption (ConBE). In this new primitive, a group of members negotiate a common public encryption key
while each member holds a decryption key. A sender seeing the public group encryption key can limit the decryption to a subset of members of his choice. Following this model, we propose a ConBE scheme with short cipher texts. The scheme is proven to be fully collusion.

| NS16NXT02 | TITLE: Building an intrusion detection system using a filter-based feature selection algorithm  
ABSTRACT: Redundant and irrelevant features in data have caused a long-term problem in network traffic classification. These features not only slow down the process of classification but also prevent a classifier from making accurate decisions, especially when coping with big data. In this paper, we propose a mutual information based algorithm that analytically selects the optimal feature for classification. This mutual information based feature selection algorithm can handle linearly and nonlinearly dependent data features. Its effectiveness is evaluated in the cases of network intrusion detection. An Intrusion Detection System (IDS), named Least Square Support Vector Machine based IDS (LSSVM-IDS), is built using the features selected by our proposed feature selection algorithm. The performance of LSSVM-IDS is evaluated using three intrusion detection evaluation datasets, namely KDD Cup 99, NSL-KDD and Kyoto 2006+ dataset. The evaluation results show that our feature selection algorithm contributes more critical features for LSSVM-IDS to achieve better accuracy and lower computational cost compared with the state-of-the-art methods. |

| NS16NXT03 | TITLE: Secure Routing in Multihop Wireless Ad-Hoc Networks With Decode-and-Forward Relaying  
ABSTRACT: In this paper, we study the problem of secure routing in a multihop wireless ad-hoc network in the presence of randomly distributed eavesdroppers. Specifically, the locations of the eavesdroppers are modeled as a homogeneous Poisson point process (PPP) and the source-destination pair is assisted by intermediate relays using the decode-and-forward (DF) strategy. We analytically characterize the physical layer security performance of any chosen multihop |
path using the end-to-end secure connection probability (SCP) for both colluding and noncolluding eavesdroppers. To facilitate finding an efficient solution to secure routing, we derive accurate approximations of the SCP. Based on the SCP approximations, we study the secure routing problem, which is defined as finding the multihop path having the highest SCP. A revised Bellman-Ford algorithm is adopted to find the optimal path in a distributed manner. Simulation results demonstrate that the proposed secure routing scheme achieves nearly the same performance as exhaustive search.

| NS16NXT04 | **TITLE:** Secure Beamforming in Wireless-Powered Cooperative Cognitive Radio Networks  
**ABSTRACT:** In wireless-powered cognitive radio networks, secure information transmission is of paramount importance for the primary system. This letter considers cooperation between a primary system and a wireless-powered secondary system. In particular, we focus on secure information transmission for the primary system when the secondary users (SUs) are the potential eavesdroppers. We aim to jointly design power splitting and secure beamforming to maximize the secondary system data rate subject to the secrecy QoS requirement of the primary system and the secondary transmitter (ST) power constraint. To solve this non convex problem, we propose a two-stage optimization approach. Simulation results demonstrate that our proposed scheme achieves a significant transmission rate of the secondary system while provides a high secrecy rate for the primary system compared to the scheme without energy harvesting. |
| NS16NXT05 | **TITLE:** Novel Design of Secure End-to-End Routing Protocol in Wireless Sensor Networks  
**ABSTRACT:** In wireless sensor networks, the secure end-to-end data communication is needed to collect data from source to destination. Collected data are transmitted in a path consisting of connected links. All existing end-to-end routing protocols propose solutions in which each link uses a pair wise |
shared key to protect data. In this paper, we propose a novel design of secure end-to-end data communication. We adopt a newly published group key pre-distribution scheme in our design, such that there is a unique group key, called path key, to protect data transmitted in the entire routing path. Specifically, instead of using multiple pair wise shared keys to repeatedly perform encryption and decryption over every link, our proposed scheme uses a unique end-to-end path key to protect data transmitted over the path. Our protocol can authenticate sensors to establish the path and to establish the path key. The main advantage using our protocol is to reduce the time needed to process data by intermediate sensors. Moreover, our proposed authentication scheme has complexity $O(n)$, where $n$ is the number of sensors in a communication path, which is different from all existing authentication schemes which are one-to-one authentications with complexity $O(n^2)$. The security of the protocol is computationally secure.

**NS16NXT06**

**TITLE: Secure Transmission Against Pilot Spoofing Attack: A Two-Way Training-Based Scheme**

**ABSTRACT:** The pilot spoofing attack is one kind of active eavesdropping activities conducted by a malicious user during the channel training phase. By transmitting the identical pilot (training) signals as those of the legal users, such an attack is able to manipulate the channel estimation outcome, which may result in a larger channel rate for the adversary but a smaller channel rate for the legitimate receiver. With the intention of detecting the pilot spoofing attack and minimizing its damages, we design a two-way training-based scheme. The effective detector exploits the intrusive component created by the adversary, followed by a secure beam forming-assisted data transmission. In addition to the solid detection performance, this scheme is also capable of obtaining the estimations of both legitimate and illegitimate channels, which allows the users to achieve secure communication in the presence of pilot spoofing attack. The detection probability is evaluated based on the derived test threshold at a given requirement on the probability of false alarming. The achievable secrecy rate is
utilized to measure the security level of the data transmission. Our analysis shows that even without any pre-assumed knowledge of eavesdropper, the proposed scheme is still able to achieve the maximal secrecy rate in certain cases. Numerical results are provided to show that our scheme could achieve a high detection probability as well as secure transmission.

**NS16NXT07**

**TITLE:** Multiantenna Secure Cognitive Radio Networks With Finite-Alphabet Inputs: A Global Optimization Approach for Precoder Design

**ABSTRACT:** This paper considers the precoder design for multiantenna secure cognitive radio networks. We use finite-alphabet inputs as the signaling and exploit statistical channel state information (CSI) at the transmitter. We maximize the secrecy rate of the secondary user and control the transmit power and the power leakage to the primary receivers that share the same frequency spectrum. The secrecy rate maximization is important for practical systems, but challenging to solve, mainly due to two reasons. First, the secrecy rate with statistical CSI is computationally prohibitive to evaluate. Second, the optimization over the precoder is a nondeterministic polynomial-time hard (NP-hard) problem. We utilize an accurate approximation of the secrecy rate to reduce the computational effort and then propose a global optimization approach based on branch-and-bound method. The idea is to define a simplex and transform the secrecy rate into a concave function. The derived concave function converges to the secrecy rate when the defined simplex shrinks down. Using this feature, we solve a sequence of concave maximization problems over iteratively shrinking simplices and eventually attain the globally optimal solution that maximizes the approximation of the secrecy rate. When the complexity is concerned, a low-complexity variant with limited number of iterations can be used in practice. We demonstrate the performance gains when compared with others through numerical examples.
**NS16NXT08**

**TITLE:** Secure Data Aggregation with Homomorphic Primitives in Wireless Sensor Networks: A Critical Survey and Open Research Issues

**ABSTRACT:** In wireless sensor networks, data aggregation plays an important role in energy conservation of sensors. However, these networks are typically deployed in hostile environments and in which privacy and data integrity are widely desired. Because of their design, wireless sensors can be easily captured. Also, nodes that perform the aggregation function are most attractive to attackers. Therefore, in order to deal with these security threats, the research on data aggregation security is essential. In this context, several solutions have been proposed to secure data aggregation in sensor networks, based on several encryption techniques. Among these, there is the homomorphic encryption. Compared to other techniques, in homomorphic encryption all the sensor nodes participate in the aggregation, without seeing any intermediate or final result, while still maintaining an effective and efficient aggregation process. In this paper, we present a survey of some secure data aggregation schemes that use homomorphic encryption properties, and then we compared them based on some criteria. Finally, we present and discuss some open issues that need to be looked in future studies in order to improve the security of data aggregation in wireless sensor networks.

**NS16NXT09**

**TITLE:** Securing SIFT: Privacy-Preserving Outsourcing Computation of Feature Extractions Over Encrypted Image Data

**ABSTRACT:** Advances in cloud computing have greatly motivated data owners to outsource their huge amount of personal multimedia data and/or computationally expensive tasks onto the cloud by leveraging its abundant resources for cost saving and flexibility. Despite the tremendous benefits, the outsourced multimedia data and its originated applications may reveal the data owner's private information, such as the personal identity, locations, or even financial profiles. This observation has recently aroused new research interest on privacy-preserving computations over outsourced multimedia data. In this paper,
we propose an effective and practical privacy-preserving computation outsourcing protocol for the prevailing scale-invariant feature transform (SIFT) over massive encrypted image data. We first show that the previous solutions to this problem have either efficiency/security or practicality issues, and none can well preserve the important characteristics of the original SIFT in terms of distinctiveness and robustness. We then present a new scheme design that achieves efficiency and security requirements simultaneously with the preservation of its key characteristics, by randomly splitting the original image data, designing two novel efficient protocols for secure multiplication and comparison, and carefully distributing the feature extraction computations onto two independent cloud servers. We both carefully analyze and extensively evaluate the security and effectiveness of our design. The results show that our solution is practically secure, outperforms the state-of-the-art, and performs comparably with the original SIFT in terms of various characteristics, including rotation invariance, image scale invariance, robust matching across affine distortion, and addition of noise and change in 3D viewpoint and illumination.

