# Parallel and distributed system for java / dot net

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

### Synopsis:

Cloud data center management is a key problem due to the numerous and heterogeneous strategies that can be applied, ranging from the VM placement to the federation with other clouds. Performance evaluation of cloud computing infrastructures is required to predict and quantify the cost-benefit of a strategy portfolio and the corresponding quality of service (QoS) experienced by users. Such analyses are not feasible by simulation or on-the-field experimentation, due to the great number of parameters that have to be investigated. In this paper, we present an analytical model, based on stochastic reward nets (SRNs), that is both scalable to model systems composed of thousands of resources and flexible to represent different policies and cloud-specific strategies. Several performance metrics are defined and evaluated to analyze the behavior of a cloud data center: utilization, availability, waiting time, and responsiveness. A resiliency analysis is also provided to take into account load bursts. Finally, a general approach is presented that, starting from the concept of system capacity, can help system managers to opportunely set the data center parameters under different working conditions.

## 2. A Privacy Leakage Upper Bound Constraint-Based Approach for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud.

### Synopsis:

Cloud computing provides massive computation power and storage capacity which enable users to deploy computation and data-intensive applications without infrastructure investment. Along the processing of such applications, a large volume of intermediate data sets will be generated, and often stored to save the cost of recomputing them. However, preserving the privacy of intermediate data sets becomes a challenging problem because adversaries may recover privacy-sensitive information by analyzing multiple intermediate data sets. Encrypting ALL data sets in cloud is widely adopted in existing approaches to address this challenge. But we argue that encrypting all intermediate data sets are neither efficient nor cost-effective because it is very time consuming and costly for data-intensive applications to en/decrypt data sets frequently while performing any operation on them. In this paper, we propose a novel upper bound privacy leakage constraint-based approach to identify which intermediate data sets need to be encrypted and which do not, so that privacy-preserving cost can be saved while the privacy requirements of data holders can still be satisfied. Evaluation results demonstrate that the privacy-preserving cost of intermediate data sets can be significantly reduced with our approach over existing ones where all data sets are encrypted.

## 3. Capacity of Data Collection in Arbitrary Wireless Sensor Networks.

### Synopsis:

Data collection is a fundamental function provided by wireless sensor networks. How to efficiently collect sensing data from all sensor nodes is critical to the performance of sensor networks. In this paper, we aim to understand the theoretical limits of data collection in a TDMA-based sensor network in terms of possible and achievable maximum capacity. Previously, the study of data collection capacity has concentrated on large-scale random networks. However, in most of the practical sensor applications, the sensor network is not uniformly deployed and the number of sensors may not be as huge as in theory. Therefore, it is necessary to study the capacity of data collection in an arbitrary network. In this paper, we first derive the upper and lower bounds for data collection capacity in arbitrary networks under protocol interference and disk graph models. We show that a simple BFS tree-based method can lead to order-optimal performance for any arbitrary sensor networks. We then study the capacity bounds of data collection under a general graph model, where two nearby nodes may be unable to communicate due to barriers or path fading, and discuss performance implications. Finally, we provide discussions on the design of data collection under a physical interference model or a Gaussian channel model.

## 4. A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis.

### Synopsis:

Interconnected systems, such as Web servers, database servers, cloud computing servers and so on, are now under threads from network attackers. As one of most common and aggressive means, denial-of-service (DoS) attacks cause serious impact on these computing systems. In this paper, we present a DoS attack detection system that uses multivariate correlation analysis (MCA) for accurate network traffic characterization by extracting the geometrical correlations between network traffic features. Our MCA-based

DoS attack detection system employs the principle of anomaly based detection in attack recognition. This makes our solution capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic only. Furthermore, a triangle-area-based technique is proposed to enhance and to speed up the process of MCA. The effectiveness of our proposed detection system is evaluated using KDD Cup 99 data set, and the influences of both non-normalized data and normalized data on the performance of the proposed detection system are examined. The results show that our system outperforms two other previously developed state-of-the-art approaches in terms of detection accuracy.