NS16NXT10

TITLE: A Secure-Efficient Data Collection Algorithm Based on Self-Adaptive Sensing Model in Mobile Internet of Vehicles

ABSTRACT: Existing research on data collection using wireless mobile vehicle network emphasizes the reliable delivery of information. However, other performance requirements such as life cycle of nodes, stability and security are not set as primary design objectives. This makes data collection ability of vehicular nodes in real application environment inferior. By considering the features of nodes in wireless 10V, such as large scales of deployment, volatility and low time delay, an efficient data collection algorithm is proposed for mobile vehicle network environment. An adaptive sensing model is designed to establish vehicular data collection protocol. The protocol adopts group management in model communication. The vehicular sensing node in group can adjust network sensing chain according to sensing distance threshold with surrounding nodes. It will dynamically choose a combination of network sensing
chains on basis of remaining energy and location characteristics of surrounding nodes. In addition, secure data collection between sensing nodes is undertaken as well. The simulation and experiments show that the vehicular node can realize secure and real-time data collection. Moreover, the proposed algorithm is superior in vehicular network life cycle, power consumption and reliability of data collection by comparing to other algorithms.

**NS16NXT11**

**TITLE: A Clean Slate Approach to Secure Ad Hoc Wireless Networking - Open Unsynchronized Networks**

**ABSTRACT:** Distributed cyber physical systems depend on secure wireless ad-hoc networks to ensure that the sensors, controllers, and actuators (or nodes) in the system can reliably communicate. Such networks are difficult to design because, being inherently complex, they are vulnerable to attack. As a result, the current process of designing secure protocols for wireless ad-hoc networks is effectively an arms race between discovering attacks and creating fixes. At no point in the process is it possible to make provable performance and security guarantees. This paper proposes a system-theoretic framework for the design of secure open wireless ad-hoc networks, that provides precisely such guarantees. The nodes are initially unsynchronized, and join the network at any stage of the operation. The framework consists of a zero-sum game between all protocols and adversarial strategies, in which the protocol is announced before the adversarial strategy. Each choice of protocol and adversarial strategy results in a payoff. The design imperative is to choose the protocol that achieves the optimal payoff. We propose an “edge-tally supervised” merge protocol that is theoretically significant in three ways. First, the protocol achieves the max-min payoff; the highest possible payoff since the adversarial strategy always knows the protocol a priori. Second, the protocol actually does better and achieves the minmax payoff; it is a Nash equilibrium in the space of protocols and adversarial strategies. The adversarial nodes gain no advantage from knowing the protocol a priori. Third, the adversarial nodes are effectively limited to either jamming or conforming to the protocol; more complicated behaviors yield no strategic
| NS16NXT12 | **TITLE**: A Stable Approach for Routing Queries in Unstructured P2P Networks  
**ABSTRACT**: Finding a document or resource in an unstructured peer-to-peer network can be an exceedingly difficult problem. In this paper we propose a query routing approach that accounts for arbitrary overlay topologies, nodes with heterogeneous processing capacity, e.g., reflecting their degree of altruism, and heterogenous class-based likelihoods of query resolution at nodes which may reflect query loads and the manner in which files/resources are distributed across the network. The approach is shown to be stabilize the query load subject to a grade of service constraint, i.e., a guarantee that queries’ routes meet pre-specified class-based bounds on their associated a priori probability of query resolution. An explicit characterization of the capacity region for such systems is given and numerically compared to that associated with random walk based searches. Simulation results further show the performance benefits, in terms of mean delay, of the proposed approach. Additional aspects associated with reducing complexity, estimating parameters, and adaptation to class-based query resolution probabilities and traffic loads are studied. |
| NS16NXT13 | **TITLE**: iPath: Path Inference in Wireless Sensor Networks  
**ABSTRACT**: Recent wireless sensor networks (WSNs) are becoming increasingly complex with the growing network scale and the dynamic nature of wireless communications. Many measurement and diagnostic approaches depend on per-packet routing paths for accurate and fine-grained analysis of the complex network behaviors. In this paper, we propose iPath, a novel path inference approach to reconstructing the per-packet routing paths in dynamic and large-scale networks. The basic idea of iPath is to exploit high path similarity to iteratively infer long paths from short ones. iPath starts with an initial known set of paths and performs path inference iteratively. iPath includes a novel design of a lightweight hash function for verification of the inferred paths. In order to further improve the inference capability as well as the execution efficiency, iPath |
includes a fast bootstrapping algorithm to reconstruct the initial set of paths. We also implement iPath and evaluate its performance using traces from large-scale WSN deployments as well as extensive simulations. Results show that iPath achieves much higher reconstruction ratios under different network settings compared to other state-of-the-art approaches.

| NS16NXT14 | TITLE: Opportunistic Routing With Congestion Diversity in Wireless Ad Hoc Networks  
| | ABSTRACT: We consider the problem of routing packets across a multi-hop network consisting of multiple sources of traffic and wireless links while ensuring bounded expected delay. Each packet transmission can be overheard by a random subset of receiver nodes among which the next relay is selected opportunistically. The main challenge in the design of minimum-delay routing policies is balancing the trade-off between routing the packets along the shortest paths to the destination and distributing the traffic according to the maximum backpressure. Combining important aspects of shortest path and backpressure routing, this paper provides a systematic development of a distributed opportunistic routing policy with congestion diversity (D-ORCD). D-ORCD uses a measure of draining time to opportunistically identify and route packets along the paths with an expected low overall congestion. D-ORCD with single destination is proved to ensure a bounded expected delay for all networks and under any admissible traffic, so long as the rate of computations is sufficiently fast relative to traffic statistics. Furthermore, this paper proposes a practical implementation of D-ORCD which empirically optimizes critical algorithm parameters and their effects on delay as well as protocol overhead. Realistic QualNet simulations for 802.11-based networks demonstrate a significant improvement in the average delay over comparable solutions in the literature. |

| NS16NXT15 | TITLE: Spatial Reusability-Aware Routing in Multi-Hop Wireless Networks  
| | ABSTRACT: In the problem of routing in multi-hop wireless networks, to achieve high end-to-end throughput, it is crucial to find the “best” path from the source
node to the destination node. Although a large number of routing protocols have been proposed to find the path with minimum total transmission count/time for delivering a single packet, such transmission count/time minimizing protocols cannot be guaranteed to achieve maximum end-to-end throughput. In this paper, we argue that by carefully considering spatial reusability of the wireless communication media, we can tremendously improve the end-to-end throughput in multi-hop wireless networks. To support our argument, we propose spatial reusability-aware single-path routing (SASR) and any path routing (SAAR) protocols, and compare them with existing single-path routing and any path routing protocols, respectively. Our evaluation results show that our protocols significantly improve the end-to-end throughput compared with existing protocols. Specifically, for single-path routing, the median throughput gain is up to 60 percent, and for each source-destination pair, the throughput gain is as high as 5.3x; for any path routing, the maximum per-flow throughput gain is 71.6 percent, while the median gain is up to 13.2 percent.

**NS16NXT16**

**TITLE:** STAMP: Enabling Privacy-Preserving Location Proofs for Mobile Users

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

**NS16NXT17**

**TITLE:** 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 smart phone 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.

**NS16NXT18**

**TITLE:** Detecting Malicious Facebook Applications

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

NS16NXT19

TITLE: Carrier Aggregation Between Operators in Next Generation Cellular Networks: A Stable Roommate Market

ABSTRACT: This paper studies carrier aggregation between multiple mobile network operators (MNOs), referred to as inter operator carrier aggregation (IO-CA). In IO-CA, each MNO can transmit on its own licensed spectrum and aggregate the spectrum licensed to other MNOs. We focus on the case that MNOs are partitioned and distributed into small groups, called IO-CA pairs, each of which consists of two MNOs that mutually agree to share their spectrum with each other. We model the IO-CA pairing problem between MNOs as a stable room mate market and derive a condition for which a stable matching structure among all MNOs exist. We propose an algorithm that achieves a stable matching if it exists. Otherwise, the algorithm results in a stable partition. For each IO-CA pair, we derive the optimal transmit power for each spectrum aggregator and establish a Stackelberg game model to analyze the interaction between the
licensed subscribers and aggregators in the spectrum of each MNO. We derive
the Stackelberg equilibrium of our proposed game and then develop a joint
optimization algorithm that achieves the stable matching structure among MNOs
as well as the optimal transmit powers for the aggregators and prices for the
subscribers of each MNO’.