## 5. Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment.

### Synopsis:

Cloud computing allows business customers to scale up and down their resource usage based on needs. Many of the touted gains in the cloud model come from resource multiplexing through virtualization technology. In this paper, we present a system that uses virtualization technology to allocate data center resources dynamically based on application demands and support green computing by optimizing the number of servers in use. We introduce the concept of "skewness” to measure the unevenness in the multidimensional resource utilization of a server. By minimizing skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources. We develop a set of heuristics that prevent overload in the system effectively while saving energy used. Trace driven simulation and experiment results demonstrate that our algorithm achieves good performance.

## 6. Harnessing the Cloud for Securely Outsourcing Large-Scale Systems of Linear Equations.

### Synopsis:

Cloud computing economically enables customers with limited computational resources to outsource large-scale computations to the cloud. However, how to protect customers' confidential data involved in the computations then becomes a major security concern. In this paper, we present a secure outsourcing mechanism for solving large-scale systems of linear equations (LE) in cloud. Because applying traditional approaches like Gaussian elimination or LU decomposition (aka. direct method) to such large-scale LEs would be prohibitively expensive, we build the secure LE outsourcing mechanism via a completely different approach-iterative method, which is much easier to implement in practice and only demands relatively simpler matrix-vector operations. Specifically, our mechanism enables a customer to securely harness the cloud for iteratively finding successive approximations to the LE solution, while keeping both the sensitive input and output of the computation private. For robust cheating detection, we further explore the algebraic property of matrix-vector operations and propose an efficient result verification mechanism, which allows the customer to verify all answers received from previous iterative approximations in one batch with high probability. Thorough security analysis and prototype experiments on Amazon EC2 demonstrate the validity and practicality of our proposed design.

## 7. A Two-Stage Deanonymization Attack against Anonymized Social Networks.

### Synopsis:

Digital traces left by users of online social networking services, even after anonymization, are susceptible to privacy breaches. This is exacerbated by the increasing overlap in user-bases among various services. To alert fellow researchers in both the academia and the industry to the feasibility of such an attack, we propose an algorithm, Seed-and-Grow, to identify users from an anonymized social graph, based solely on graph structure. The algorithm first identifies a seed subgraph, either planted by an attacker or divulged by a collusion of a small group of users, and then grows the seed larger based on the attacker's existing knowledge of the users' social relations. Our work identifies and relaxes implicit assumptions taken by previous works, eliminates arbitrary parameters, and improves identification effectiveness and accuracy. Simulations on real-world collected data sets verify our claim.

## 8. Efficient Two-Server Password-Only Authenticated Key Exchange.

### Synopsis:

Password-authenticated key exchange (PAKE) is where a client and a server, who share a password, authenticate each other and meanwhile establish a cryptographic key by exchange of messages. In this setting, all the passwords necessary to authenticate clients are stored in a single server. If the server is compromised, due to, for example, hacking or even insider attack, passwords stored in the server are all disclosed. In this paper, we consider a scenario where two servers cooperate to authenticate a client and if one server is compromised, the attacker still cannot pretend to be the client with the information from the compromised server. Current solutions for two-server PAKE are either symmetric in the sense that two peer servers equally contribute to the authentication or asymmetric in the sense that one server authenticates the client with the help of another server. This paper presents a symmetric solution for two-server PAKE, where the client can establish different cryptographic keys with the two servers, respectively. Our protocol runs in parallel and is more efficient than existing symmetric two-server PAKE protocol, and even more efficient than existing asymmetric two-server PAKE protocols in terms of parallel computation.

## 9. Payments for Outsourced Computations

### Synopsis:

With the recent advent of cloud computing, the concept of outsourcing computations, initiated by volunteer computing efforts, is being revamped. While the two paradigms differ in several dimensions, they also share challenges, stemming from the lack of trust between outsourcers and workers. In this work, we propose a unifying trust framework, where correct participation is financially rewarded: neither participant is trusted, yet outsourced computations are efficiently verified and validly remunerated. We propose three solutions for this problem, relying on an offline bank to generate and redeem payments; the bank is oblivious to interactions between outsourcers and workers. We propose several attacks that can be launched against our framework and study the effectiveness of our solutions. We implemented our most secure solution and our experiments show that it is efficient: the bank can perform hundreds of payment transactions per second and the overheads imposed on outsourcers and workers are negligible.