NS16NXT20  TITLE: Secure Routing Based on Social Similarity in Opportunistic Networks

ABSTRACT: The lack of pre-existing infrastructure or dynamic topology makes it
impossible to establish end-to-end connections in opportunistic networks
(OppNets). Instead, a store-and-forward strategy can be employed. However,
such loosely knit routing paths depend heavily on the cooperation among
participating nodes. Selfish or malicious behaviors of nodes impact greatly on the
network performance. In this paper, we design and validate a dynamic trust
management model for secure routing optimization. We propose the concept of
incorporating social trust into the routing decision process and design a trust
routing based on social similarity (TRSS) scheme. TRSS is based on the
observation that nodes move around and contact each other according to their
common interests or social similarities. A node sharing more social features in
social history record with the destination is more likely to travel close to the
latter in the near future and should be chosen as the next-hop forwarder.
Furthermore, social trust can be established based on an observed node's
trustworthiness and its encounter history. Based on direct and recommended
trust, those untrustworthy nodes will be detected and purged from the trusted
list. Since only trusted nodes' packets will be forwarded, the selfish nodes have
the incentives to behave well again. Simulation evaluation demonstrates that
TRSS is very effective in detecting selfish or even malicious nodes and achieving
better performance.
TITLE: Impact of mobile nodes for few mobility models on delay-tolerant network routing protocols

ABSTRACT: Delay-Tolerant Networks (DTNs) are sparse dynamic wireless networks, where most of the time a complete end-to-end path from the source to the destination does not exist. There are many real networks that follow this model, for example, military networks, vehicular ad-hoc networks (VANETs), wildlife tracking sensor networks, etc. In this context, conventional mobile ad-hoc routing schemes would fail, because they try to establish complete end-to-end paths, before any data sent. Therefore, performance analysis of different DTN routing mechanisms plays an important role in understanding the design of DTNs that encourages one to choose proper routing protocol for a particular scenario. This paper investigates the performance of replication-based DTN routing protocols, namely Epidemic, Probabilistic Routing Protocol using History of Encounters and Transitivity (PRoPHET), MaxProp, Resource Allocation Protocol for Intentional DTN (RAPID), Binary Spray-and-Wait (B-SNW), and Spray-and-Focus (SNF) in DTN scenario against varying number of mobile nodes for three mobility models, namely Random Walk (RW), Random Direction (RD) and Shortest Path Map Based (SPMB) movement. Performance has been evaluated and analyzed using Opportunistic Network Environment (ONE) simulator by considering three metrics: delivery probability, average latency and overhead ratio. Simulation results show that the investigated DTN routing protocols, in general, exhibit better performance in the SPMB movement model than other movement models, i.e., RW and RD, because they yield maximum message delivery probability, minimum overhead ratio (for B-SNW and SNF), and minimum average latency.

TITLE: Biosec: a biometric based approach for securing communication in wireless networks of biosensors implanted in the human body

ABSTRACT: Advances in microelectronics, material science and wireless technology have led to the development of sensors that can be used for accurate
monitoring of inaccessible environments. Health monitoring, telemedicine, military and environmental monitoring are some of the applications where sensors can be used. The sensors implanted inside the human body to monitor parts of the body are called biosensors. These biosensors form a network and collectively monitor the health condition of their carrier or host. Health monitoring involves collection of data about vital body parameters from different parts of the body and making decisions based on it. This information is of personal nature and is required to be secured. Insecurity may also lead to dangerous consequences. Due to the extreme constraints of energy, memory and computation securing the communication among the biosensors is not a trivial problem. Key distribution is central to any security mechanism. In this paper we propose an approach wherein, biometrics derived from the body are used for securing the keying material. This method obviates the need for expensive computation and avoids unnecessary communication making our approach novel compared to existing approaches.

<table>
<thead>
<tr>
<th>NS16NXT23</th>
<th>TITLE: Interference-aware relay in hybrid wireless networks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> We develop an interference-aware relaying scheme in hybrid wireless networks. The scheme is derived from the optimization model which generalizes the multi-hop package delivery problem in hybrid wireless networks. It aims at minimizing the network costs via optimal power loading and relaying route. Besides, the new cost model considering interference impact is an advance on the basis of our previous work. The advanced model takes non-uniform traffic and interference distribution into account. As a result, the relaying scheme is applicable to the scenarios which are more general. The simulation result shows that the proposed scheme is adaptive to traffic with different features. It can automatically balance the traffic loads and significantly improve the performance of the hybrid network. In the best case, the spectral efficiency of the network employing the hybrid scheme reaches 4.3 bit/s/Hz/cell compared to 2.4 bit/s/Hz/cell for the traditional cellular mode in the same case.</td>
<td></td>
</tr>
</tbody>
</table>
TITLE: Edu-firewall device: An advanced firewall hardware device for information security education

ABSTRACT: Firewalls are security devices used to apply an organization's security policy. However, commercial firewalls, such as Juniper Networks and Cisco ASA firewall devices, are mainly designed to be used by networking and security professionals, are not very appropriate for the academia environment, and lack simplicity. In addition, commercial firewalls are usually considered high-cost hardware devices. However, academic institutions usually have tight budget and few financial resources for purchasing networking and security devices for their laboratories. To overcome the aforementioned limitation of commercial firewalls, an educational firewall hardware device, called Edu-Firewall, is discussed and demonstrated. Edu-Firewall main objectives are to offer advanced educational security functions that are not commonly available in current commercial firewalls, and an easy-to-use friendly graphical interface to configure the firewall and manipulate the filtering rules. In addition, Edu-Firewall's objective is to offer an affordable educational device, compared to the available commercial firewall devices. Edu-Firewall allows students to implement hands-on lab exercises on firewalls and better anatomize network traffic filtering concepts and firewall configuration, and contributes to enhance students' hands-on security skills.

TITLE: Privacy and security problems of national health data warehouse: a convenient solution for developing countries

ABSTRACT: Healthcare providers and researchers can discover the hidden knowledge from different health repositories if integration of data is performed by data warehousing. Integration of health records requires linkage of patients' data in different heterogeneous sources. Preserving record linkage in National Health Data Warehouse, by retaining identifiable attributes, is essential for effective data mining as well. In contrast identifiable health data have high risk to patient privacy and make the health information systems security vulnerable to hackers. In this paper, we have provided a practical solution: Global Patient
Identification Technique (GPIT) that can anonymize identifiable private data of the patients while maintaining record linkage in integrated health repositories to facilitate knowledge discovery process. We have used encrypted mobile number, gender and NAMEVALUE of patients to produce Global Patient Identification Key. This system is being implemented in Bangladesh to develop National Health Data Warehouse.

SOFTWARE ENGINEERING

SE16NXT01  TITLE: A Tool-Supported Methodology for Validation and Refinement of Early-Stage Domain Models

ABSTRACT: Model-driven engineering (MDE) promotes automated model transformations along the entire development process. Guaranteeing the quality of early models is essential for a successful application of MDE techniques and related tool-supported model refinements. Do these models properly reflect the requirements elicited from the owners of the problem domain? Ultimately, this question needs to be asked to the domain experts. The problem is that a gap exists between the respective backgrounds of modeling experts and domain experts. MDE developers cannot show a model to the domain experts and simply ask them whether it is correct with respect to the requirements they had in mind. To facilitate their interaction and make such validation more systematic, we propose a methodology and a tool that derive a set of customizable questionnaires expressed in natural language from each model to be validated. Unexpected answers by domain experts help to identify those portions of the models requiring deeper attention. We illustrate the methodology and the current status of the developed tool MOTHIA, which can handle UML Use Case, Class, and Activity diagrams. We assess MOTHIA effectiveness in reducing the gap between domain and modeling experts, and in detecting modeling faults on the European Project CHOReOS.
| SE16NXT02 | TITLE: Trust Agent-Based Behavior Induction in Social Networks  
ABSTRACT: The essence of social networks is that they can influence people's public opinions and group behaviors form quickly. Negative group behavior influences societal stability significantly, but existing behavior-induction approaches are too simple and inefficient. To automatically and efficiently induct behavior in social networks, this article introduces trust agents and designs their features according to group behavior features. In addition, a dynamics control mechanism can be generated to coordinate participant behaviors in social networks to avoid a specific restricted negative group behavior. |
| SE16NXT03 | TITLE: A Multi-Site Joint Replication of a Design Patterns Experiment Using Moderator Variables to Generalize across Contexts  
ABSTRACT: Several empirical studies have explored the benefits of software design patterns, but their collective results are highly inconsistent. Resolving the inconsistencies requires investigating moderators—i.e., variables that cause an effect to differ across contexts. Objectives. Replicate a design patterns experiment at multiple sites and identify sufficient moderators to generalize the results across prior studies. Methods. We perform a close replication of an experiment investigating the impact (in terms of time and quality) of design patterns (Decorator and Abstract Factory) on software maintenance. The experiment was replicated once previously, with divergent results. We execute our replication at four universities—spanning two continents and three countries—using a new method for performing distributed replications based on closely coordinated, small-scale instances (“joint replication”). We perform two analyses: 1) a post-hoc analysis of moderators, based on frequentist and Bayesian statistics; 2) an a priori analysis of the original hypotheses, based on frequentist statistics. Results. The main effect differs across the previous instances of the experiment and across the sites in our distributed replication. Our analysis of moderators (including developer experience and pattern knowledge) resolves the differences sufficiently to allow for cross-context (and cross-study) conclusions. The final conclusions represent 126 participants from
five universities and 12 software companies, spanning two continents and at least four countries. Conclusions. The Decorator pattern is found to be preferable to a simpler solution during maintenance, as long as the developer has at least some prior knowledge of the pattern. For Abstract Factory, the simpler solution is found to be mostly equivalent to the pattern solution. Abstract Factory is shown to require a higher level of knowledge and/or experience than Decorator for the pattern to be beneficial.