## 10. Application-Aware Local-Global Source Deduplication for Cloud Backup Services of Personal Storage.

### Synopsis:

In personal computing devices that rely on a cloud storage environment for data backup, an imminent challenge facing source deduplication for cloud backup services is the low deduplication efficiency due to a combination of the resource-intensive nature of deduplication and the limited system resources. In this paper, we present ALG-Dedupe, an Application-aware Local-Global source deduplication scheme that improves data deduplication efficiency by exploiting application awareness, and further combines local and global duplicate detection to strike a good balance between cloud storage capacity saving and deduplication time reduction. We perform experiments via prototype implementation to demonstrate that our scheme can significantly improve deduplication efficiency over the state-of-the-art methods with low system overhead, resulting in shortened backup window, increased power efficiency and reduced cost for cloud backup services of personal storage.

## 11. Enabling Dynamic Data and Indirect Mutual Trust for Cloud Computing Storage Systems.

### Synopsis:

Storage-as-a-service offered by cloud service providers (CSPs) is a paid facility that enables organizations to outsource their sensitive data to be stored on remote servers. In this paper, we propose a cloud-based storage scheme that allows the data owner to benefit from the facilities offered by the CSP and enables indirect mutual trust between them. The proposed scheme has four important features: 1) it allows the owner to outsource sensitive data to a CSP, and perform full block-level dynamic operations on the outsourced data, i.e., block modification, insertion, deletion, and append, 2) it ensures that authorized users (i.e., those who have the right to access the owner's file) receive the latest version of the outsourced data, 3) it enables indirect mutual trust between the owner and the CSP, and 4) it allows the owner to grant or revoke access to the outsourced data. We discuss the security issues of the proposed scheme. Besides, we justify its performance through theoretical analysis and a prototype implementation on Amazon cloud platform to evaluate storage, communication, and computation overheads.

## 12. Privacy-Preserving Decentralized Key-Policy Attribute-Based Encryption.

### Synopsis:

Decentralized attribute-based encryption (ABE) is a variant of a multiauthority ABE scheme where each authority can issue secret keys to the user independently without any cooperation and a central authority. This is in contrast to the previous constructions, where multiple authorities must be online and setup the system interactively, which is impractical. Hence, it is clear that a decentralized ABE scheme eliminates the heavy communication cost and the need for collaborative computation in the setup stage. Furthermore, every authority can join or leave the system freely without the necessity of reinitializing the system. In contemporary multiauthority ABE schemes, a user's secret keys from different authorities must be tied to his global identifier (GID) to resist the collusion attack. However, this will compromise the user's privacy. Multiple authorities can collaborate to trace the user by his GID, collect his attributes, then impersonate him. Therefore, constructing a decentralized ABE scheme with privacy-preserving remains a challenging research problem. In this paper, we propose a privacy-preserving decentralized key-policy ABE scheme where each authority can issue secret keys to a user independently without knowing anything about his GID. Therefore, even if multiple authorities are corrupted, they cannot collect the user's attributes by tracing his GID. Notably, our scheme only requires standard complexity assumptions (e.g., decisional bilinear Diffie-Hellman) and does not require any cooperation between the multiple authorities, in contrast to the previous comparable scheme that requires nonstandard complexity assumptions (e.g., q-decisional Diffie-Hellman inversion) and interactions among multiple authorities. To the best of our knowledge, it is the first decentralized ABE scheme with privacy-preserving based on standard complexity assumptions.