SE16NXT04  TITLE: Mapping Bug Reports to Relevant Files: A Ranking Model, a Fine-Grained Benchmark, and Feature Evaluation

ABSTRACT: When a new bug report is received, developers usually need to reproduce the bug and perform code reviews to find the cause, a process that can be tedious and time consuming. A tool for ranking all the source files with respect to how likely they are to contain the cause of the bug would enable developers to narrow down their search and improve productivity. This paper introduces an adaptive ranking approach that leverages project knowledge through functional decomposition of source code, API descriptions of library components, the bug-fixing history, the code change history, and the file dependency graph. Given a bug report, the ranking score of each source file is computed as a weighted combination of an array of features, where the weights are trained automatically on previously solved bug reports using a learning-to-rank technique. We evaluate the ranking system on six large scale open source Java projects, using the before-fix version of the project for every bug report. The experimental results show that the learning-to-rank approach outperforms three recent state-of-the-art methods. In particular, our method makes correct recommendations within the top 10 ranked source files for over 70 percent of the bug reports in the Eclipse Platform and Tomcat projects.

SE16NXT05  TITLE: RELAI Testing: A Technique to Assess and Improve Software Reliability

ABSTRACT: Testing software to assess or improve reliability presents several practical challenges. Conventional operational testing is a fundamental strategy
that simulates the real usage of the system in order to expose failures with the highest occurrence probability. However, practitioners find it unsuitable for assessing/achieving very high reliability levels; also, they do not see the adoption of a “real” usage profile estimate as a sensible idea, being it a source of non-quantifiable uncertainty. Oppositely, debug testing aims to expose as many failures as possible, but regardless of their impact on runtime reliability. These strategies are used either to assess or to improve reliability, but cannot improve and assess reliability in the same testing session. This article proposes Reliability Assessment and Improvement (RELAI) testing, a new technique thought to improve the delivered reliability by an adaptive testing scheme, while providing, at the same time, a continuous assessment of reliability attained through testing and fault removal. The technique also quantifies the impact of a partial knowledge of the operational profile. RELAI is positively evaluated on four software applications compared, in separate experiments, with techniques conceived either for reliability improvement or for reliability assessment, demonstrating substantial improvements in both cases.

SE16NXT06

TITLE: GALE: Geometric Active Learning for Search-Based Software Engineering

ABSTRACT: Multi-objective evolutionary algorithms (MOEAs) help software engineers find novel solutions to complex problems. When automatic tools explore too many options, they are slow to use and hard to comprehend. GALE is a near-linear time MOEA that builds a piecewise approximation to the surface of best solutions along the Pareto frontier. For each piece, GALE mutates solutions towards the better end. In numerous case studies, GALE finds comparable solutions to standard methods (NSGA-II, SPEA2) using far fewer evaluations (e.g. 20 evaluations, not 1,000). GALE is recommended when a model is expensive to evaluate, or when some audience needs to browse and understand how an MOEA has made its conclusions.
<table>
<thead>
<tr>
<th>SE16NXT07</th>
<th>TITLE: Round-Up: Runtime Verification of Quasi Linearizability for Concurrent Data Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> We propose a new method for runtime checking of a relaxed consistency property called quasi linearizability for concurrent data structures. Quasi linearizability generalizes the standard notion of linearizability by introducing nondeterminism into the parallel computations quantitatively and then exploiting such nondeterminism to improve the runtime performance. However, ensuring the quantitative aspects of this correctness condition in the low-level code of the concurrent data structure implementation is a difficult task. Our runtime verification method is the first fully automated method for checking quasi linearizability in the C/C++ code of concurrent data structures. It guarantees that all the reported quasi linearizability violations manifested by the concurrent executions are real violations. We have implemented our method in a software tool based on the LLVM compiler and a systematic concurrency testing tool called Inspect. Our experimental evaluation shows that the new method is effective in detecting quasi linearizability violations in the source code implementations of concurrent data structures.</td>
</tr>
<tr>
<td>SE16NXT08</td>
<td>TITLE: Mining Workflow Models from Web Applications</td>
</tr>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> Modern business applications predominantly rely on web technology, enabling software vendors to efficiently provide them as a service, removing some of the complexity of the traditional release and update process. While this facilitates shorter, more efficient and frequent release cycles, it requires continuous testing. Having insight into application behavior through explicit models can largely support development, testing and maintenance. Model-based testing allows efficient test creation based on a description of the states the application can be in and the transitions between these states. As specifying behavior models that are precise enough to be executable by a test automation tool is a hard task, an alternative is to extract them from running applications. However, mining such models is a challenge, in particular because</td>
</tr>
</tbody>
</table>
one needs to know when two states are equivalent, as well as how to reach that state. We present Process Crawler (ProCrawl), a tool to mine behavior models from web applications that support multi-user workflows. ProCrawl incrementally learns a model by generating program runs and observing the application behavior through the user interface. In our evaluation on several real-world web applications, ProCrawl extracted models that concisely describe the implemented workflows and can be directly used for model-based testing.

**SE16NXT09**

**TITLE: Self-Adapting Reliability in Distributed Software Systems**

**ABSTRACT:** Developing modern distributed software systems is difficult in part because they have little control over the environments in which they execute. For example, hardware and software resources on which these systems rely may fail or become compromised and malicious. Redundancy can help manage such failures and compromises, but when faced with dynamic, unpredictable resources and attackers, the system reliability can still fluctuate greatly. Empowering the system with self-adaptive and self-managing reliability facilities can significantly improve the quality of the software system and reduce reliance on the developer predicting all possible failure conditions. We present iterative redundancy, a novel approach to improving software system reliability by automatically injecting redundancy into the system's deployment. Iterative redundancy self-adapts in three ways: (1) by automatically detecting when the resource reliability drops, (2) by identifying unlucky parts of the computation that happen to deploy on disproportionately many compromised resources, and (3) by not relying on a priori estimates of resource reliability. Further, iterative redundancy is theoretically optimal in its resource use: Given a set of resources, iterative redundancy guarantees to use those resources to produce the most reliable version of that software system possible; likewise, given a desired increase in the system's reliability, iterative redundancy guarantees achieving that reliability using the least resources possible. Iterative redundancy handles even the Byzantine threat model, in which compromised resources collude to attack the system. We evaluate iterative redundancy in three ways. First, we
formally prove its self-adaptation, efficiency, and optimality properties. Second, we simulate it at scale using discrete event simulation. Finally, we modify the existing, open-source, volunteer-computing BOINC software system and deploy it on the globally-distributed Planet Lab tested network to empirically evaluate that iterative redundancy is self-adaptive and more efficient than existing techniques.

<table>
<thead>
<tr>
<th>SE16NXT10</th>
<th>TITLE: Reducing Feedback Delay of Software Development Tools via Continuous Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: During software development, the sooner a developer learns how code changes affect program analysis results, the more helpful that analysis is. Manually invoking an analysis may interrupt the developer's workflow or cause a delay before the developer learns the implications of the change. A better approach is continuous analysis tools that always provide up-to-date results. We present Codebase Replication, a technique that eases the implementation of continuous analysis tools by converting an existing offline analysis into an IDE-integrated, continuous tool with two desirable properties: isolation and currency. Codebase Replication creates and keeps in sync a copy of the developer's codebase. The analysis runs on the copy codebase without disturbing the developer and without being disturbed by the developer's changes. We developed Solstice, an open-source, publicly-available Eclipse plug-in that implements Codebase Replication. Solstice has less than 2.5 milliseconds overhead for most common developer actions. We used Solstice to implement four Eclipse-integrated continuous analysis tools based on the offline versions of Find Bugs, PMD, data race detection, and unit testing. Each conversion required on average 710 LoC and 20 hours of implementation effort. Case studies indicate that Solstice-based continuous analysis tools are intuitive and easy-to-use.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SE16NXT11</th>
<th>TITLE: Detecting, Tracing, and Monitoring Architectural Tactics in Code</th>
</tr>
</thead>
</table>
| ABSTRACT: Software architectures are often constructed through a series of design decisions. In particular, architectural tactics are selected to satisfy specific
quality concerns such as reliability, performance, and security. However, the knowledge of these tactical decisions is often lost, resulting in a gradual degradation of architectural quality as developers modify the code without fully understanding the underlying architectural decisions. In this paper we present a machine learning approach for discovering and visualizing architectural tactics in code, mapping these code segments to tactic traceability patterns, and monitoring sensitive areas of the code for modification events in order to provide users with up-to-date information about underlying architectural concerns. Our approach utilizes a customized classifier which is trained using code extracted from fifty performance-centric and safety-critical open source software systems. Its performance is compared against seven off-the-shelf classifiers. In a controlled experiment all classifiers performed well; however our tactic detector outperformed the other classifiers when used within the larger context of the Hadoop Distributed File System. We further demonstrate the viability of our approach for using the automatically detected tactics to generate viable and informative messages in a simulation of maintenance events mined from Hadoop’s change management system.

SE16NXT12  TITLE: A Probabilistic Analysis of the Efficiency of Automated Software Testing

ABSTRACT: We study the relative efficiencies of the random and systematic approaches to automated software testing. Using a simple but realistic set of assumptions, we propose a general model for software testing and define sampling strategies for random (R) and systematic (S₀) testing, where each sampling is associated with a sampling cost: 1 and c units of time, respectively. The two most important goals of software testing are: (i) achieving in minimal time a given degree of confidence x in a program’s correctness and (ii) discovering a maximal number of errors within a given time bound n̂. For both (i) and (ii), we show that there exists a bound on c beyond which R performs better than S₀ on the average. Moreover for (i), this bound depends asymptotically only on x. We also show that the efficiency of R can be fitted to the exponential curve. Using these results we design a hybrid strategy H that starts with R and
switches to S₀ when S₀ is expected to discover more errors per unit time. In our
experiments we find that H performs similarly or better than the most efficient
of both and that S₀ may need to be significantly faster than our bounds suggest
to retain efficiency over R.

TITLE: Using Reduced Execution Flow Graph to Identify Library Functions in
Binary Code

ABSTRACT: Discontinuity and polymorphism of a library function create two
challenges for library function identification, which is a key technique in reverse
engineering. A new hybrid representation of dependence graph and control flow
graph called Execution Flow Graph (EFG) is introduced to describe the semantics
of binary code. Library function identification turns to be a subgraph
isomorphism testing problem since the EFG of a library function instance is
isomorphic to the sub-EFG of this library function. Subgraph isomorphism
detection is time-consuming. Thus, we introduce a new representation called
Reduced Execution Flow Graph (REFG) based on EFG to speed up the
isomorphism testing. We have proved that EFGs are subgraph isomorphic as long
as their corresponding REFGs are subgraph isomorphic. The high efficiency of the
REFG approach in subgraph isomorphism detection comes from fewer nodes and
edges in REFGs and new lossless filters for excluding the unmatched subgraphs
before detection. Experimental results show that precisions of both the EFG and
REFG approaches are higher than the state-of-the-art tool and the REFG
approach sharply decreases the processing time of the EFG approach with
consistent precision and recall.

TITLE: Seer: A Lightweight Online Failure Prediction Approach

ABSTRACT: Online failure prediction approaches aim to predict the
manifestation of failures at runtime before the failures actually occur. Existing
approaches generally refrain themselves from collecting internal execution data,
which can further improve the prediction quality. One reason behind this general
trend is the runtime overhead incurred by the measurement instruments that
collect the data. Since these approaches are targeted at deployed software systems, excessive runtime overhead is generally undesirable. In this work we conjecture that large cost reductions in collecting internal execution data for online failure prediction may derive from pushing the substantial parts of the data collection work onto the hardware. To test this hypothesis, we present a lightweight online failure prediction approach, called Seer, in which most of the data collection work is performed by fast hardware performance counters. The hardware-collected data is augmented with further data collected by a minimal amount of software instrumentation that is added to the systems software. In our empirical evaluations conducted on three open source projects, Seer performed significantly better than other related approaches in predicting the manifestation of failures.

<table>
<thead>
<tr>
<th>SE16NXT15</th>
<th>TITLE: Automatic Source Code Summarization of Context for Java Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ABSTRACT:</strong> Source code summarization is the task of creating readable summaries that describe the functionality of software. Source code summarization is a critical component of documentation generation, for example as Java docs formed from short paragraphs attached to each method in a Java program. At present, a majority of source code summarization is manual, in that the paragraphs are written by human experts. However, new automated technologies are becoming feasible. These automated techniques have been shown to be effective in select situations, though a key weakness is that they do not explain the source code’s context. That is, they can describe the behavior of a Java method, but not why the method exists or what role it plays in the software. In this paper, we propose a source code summarization technique that writes English descriptions of Java methods by analyzing how those methods are invoked. We then performed two user studies to evaluate our approach. First, we compared our generated summaries to summaries written manually by experts. Then, we compared our summaries to summaries written by a state-of-the-art automatic summarization tool. We found that while our approach does not reach the quality of human-written summaries, we do improve over the</td>
</tr>
<tr>
<td>TITLE: Crossover Designs in Software Engineering Experiments: Benefits and Perils</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>ABSTRACT:</strong> In experiments with crossover design subjects apply more than one treatment. Crossover designs are widespread in software engineering experimentation: they require fewer subjects and control the variability among subjects. However, some researchers disapprove of crossover designs. The main criticisms are: the carryover threat and its troublesome analysis. Carryover is the persistence of the effect of one treatment when another treatment is applied later. It may invalidate the results of an experiment. Additionally, crossover designs are often not properly designed and/or analysed, limiting the validity of the results. In this paper, we aim to make SE researchers aware of the perils of crossover experiments and provide risk avoidance good practices. We study how another discipline (medicine) runs crossover experiments. We review the SE literature and discuss which good practices tend not to be adhered to, giving advice on how they should be applied in SE experiments. We illustrate the concepts discussed analysing a crossover experiment that we have run. We conclude that crossover experiments can yield valid results, provided they are properly designed and analysed, and that, if correctly addressed, carryover is no worse than other validity threats.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TITLE: GoPrime: A Fully Decentralized Middleware for Utility-Aware Service Assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Modern applications, e.g., for pervasive computing scenarios, are increasingly reliant on systems built from multiple distributed components, which must be suitably composed to meet some specified functional and non-functional requirements. A key challenge is how to efficiently and effectively manage such complex systems. The use of self-management capabilities has been suggested as a possible way to address this challenge. To cope with the...</td>
</tr>
</tbody>
</table>
scalability and robustness issues of large distributed systems, self-management should ideally be architected in a decentralized way, where the overall system behavior emerges from local decisions and interactions. Within this context, we propose GOPRIME, a fully decentralized middleware solution for the adaptive self-assembly of distributed services. The GOPRIME goal is to build and maintain an assembly of services that, besides functional requirements, fulfills also global quality-of-service and structural requirements. The key aspect of GOPRIME is the use of a gossip protocol to achieve decentralized information dissemination and decision making. To show the validity of our approach, we present results from the experimentation of a prototype implementation of GOPRIME in a mobile health application, and an extensive set of simulation experiments that assess the effectiveness of GOPRIME in terms of scalability, robustness and convergence speed.

SE16NXT18

TITLE: Probabilistic Model Checking of Regenerative Concurrent Systems

ABSTRACT: We consider the problem of verifying quantitative reachability properties in stochastic models of concurrent activities with generally distributed durations. Models are specified as stochastic time Petri nets and checked against Boolean combinations of interval until operators imposing bounds on the probability that the marking process will satisfy a goal condition at some time in the interval \([\alpha, \beta]\) after an execution that never violates a safety property. The proposed solution is based on the analysis of regeneration points in model executions: a regeneration is encountered after a discrete event if the future evolution depends only on the current marking and not on its previous history, thus satisfying the Markov property. We analyze systems in which multiple generally distributed timers can be started or stopped independently, but regeneration points are always encountered with probability 1 after a bounded number of discrete events. Leveraging the properties of regeneration points in probability spaces of execution paths, we show that the problem can be reduced to a set of Volterra integral equations, and we provide algorithms to compute their parameters through the enumeration of finite sequences of stochastic state.
classes encoding the joint probability density function (PDF) of generally
distributed timers after each discrete event. The computation of symbolic PDFs is
limited to discrete events before the first regeneration, and the repetitive
structure of the stochastic process is exploited also before the lower bound $\alpha$,
providing crucial benefits for large time bounds. A case study is presented
through the probabilistic formulation of Fischer's mutual exclusion protocol, a
well-known real-time verification benchmark.