## 13. Asymmetric Social Proximity Based Private Matching Protocols for Online Social Networks.

### Synopsis:

The explosive growth of Online Social Networks (OSNs) over the past few years has redefined the way people interact with existing friends and especially make new friends. Some works propose to let people become friends if they have similar profile attributes. However, profile matching involves an inherent privacy risk of exposing private profile information to strangers in the cyberspace. The existing solutions to the problem attempt to protect users' privacy by privately computing the intersection or intersection cardinality of the profile attribute sets of two users. These schemes have some limitations and can still reveal users' privacy. In this paper, we leverage community structures to redefine the OSN model and propose a realistic asymmetric social proximity measure between two users. Then, based on the proposed asymmetric social proximity, we design three private matching protocols, which provide different privacy levels and can protect users' privacy better than the previous works. We also analyze the computation and communication cost of these protocols. Finally, we validate our proposed asymmetric proximity measure using real social network data and conduct extensive simulations to evaluate the performance of the proposed protocols in terms of computation cost, communication cost, total running time, and energy consumption. The results show the efficacy of our proposed proximity measure and better performance of our protocols over the state-of-the-art protocols.

.

## 14. Error-Tolerant Resource Allocation and Payment Minimization for Cloud System.

### Synopsis:

With virtual machine (VM) technology being increasingly mature, compute resources in cloud systems can be partitioned in fine granularity and allocated on demand. We make three contributions in this paper: 1) We formulate a deadline-driven resource allocation problem based on the cloud environment facilitated with VM resource isolation technology, and also propose a novel solution with polynomial time, which could minimize users' payment in terms of their expected deadlines. 2) By analyzing the upper bound of task execution length based on the possibly inaccurate workload prediction, we further propose an error-tolerant method to guarantee task's completion within its deadline. 3) We validate its effectiveness over a real VM-facilitated cluster environment under different levels of competition. In our experiment, by tuning algorithmic input deadline based on our derived bound, task execution length can always be limited within its deadline in the sufficient-supply situation; the mean execution length still keeps 70 percent as high as user-specified deadline under the severe competition. Under the original-deadline-based solution, about 52.5 percent of tasks are completed within 0.95-1.0 as high as their deadlines, which still conforms to the deadline-guaranteed requirement. Only 20 percent of tasks violate deadlines, yet most (17.5 percent) are still finished within 1.05 times of deadlines.

## 15. Community home-based multi-copy routing in mobile social networks.

### Synopsis:

A mobile social network (MSN) is a special delay tolerant network (DTN) composed of mobile nodes with social characteristics. Mobile nodes in MSNs generally visit community homes frequently, while other locations are visited less frequently. We propose a novel zero-knowledge MSN routing algorithm, homing spread (HS). The community homes have a higher priority to spread messages into the network. Theoretical analysis shows that the proposed algorithm can spread a given number of message copies in an optimal way when the inter-meeting times between any two nodes and between a node and a community home follow exponential distributions. We also calculate the expected delivery delay of HS. In addition, extensive simulations are conducted. Results show that community homes are important factors in efficient message spreading. By using homes to spread messages faster, HS achieves a better performance than existing zero-knowledge MSN routing algorithms, including Epidemic, with a given number of copies, and Spray&Wait.

## 16. Geocommunity-Based Broadcasting for Data Dissemination in Mobile Social Networks.

### Synopsis:

In this paper, we consider the issue of data broadcasting in mobile social networks (MSNets). The objective is to broadcast data from a superuser to other users in the network. There are two main challenges under this paradigm, namely 1) how to represent and characterize user mobility in realistic MSNets; 2) given the knowledge of regular users' movements, how to design an efficient superuser route to broadcast data actively. We first explore several realistic data sets to reveal both geographic and social regularities of human mobility, and further propose the concepts of geocommunity and geocentrality into MSNet analysis. Then, we employ a semi-Markov process to model user mobility based on the geocommunity structure of the network. Correspondingly, the geocentrality indicating the “dynamic user density” of each geocommunity can be derived from the semi-Markov model. Finally, considering the geocentrality information, we provide different route algorithms to cater to the superuser that wants to either minimize total duration or maximize dissemination ratio. To the best of our knowledge, this work is the first to study data broadcasting in a realistic MSNet setting. Extensive trace-driven simulations show that our approach consistently outperforms other existing superuser route design algorithms in terms of dissemination ratio and energy efficiency.