**SE16NXT19**

**TITLE:** Achieving High Energy Efficiency and Physical-Layer Security in AF Relaying

**ABSTRACT:** For transmitting data in a secret and energy-efficient manner in
collaborative amplify-and-forward relay networks, the secure energy efficiency
(EE) defined as the secret bits transferred with unit energy is maximized to
satisfy each node power constraint and target secrecy rate requirement, based
on physical security framework. The secure EE is maximized by joint source and
relay power allocation, which is a non convex optimization problem. To cope
with this difficulty, a solution scheme and corresponding algorithms are
developed by jointly applying fractional programming, exact penalty, alternate
search, and difference of convex functions programming. The key idea of the
scheme is to convert the primal problem into simple sub problems step by step,
such that related methods are adopted. It is verified that, compared with secrecy
rate maximization, the proposed scheme improves the secure EE significantly yet
with a certain loss of the secrecy rate due to the tradeoff between secure EE and
secrecy rate. Furthermore, the proposed scheme achieves higher secure EE and
secrecy rate than total transmission power minimization does, while with a
certain increase of power consumption. These results indicate that a reasonable
balance among secure EE, secrecy rate, and power consumption can be reached
by the proposed scheme.
| SE16NXT20 | TITLE: Greening the Airwaves With Collaborating Mobile Network Operators  
**ABSTRACT:** Base station sharing is currently considered one of the most promising solutions for reducing the energy consumption costs of cellular networks. This paper presents a game theoretic framework for the study of such cooperative solutions where different mobile network operators (MNOs) decide to switch off subsets of their base stations during off-peak hours and roam their traffic to the remaining stations. The solution is based on a detailed optimization framework that determines exactly which base stations should remain active and how much traffic each one of them should serve, so as to maximize the aggregate energy savings. Accordingly, using the axiomatic Shapley value rule, it is determined how the benefits from the cooperation, i.e., the cost savings, should be dispersed among the cooperating MNOs. It is proved that this coalitional game with transferrable utilities has a nonempty core, and thus there exists a cooperation solution that incentivizes the participation of all operators. Moreover, using a thorough numerical analysis, it is shown that the benefits achieved with the implementation of the cooperation strategy depend mainly on the power consumption characteristics of the MNOs, which in turn are related to the number, type, and technology of their base stations. Overall, the energy savings are found to be most sensitive to the technology of the used base stations, and more precisely to the no-load base station energy consumption which defines the energy waste in a network. |
| SE16NXT21 | TITLE: Exploring topic models in software engineering data analysis: A survey  
**ABSTRACT:** Topic models are shown to be effective to mine unstructured software engineering (SE) data. In this paper, we give a simple survey of exploring topic models to support various SE tasks between 2003 and 2015. The survey results show that there is an increasing concern in this area. Among the SE tasks, source code comprehension and software history comprehension are the mostly studied, followed by software defects prediction. However, there is still only a few studies on other SE tasks, such as feature location and regression.
| SE16NXT22 | **TITLE:** Automatic support for formal specification construction using pattern knowledge  
**ABSTRACT:** Although formal specification is considered as a potential technique for improving the accuracy of requirements documentation and the quality of software product, the difficulty of using formal notations leads to the gap between this technique and the practice of software development. Many approaches for solving this problem were proposed. Most of them provide automatic transformation from informal requirements into formal specifications. However, rather than clarifying and formalizing requirements on the semantic level, they only use syntactic rules to translate between different languages. To handle the challenge, this paper describes an approach for formal specification construction based on pattern knowledge. The knowledge is composed of a set of inter-related specification patterns. Each pattern defines the method for formalizing one kind of function, including derivation knowledge for guiding the clarification of the function and transformation knowledge for formally representing the clarified function. A supporting tool is also described in the paper which derives necessary function details of the intended requirement through interactions by applying the derivation knowledge and transforms these details into formal specifications by applying the transformation knowledge. An experiment on the tool is held and the result shows that the tool can help formalize requirements efficiently and enhance the quality of the resultant formal specifications. |
<table>
<thead>
<tr>
<th>SE16NXT23</th>
<th>TITLE: Social service innovations with information technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Summary form only given. Service science research has been developing rapidly in this decade and provided a lot of contribution to enhance value of information systems. Most of IT engineers and scientists are promoting researches based on the concept what information systems make an effective, an optimal, and comfortable society. Since the e-services are provided to users, we need to define their values on user-friendliness, sustainability, economical efficiencies, effectiveness, satisfaction, and some other viewpoints. In this talk, I introduce the service fit model to realize what the users utilities are enhanced by e-services. The former half, we discuss the irrational issues of human activity and thinking. Even though some splendid services are provided by the information systems, people sometimes do not think the service is valuable because their thinking and services are not matched. In latter half, I introduce some examples and actual cases where the systems provide smart services and functions. IT solutions in tourism industry and semiconductors industry are introduced in end of this talk.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SE16NXT24</th>
<th>TITLE: Verifying RTuinOS using VCC: From approach to practice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> The correctness of implementation codes is important especially for safety-critical software which is usually written in C programming language. In this paper, we present a C code correctness verification framework (CVF for short) which is based on an automatic theorem proving tool-VCC, and pro-pose a specification checking method to improve the correctness of verification specification codes. We use a real-time OS, RTuinOS1.0.2, to evaluate our CVF framework. And the result shows it is feasible and effective to apply CVF to a real system software. Besides, the experiments also show that the specification checking method is effective.</td>
<td></td>
</tr>
</tbody>
</table>
| SE16NXT25 | TITLE: A streaming-based network monitoring and threat detection system.  
ABSTRACT: The unyielding trend of increasing cyber threats has made cyber security paramount in protecting personal and private intellectual property. In order to provide the most highly secured network environment, network traffic monitoring and threat detection systems must handle real-time data from varied and branching places in enterprise networks. Though numerous investigations have yielded real-time threat detection systems, in this paper we addressed the issue of handling the large volumes of network traffic data of enterprise systems, while simultaneously providing real-time monitoring and detection remain unsolved. Particularly, we introduced and evaluated a streaming-based threat detection system that can rapidly analyze highly intensive network traffic data in real-time, utilizing the streaming-based clustering algorithms to detect abnormal network activities. The developed system integrates the streaming and high-performance data analysis capabilities of Flume, Sharp, and Hadoop into a cloud-computing environment to provide network monitoring and intrusion detection. Our performance evaluation and experimental results demonstrate that the developed system can cope with a significant volume streaming data with high detection accuracy and good system performance. |

| PDS16NXT01 | TITLE: Cost-Effective Request Scheduling for Greening Cloud Data Centers  
ABSTRACT: With the popularity of cloud computing, many cloud service providers deploy regional data centers to offer services and applications. These large-scale data centers have drawn extensive attention in terms of the huge energy demand and carbon emission. Thus, how to make use of their spatial diversities to green data centers and reduce cloud provider's costs is an important concern. In this paper, we integrate service reward, electricity cost, carbon taxes and service performance to study cost-effective request scheduling for cloud data centers. We propose an online and distributed scheduling algorithm CESA to achieve the flexible tradeoff between these conflicting... |
objectives. The time complexity of CESA is polynomial, and it can be implemented in a parallel way. CESA requires no prior knowledge of the statistics of request arrivals or future electricity prices, yet it provably approximates the optimal system profit while bounding the queue length. Real-trace based simulations are conducted which verify the effectiveness of our CESA algorithm.

PDS16NXT02 | TITLE: Scalable Distributed Meta-Data Management in Parallel NFS

ABSTRACT: High performance computing (HPC) applications demand high I/O throughput from file systems to match their fast processing requirements. These HPC applications create large amounts of file meta-data that can overwhelm current single meta-data server file systems leading to performance bottlenecks. We address this problem with a multiple meta-data server (M-MDS) design using standard parallel NFS (pNFS). We utilize NFS directory referrals mechanism with hashing to efficiently distribute meta-data across a cluster of metadata servers. We show through a large scale setup in Amazon EC2 that our pNFS M-MDS can scale almost linearly and outperform Lustre CMD (Clustered Metadata) by up to 3 times in some of the file system meta-data operation benchmarks.

PDS16NXT03 | TITLE: Performance improvement of the parallel smith waterman algorithm implementation using Hybrid MPI-OpenMP model

ABSTRACT: This paper applies the hybrid parallel model that combines both shared and distributed memory architectures to improve the performance of the Smith waterman algorithm (SW). The hybrid model uses both MPI and OpenMp as programming techniques for different memory architectures. Our improved implementation executes a parallel version of SW algorithm with a row wise computation of the alignment matrix, which mainly optimizes the memory usage. We applied the parallel SW implementation and tested the system scalability on a homogenous cluster of up to eight nodes each of twenty four cores. We used the SWISS-PROT protein knowledgebase to test our implementation which achieved a tremendous reduction in the running time using the Hybrid MPI-OpenMP over the OpenMP and sequential
implementations. The Hybrid MPI-OpenMP achieved a speed up of 14X and 50X over the OpenMP and sequential implementations respectively when tested against all the SWISS-PROT protein knowledgebase entries.

PDS16NXT04

TITLE: Integration of DG systems composed of photovoltaic and a micro-turbine in remote areas

ABSTRACT: Utility side consumers are in the remote areas who are not linked by the central electrical grid network, hybrid systems such as Photovoltaic/Microturbine have been considered as the most reliable, desired and attractive unconventional source of power supply. This documents the analysis and simulation of a Photovoltaic/Microturbine hybrid system using MATLAB/SIMULINK, simpower system block sets as its simulation software. The system is designed entirely based on the concept of a parallel hybrid configuration. The software modeling of a Photovoltaic/Microturbine system provides an in-depth understanding of the system operation before building the actual system and also the testing and experiments of system operation under disturbances is not possible on the actual system.

PDS16NXT05

TITLE: Hierarchical Architecture for Integration of Rooftop PV in Smart Distribution Systems

ABSTRACT: The paper presents a novel hierarchical multilevel decentralized optimal power flow (OPF) for power loss minimization in three-phase unbalanced large-scale distribution systems via optimal reactive power scheduling of rooftop photovoltaics (PV) generators. The system is decomposed into three levels corresponding to the primary, the lateral and the secondary feeder sub-networks of distribution systems. A distributed sequential coordination scheme based on analytical target cascading (ATC) method is developed to minimize power losses while considering the operational constraints. Results based on the proposed method are compared with centralized OPF for validation. Control based on the proposed method is compared with no reactive power control and local reactive power control to
identify its effectiveness. Further, a virtual feeder is introduced to separate the coupled sub-networks into decomposed layers, which enables parallel processing of the optimization problems to reduce computational complexity and provide faster solution. A 559-node large-scale distribution network built based on the IEEE 37 node test system is used to demonstrate performance of the proposed algorithm.