## 17. Distributed, Concurrent, and Independent Access to Encrypted Cloud Databases.

### Synopsis:

Placing critical data in the hands of a cloud provider should come with the guarantee of security and availability for data at rest, in motion, and in use. Several alternatives exist for storage services, while data confidentiality solutions for the database as a service paradigm are still immature. We propose a novel architecture that integrates cloud database services with data confidentiality and the possibility of executing concurrent operations on encrypted data. This is the first solution supporting geographically distributed clients to connect directly to an encrypted cloud database, and to execute concurrent and independent operations including those modifying the database structure. The proposed architecture has the further advantage of eliminating intermediate proxies that limit the elasticity, availability, and scalability properties that are intrinsic in cloud-based solutions. The efficacy of the proposed architecture is evaluated through theoretical analyses and extensive experimental results based on a prototype implementation subject to the TPC-C standard benchmark for different numbers of clients and network latencies.

## 18. Identity-Based Secure DistributedData Storage Schemes.

### Synopsis:

Secure distributed data storage can shift the burden of maintaining a large number of files from the owner to proxy servers. Proxy servers can convert encrypted files for the owner to encrypted files for the receiver without the necessity of knowing the content of the original files. In practice, the original files will be removed by the owner for the sake of space efficiency. Hence, the issues on confidentiality and integrity of the outsourced data must be addressed carefully. In this paper, we propose two identity-based secure distributed data storage (IBSDDS) schemes. Our schemes can capture the following properties: (1) The file owner can decide the access permission independently without the help of the private key generator (PKG); (2) For one query, a receiver can only access one file, instead of all files of the owner; (3) Our schemes are secure against the collusion attacks, namely even if the receiver can compromise the proxy servers, he cannot obtain the owner's secret key. Although the first scheme is only secure against the chosen plaintext attacks (CPA), the second scheme is secure against the chosen ciphertext attacks (CCA). To the best of our knowledge, it is the first IBSDDS schemes where an access permission is made by the owner for an exact file and collusion attacks can be protected in the standard model.

## 19. Expressive, Efficient, and Revocable Data Access Control for Multi-Authority Cloud Storage.

### Synopsis:

Data access control is an effective way to ensure the data security in the cloud. Due to data outsourcing and untrusted cloud servers, the data access control becomes a challenging issue in cloud storage systems. Ciphertext-Policy Attribute-based Encryption (CP-ABE) is regarded as one of the most suitable technologies for data access control in cloud storage, because it gives data owners more direct control on access policies. However, it is difficult to directly apply existing CP-ABE schemes to data access control for cloud storage systems because of the attribute revocation problem. In this paper, we design an expressive, efficient and revocable data access control scheme for multi-authority cloud storage systems, where there are multiple authorities co-exist and each authority is able to issue attributes independently. Specifically, we propose a revocable multi-authority CP-ABE scheme, and apply it as the underlying techniques to design the data access control scheme. Our attribute revocation method can efficiently achieve both forward security and backward security. The analysis and simulation results show that our proposed data access control scheme is secure in the random oracle model and is more efficient than previous works.

## 20. IP-Geolocation Mapping for Moderately Connected Internet Regions.

### Synopsis:

Most IP-geolocation mapping schemes [14], [16], [17], [18] take delay-measurement approach, based on the assumption of a strong correlation between networking delay and geographical distance between the targeted client and the landmarks. In this paper, however, we investigate a large region of moderately connected Internet and find the delay-distance correlation is weak. But we discover a more probable rule - with high probability the shortest delay comes from the closest distance. Based on this closest-shortest rule, we develop a simple and novel IP-geolocation mapping scheme for moderately connected Internet regions, called GeoGet. In GeoGet, we take a large number of webservers as passive landmarks and map a targeted client to the geolocation of the landmark that has the shortest delay. We further use JavaScript at targeted clients to generate HTTP/Get probing for delay measurement. To control the measurement cost, we adopt a multistep probing method to refine the geolocation of a targeted client, finally to city level. The evaluation results show that when probing about 100 landmarks, GeoGet correctly maps 35.4 percent clients to city level, which outperforms current schemes such as GeoLim [16] and GeoPing [14] by 270 and 239 percent, respectively, and the median error distance in GeoGet is around 120 km, outperforming GeoLim and GeoPing by 37 and 70 percent, respectively.