**TITLE:** Transient voltage and frequency stability of an isolated microgrid based on energy storage systems

**ABSTRACT:** Microgrid (MG) based on energy storage systems to share the load between distributed generation plants in operation mode is the main issue in MG. Stability is an important component in energy management and planning of MG especially in duration of system's operation such as fault occurrence in system. In this article, frequency and voltage control based on active and reactive power control methods are analyzed. Studies demonstrate that MG stable operation in cases of proper use of control strategies is existing. MG system control due to stability improvement in islanding mode (autonomous mode) after fault occurrence at upstream network are been studying. MG system in this article included two distributed generation (DG) units. All of DG units and loads are connected in parallel at point of common coupling (PCC). In islanding mode, according to violence dependence of system's dynamic to local load changes and stability improvement after fault occurrence, design of controller algorithm is necessary. In this article, demonstrate that to frequency-load control, one of DG units is master and the other one is slave. Proposed controller based on energy storage system is design according to load uncertain. In final section of article, due to demonstrating the improvement and superior robustness of proposed controller to load dynamic, fault occurrence in system and controller capability in over demand supply and decrease short term produced power, frequency and voltage control by energy storage system cooperation that is one of novelty in this article, consider a comparison between classic and proposed controller. Proposed control strategy under two scenarios
(load change and fault occurrence) has a good performance. Finally, propose controller superior robustness performance evaluated by MATLAB software.

PDS16NXT07

TITLE: Improved parallel operation mode control of parallel connected 12-pulse thyristor dual converter for urban railway power substations

ABSTRACT: Parallel-connected 12-pulse thyristor dual converters are connected to the up/down trolley line in urban power substations. When there are high-traffic trains on the up trolley line or the down trolley line, the electric energy, i.e., energy consumed and regenerative energy, increases. During this time, two converters are connected in parallel to supply power or feed energy back to the distribution grid. A parallel operation control scheme has two operating modes depending on the energy flow of trains. When the energy is less than the rated power of a converter, only one converter operates in a single operation mode. If the energy exceeds the rated power of a converter, the other converter is operated in parallel mode operation to distribute the excess energy. During the switching from the single operation mode to the parallel operation mode, overshoots and undershoots occur in the DC trolley voltage and current of each converter due to the large error in the PI controller. To solve these problems, this study proposes improved control method to prevent large overshoots and undershoots. The effectiveness of this proposed control is verified by simulation and experiment based on the comparison of the performance of conventional and proposed control method.

PDS16NXT08

TITLE: Integration of DG systems composed of Photovoltaic and a Microturbine in remote areas

ABSTRACT: Utility side consumers are in the remote areas who are not linked by the central electrical grid network, hybrid systems such as Photovoltaic/Microturbine have been considered as the most reliable, desired and attractive unconventional source of power supply. This documents the analysis and simulation of a Photovoltaic/Microturbine hybrid system using MATLAB/SIMULINK, simpower system block sets as its simulation software. The
system is designed entirely based on the concept of a parallel hybrid configuration. The software modeling of a Photovoltaic/Microturbine system provides an in-depth understanding of the system operation before building the actual system and also the testing and experiments of system operation under disturbances is not possible on the actual system.

**PDS16NXT09**

**TITLE:** A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment

**ABSTRACT:** With the emergence of the big data age, the issue of how to obtain valuable knowledge from a dataset efficiently and accurately has attracted increasingly attention from both academia and industry. This paper presents a Parallel Random Forest (PRF) algorithm for big data on the Apache Spark platform. The PRF algorithm is optimized based on a hybrid approach combining data-parallel and task-parallel optimization. From the perspective of data-parallel optimization, a vertical data-partitioning method is performed to reduce the data communication cost effectively, and a data-multiplexing method is performed to allow the training dataset to be reused and diminish the volume of data. From the perspective of task-parallel optimization, a dual parallel approach is carried out in the training process of RF, and a task Directed Acyclic Graph (DAG) is created according to the parallel training process of PRF and the dependence of the Resilient Distributed Datasets (RDD) objects. Then, different task schedulers are invoked for the tasks in the DAG. Moreover, to improve the algorithm’s accuracy for large, high-dimensional, and noisy data, we perform a dimension-reduction approach in the training process and a weighted voting approach in the prediction process prior to parallelization. Extensive experimental results indicate the superiority and notable advantages of the PRF algorithm over the relevant algorithms implemented by Spark MLLib and other studies in terms of the classification accuracy, performance, and scalability.
TITLE: Data-driven fault detection for networked control system based on implicit model approach

ABSTRACT: In this paper, we propose a data-driven fault detection algorithm for the nonlinear system with network-induced time delay. Double-layer Quasi Takagi-Sugeno fuzzy modeling method is employed to represent the system. Quasi fuzzy model addresses the issue of time delay on the inner layer. Classical fuzzy model depicts the nonlinearity of the networked control system (NCS) on the outer layer. We apply implicit model approach to achieve the data-driven fault detection for the system. The implicit model approach provides an alternative perspective of the recently developed subspace-based data-driven approach. Inspired by the method of parallel distributed compensation, we construct the overall data-driven fault detection algorithm for the nonlinear NCS. The validation of the proposed algorithm is proved on the well-established Tennessee-Eastman process.

TITLE: Output tracking of boiler-turbine system by fuzzy guaranteed cost control

ABSTRACT: This paper proposes a controller design strategy of output tracking by guaranteed cost control (GCC) for Takagi-Sugeno (T-S) fuzzy systems and applies it to a boiler-turbine system. Firstly, based on the original T-S fuzzy system, an augmented system with integral action is established to eliminate steady tracking errors. Next, a state-feedback parallel distributed compensation (PDC) controller is designed such that the guaranteed cost function has an upper bound. Sufficient condition that guarantees quadratical stability of the resulting closed-loop system is derived by using a relaxed stability condition of T-S fuzzy systems. To obtain as good performance as possible, the design procedure is transformed into a linear matrix inequalities (LMIs) optimization problem and the upper bound of the guaranteed cost function is minimized. Finally, the simulation of output tracking control of a boiler-turbine system verifies the
effectiveness and feasibility of the proposed approach.

<table>
<thead>
<tr>
<th>PDS16NXT12</th>
<th>TITLE: Distributed cooperative control of microgrids with unbalance impedances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT: The power-sharing accuracy is highly sensitive to the output impedances of distributed generator (DG) inverters with parallel-connection capability at droop control method. The control strategy of autonomous microgrid based on coordinate rotational virtual impedance is proposed to handle those problem. The coordinate rotational virtual impedance loop is adopted to improve the impedance characteristic, to compensate the active and reactive power unbalance and eliminate the circulating current between each DG unit. Moreover, secondary control is to compensate the voltage deviation and frequency offset. As opposed to centralized control, the use of distributed cooperative control method greatly reduces the burden of system communications, improve system reliability and stability. Simulation results are presented to validate those methods.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PDS16NXT13</th>
<th>TITLE: Improved Holistic Analysis for Fork-Join Distributed Real-Time Tasks supported by the FTT-SE Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT: Modern distributed real-time embedded applications have high processing requirements associated with strict deadlines. For some applications, such constraints cannot be fulfilled by existing single-core embedded platforms. A solution is to parallelise the execution of the applications, by allowing networked nodes to distribute their workload to remote nodes with spare capacity. In that context, this paper presents a holistic timing analysis for fixedpriority fork-join Parallel/Distributed tasks (P/D tasks). Furthermore, we extend the holistic approach to consider the interaction between parallel threads and messages interchanged through a Flexible Time Triggered - Switched Ethernet (FTT-SE) network, and we show how the pessimism on the Worst-Case Response Time computation of such tasks can be reduced by considering the pipeline effect that occurs in such distributed systems. To evaluate the</td>
</tr>
</tbody>
</table>
performance and correctness of the holistic model, this paper includes a numerical evaluation based on a real automotive application. The obtained results show that the proposed method is effective in distributing the load by different nodes, allowing a significant reduction of the worst-case response time of the tasks. Moreover, the paper also reports an implementation of the model on a Linux library, called Parallel/Distributed Real-Time, as well as the corresponding results obtained on a real testbed. The obtained results are in accordance with the predictions of the holistic timing analysis.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> On emerging multicore systems, task replication is a powerful way to achieve high reliability targets. In this paper, we consider the problem of achieving a given reliability target for a set of periodic real-time tasks running on a multicore system with minimum energy consumption. Our framework explicitly takes into account the coverage factor of the fault detection techniques and the negative impact of Dynamic Voltage Scaling (DVS) on the rate of transient faults leading to soft errors. We characterize the subtle interplay between the processing frequency, replication level, reliability, fault coverage, and energy consumption on DVS-enabled multicore systems. We first develop static solutions and then propose dynamic adaptation schemes in order to reduce the concurrent execution of the replicas of a given task and to take advantage of early completions. Our simulation results indicate that through our algorithms, a very broad spectrum of reliability targets can be achieved with minimum energy consumption thanks to the judicious task replication and frequency assignment.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PDS16NXT15</th>
<th>TITLE: Distributed and parallel real-time control system equipped FPGA-Zynq and EPICS middleware</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT:</strong> Zynq series of Xilinx FPGA chips are divided into Processing System (PS) and Programmable Logic (PL), as a kind of SoC (System on Chip). PS with the dual-core ARM CortexA9 processor is performing the high-level control logic at</td>
<td></td>
</tr>
</tbody>
</table>
run-time on Linux operating system. PL with the low-level Field Programmable Gate Array (FPGA) built on high-performance, low-power, and high-k metal gate process technology is connecting with a lot of I/O peripherals for real-time control system. EPICS (Experimental Physics and Industrial Control System) is a set of open-source-based software tools which supports for the Ethernet-based middleware layer. In order to configure the environment of the distributed control system, EPICS middleware is equipped on the Linux operating system of the Zynq PS. In addition, a lot of digital logic gates of the Zynq PL of FPGA-Zynq evaluation board (ZedBoard) are connected with I/O pins of the daughter board via FPGA Mezzanine Connector (FMC) of ZedBoard. An interface between the Zynq PS and PL is interconnected with AMBA4 AXI. For the organic connection both the PS and PL, it also used the Linux device driver for AXI interface. This paper describes the content and configuration of the distributed and parallel real-time control system applying FPGA-Zynq and EPICS middleware.

PDS16NXT16

TITLE: Control of parallel DC-DC converters in a DC microgrid using virtual output impedance method

ABSTRACT: Microgrid allows flexible integration of distributed generation using renewable energy sources into the conventional centralized power system with the additional benefit of islanded mode operation. Research efforts on DC microgrids have gained momentum in the recent years, mainly due to its inherent advantages compared to AC microgrids. Several DC renewable energy sources are connected to a common DC bus in a DC microgrid using DC-DC converters. When converters of different ratings are connected in parallel, it is required that the load demand is shared by the converters proportional to their power ratings. This will ensure that individual converters are not overstressed and the connected load is optimally shared. This paper investigates a droop control strategy for parallel DC-DC converters using virtual output impedance method for optimum load sharing. Simulation of the control strategy has been done for two parallel connected buck converters and the results show the
<table>
<thead>
<tr>
<th>Title</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resilient distributed computing platforms for big data analysis using Spark and Hadoop</td>
<td>This paper introduces the integration of three platforms using Apache Hive, Cloudera Impala and BDAS Spark SQL which enables to support SQL-like queries in big data environment. In order to fast respond to user's query for big data processing, the optimized system can automatically select the appropriate platform to best perform a query. In addition, the rapid data retrieval from the in-memory cache or in-disk cache has achieved for the repeated SQL command. The proposed approach improves the efficiency of data retrieval significantly.</td>
</tr>
<tr>
<td>A Look at Basics of Distributed Computing</td>
<td>This tutorial presents concepts and basics of distributed computing which are important (at least from the author’s point of view!), and should be known and mastered by Master students, researchers, and engineers. Those include: (a) a characterization of distributed computing (which is too much often confused with parallel computing); (b) the notion of a synchronous system and its associated notions of a local algorithm and message adversaries; (c) the notion of an asynchronous shared memory system and its associated notions of universality and progress conditions; and (d) the notion of an asynchronous messagepassing system with its associated broadcast and agreement abstractions, its impossibility results, and approaches to circumvent them. Hence, the tutorial can be seen as a guided tour to key elements that constitute basics of distributed computing.</td>
</tr>
<tr>
<td>On utility optimization in distributed multiple access over a multi-packet reception channel</td>
<td>This paper considers distributed medium access control in a wireless multiple access network with an unknown number of users. A multi-</td>
</tr>
</tbody>
</table>
packet reception channel is assumed in which all packets should be received successfully if and only if the number of users transmitting in parallel does not exceed a known threshold. We propose a transmission adaptation approach, where, in each time slot, a user estimates the probability of channel availability from the viewpoint of a virtual user and adjusts its transmission probability according to its utility objective and a derived user number estimate. A sufficient condition under which the system should have a unique equilibrium is obtained. Simulation results show that the proposed medium access control algorithm does help users to converge asymptotically to a near optimal transmission probability.

**PDS16NXT20**

**TITLE:** Scaling of Distributed Multi-simulations on Multi-core Clusters

**ABSTRACT:** DACCOSIM is a multi-simulation environment for continuous time systems, relying on FMI standard, making easy the design of a multi-simulation graph, and specially developed for multi-core PC clusters, in order to achieve speedup and size up. However, the distribution of the simulation graph remains complex and is still the responsibility of the simulation developer. This paper introduces DACCOSIM parallel and distributed architecture, and our strategies to achieve efficient multi-simulation graph distribution on multi-core clusters. Some performance experiments on two clusters, running up to 81 simulation components (FMU) and using up to 16 multi-core computing nodes, are shown. Performances measured on our faster cluster exhibit a good scalability, but some limitations of current DACCOSIM implementation are discussed.

**PDS16NXT21**

**TITLE:** A Declarative Optimization Engine for Resource Provisioning of Scientific Workflows in Geo-Distributed Clouds

**ABSTRACT:** Geo-distributed clouds are becoming increasingly popular for cloud providers, and data centers with different regions often offer different prices, even for the same type of virtual machines. Resource provisioning in geo-distributed clouds is an important and complicated problem for budget and performance optimizations of scientific workflows. Scientists are facing the
complexities resulted from various cloud offerings in the geo-distributed settings, severe cloud performance dynamics and evolving user requirements on performance and cost. To address those complexities, we propose a declarative optimization engine named Geco for resource provisioning of scientific workflows in geo-distributed clouds. Geco allows users to specify their workflow optimization goals and constraints of specific problems with an extended declarative language. We propose a novel probabilistic optimization approach for evaluating the declarative optimization goals and constraints to address the cloud dynamics. Additionally, we develop runtime optimizations to more effectively utilize the cloud resources at runtime. To accelerate the solution finding, Geco leverages the power of GPUs to find the solution in a fast and timely manner. Our evaluations with four common workflow provisioning problems demonstrate that, Geco is able to achieve more effective performance/cost optimizations in geo-distributed cloud environments than the state-of-the-art approaches.

### PDS16NXT22

**TITLE:** A reconfigurable parallel FPGA accelerator for the adapt-then-combine diffusion LMS algorithm

**ABSTRACT:** The combination of diffusion strategies and least-mean-square (LMS) algorithm provides many advantages for adaptive-filter to solve distributed optimization, estimation and inference problems. However, suffering from high computation complexity, software implementation of diffusion LMS algorithm is unsuitable for real-time and portable applications. In order to extend its availability, we design a reconfigurable parallel FPGA accelerator by exploring multiple dimensions of parallelism, including: parallel execution of agents state updating, data combining, data training and multi-stages pipeline to speedup the execution time. The accelerator for networks with various number of agents and different input dimensions is implemented. Results demonstrate that, it can achieve a speedup of three orders of magnitude at 100Mhz compared with C implementation for a 32-nodes network with 16-dimensional
<table>
<thead>
<tr>
<th>Title</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>An SDN-Based Multipath GridFTP for High-Speed Data Transfer</td>
<td>We demonstrate high-speed data transfer GridFTP using a multipath control mechanism with SDN (Software-Defined Networking). GridFTP is a typical tool that has been developed and widely used for bulk data transfer over a wide area network in the field. GridFTP supports a parallel high-speed data transfer scheme using multiple TCP streams. However, one of the shortest paths is used solely for data transfer in the default IP routing while there are multiple network paths (multipath) exist between widely-distributed sites. In this study, we propose a system that distributes the parallel TCP streams of GridFTP into multiple network paths by a traffic engineering technique brought by SDN. Our proposed system has achieved approximately 20% better performance than the conventional method in the best case in a global-scale real environment.</td>
</tr>
<tr>
<td>A parallel primal-dual interior-point method for DC optimal power flow</td>
<td>Optimal power flow (OPF) seeks the operation point of a power system that maximizes a given objective function and satisfies operational and physical constraints. Given real-time operating conditions and the large scale of the power system, it is demanding to develop algorithms that allow for OPF to be decomposed and efficiently solved on parallel computing systems. In our work, we develop a parallel algorithm for applying the primal-dual interior point method to solve OPF. The primal-dual interior point method has a much faster convergence rate than gradient-based algorithms but requires solving a series of large, sparse linear systems. We design efficient parallelized and iterative methods to solve such linear systems which utilize the sparsity structure of the system matrix.</td>
</tr>
</tbody>
</table>
| PDS16NXT25 | TITLE: Parallel matrix neural network training on cluster systems for dynamic FET modeling from large datasets  

**ABSTRACT:** This paper presents a powerful and general parallel artificial neural network training technique with parallel computing on cluster systems. Large numbers of training samples are distributed to multiple computers on a cluster system to achieve a high speed-up for training. The method is evaluated with respect to two examples of an advanced dynamic nonlinear simulation model for GaN transistors where the training set is measured large-signal waveform data from an NVNA. For advanced models in a high dimensional space with large training sample size, the proposed approach is demonstrated to reduce the total training time by a factor of 35. |
TRAINING AND SERVICES

❖ Training By Corporate Trainers
❖ Research Environment
❖ Coding Explanation
❖ Will Provide Full Source Code Of The Project
❖ Full Documentation And Diagrams (Min 80 Pages)
❖ Some Of Q&A Based On Final Viva
❖ Journal Publications / Conference Publications
❖ Study Materials & Certification
❖ Printing & Binding.
❖ Aptitude and Soft Skill Training.
❖ 100% Placement Assistance.
❖ Full Time/Part Time Job Assistance During Your Project
❖ Own Project Idea Also Welcome