## 21. Hop-by-Hop Message Authenticationand Source Privacy in WirelessSensor Networks.

### Synopsis:

Message authentication is one of the most effective ways to thwart unauthorized and corrupted messages from being forwarded in wireless sensor networks (WSNs). For this reason, many message authentication schemes have been developed, based on either symmetric-key cryptosystems or public-key cryptosystems. Most of them, however, have the limitations of high computational and communication overhead in addition to lack of scalability and resilience to node compromise attacks. To address these issues, a polynomial-based scheme was recently introduced. However, this scheme and its extensions all have the weakness of a built-in threshold determined by the degree of the polynomial: when the number of messages transmitted is larger than this threshold, the adversary can fully recover the polynomial. In this paper, we propose a scalable authentication scheme based on elliptic curve cryptography (ECC). While enabling intermediate nodes authentication, our proposed scheme allows any node to transmit an unlimited number of messages without suffering the threshold problem. In addition, our scheme can also provide message source privacy. Both theoretical analysis and simulation results demonstrate that our proposed scheme is more efficient than the polynomial-based approach in terms of computational and communication overhead under comparable security levels while providing message source privacy.

## 22. Nothing is for Free: Security in Searching Shared and Encrypted Data.

### Synopsis:

Most existing symmetric searchable encryption schemes aim at allowing a user to outsource her encrypted data to a cloud server and delegate the latter to search on her behalf. These schemes do not qualify as a secure and scalable solution for the multiparty setting, where users outsource their encrypted data to a cloud server and selectively authorize each other to search. Due to the possibility that the cloud server may collude with some malicious users, it is a challenge to have a secure and scalable multiparty searchable encryption (MPSE) scheme. This is shown by our analysis on the Popa-Zeldovich scheme, which says that an honest user may leak all her search patterns even if she shares only one of her documents with another malicious user. Based on our analysis, we present a new security model for MPSE by considering the worst case and average-case scenarios, which capture different server-user collusion possibilities. We then propose a MPSE scheme by employing the bilinear property of Type-3 pairings and prove its security based on the bilinear Diffie-Hellman variant and symmetric external Diffie-Hellman assumptions in the random oracle model.

## 23. On the Security of a Public Auditing Mechanism for Shared Cloud Data Service.

### Synopsis:

Recently, a public auditing protocol for shared data called Panda (IEEE Transactions on Services Computing, doi: 10.1109/TSC.2013.2295611) was proposed to ensure the correctness of the outsourced data. A distinctive feature of Panda is the support of data sharing and user revocation. Unfortunately, in this letter, we show that Panda is insecure in the sense that a cloud server can hide data loss without being detected. Specifically, we show that even some stored file blocks have been lost, the server is able to generate a valid proof by replacing a pair of lost data block and its signature with another block and signature pair. We also provide a solution to the problem while preserving all the desirable features of the original protocol.

## 24. On the Security of a Public Auditing Mechanism for Shared Cloud Data Service.

### Synopsis:

Recently, a public auditing protocol for shared data called Panda (IEEE Transactions on Services Computing, doi: 10.1109/TSC.2013.2295611) was proposed to ensure the correctness of the outsourced data. A distinctive feature of Panda is the support of data sharing and user revocation. Unfortunately, in this letter, we show that Panda is insecure in the sense that a cloud server can hide data loss without being detected. Specifically, we show that even some stored file blocks have been lost, the server is able to generate a valid proof by replacing a pair of lost data block and its signature with another block and signature pair. We also provide a solution to the problem while preserving all the desirable features of the original protocol.

## 25.Label-Embedding for Attribute-Based Classification

Attributes are an intermediate representation, which enables parameter sharing between classes, a must when training data is scarce. We propose to view attribute-based image classification as a label-embedding problem: each class is embedded in the space of attribute vectors. We introduce a function which measures the compatibility between an image and a label embedding. The parameters of this function are learned on a training set of labeled samples to ensure that, given an image, the correct classes rank higher than the incorrect ones. Results on the Animals With Attributes and Caltech-UCSD-Birds datasets show that the proposed framework outperforms the standard Direct Attribute Prediction baseline in a zero-shot learning scenario. The label embedding framework offers other advantages such as the ability to leverage alternative sources of information in addition to attributes (e.g. class hierarchies) or to transition smoothly from zero-shot learning to learning with large quantities of data.

## 26. Load Rebalancing for Distributed File Systems in Clouds.

### Synopsis:

Distributed file systems are key building blocks for cloud computing applications based on the MapReduce programming paradigm. In such file systems, nodes simultaneously serve computing and storage functions; a file is partitioned into a number of chunks allocated in distinct nodes so that MapReduce tasks can be performed in parallel over the nodes. However, in a cloud computing environment, failure is the norm, and nodes may be upgraded, replaced, and added in the system. Files can also be dynamically created, deleted, and appended. This results in load imbalance in a distributed file system; that is, the file chunks are not distributed as uniformly as possible among the nodes. Emerging distributed file systems in production systems strongly depend on a central node for chunk reallocation. This dependence is clearly inadequate in a large-scale, failure-prone environment because the central load balancer is put under considerable workload that is linearly scaled with the system size, and may thus become the performance bottleneck and the single point of failure. In this paper, a fully distributed load rebalancing algorithm is presented to cope with the load imbalance problem. Our algorithm is compared against a centralized approach in a production system and a competing distributed solution presented in the literature. The simulation results indicate that our proposal is comparable with the existing centralized approach and considerably outperforms the prior distributed algorithm in terms of load imbalance factor, movement cost, and algorithmic overhead. The performance of our proposal implemented in the Hadoop distributed file system is further investigated in a cluster environment.

## 27. Message in a Sealed Bottle: Privacy Preserving Friending in Social Networks.

### Synopsis:

Many proximity-based mobile social networks are developed to facilitate connections between any two people, or to help a user to find people with a matched profile within a certain distance. A challenging task in these applications is to protect the privacy of the participants' profiles and personal interests. In this paper, we design novel mechanisms, when given a preference-profile submitted by a user, that search persons with matching-profile in decentralized multi-hop mobile social networks. Our mechanisms also establish a secure communication channel between the initiator and matching users at the time when the matching user is found. Our rigorous analysis shows that our mechanism is privacy-preserving (no participants' profile and the submitted preference-profile are exposed), verifiable (both the initiator and the unmatched user cannot cheat each other to pretend to be matched), and efficient in both communication and computation. Extensive evaluations using real social network data, and actual system implementation on smart phones show that our mechanisms are significantly more efficient than existing solutions.

## 28. Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud

### Synopsis:

With the character of low maintenance, cloud computing provides an economical and efficient solution for sharing group resource among cloud users. Unfortunately, sharing data in a multi-owner manner while preserving data and identity privacy from an untrusted cloud is still a challenging issue, due to the frequent change of the membership. In this paper, we propose a secure multi-owner data sharing scheme, named Mona, for dynamic groups in the cloud. By leveraging group signature and dynamic broadcast encryption techniques, any cloud user can anonymously share data with others. Meanwhile, the storage overhead and encryption computation cost of our scheme are independent with the number of revoked users. In addition, we analyze the security of our scheme with rigorous proofs, and demonstrate the efficiency of our scheme in experiments.

## 29. Securing Broker-Less Publish/Subscribe Systems Using Identity-Based Encryption.

### Synopsis:

The provisioning of basic security mechanisms such as authentication and confidentiality is highly challenging in a content-based publish/subscribe system. Authentication of publishers and subscribers is difficult to achieve due to the loose coupling of publishers and subscribers. Likewise, confidentiality of events and subscriptions conflicts with content-based routing. This paper presents a novel approach to provide confidentiality and authentication in a broker-less content-based publish/subscribe system. The authentication of publishers and subscribers as well as confidentiality of events is ensured, by adapting the pairing-based cryptography mechanisms, to the needs of a publish/subscribe system. Furthermore, an algorithm to cluster subscribers according to their subscriptions preserves a weak notion of subscription confidentiality. In addition to our previous work , this paper contributes 1) use of searchable encryption to enable efficient routing of encrypted events, 2) multicredential routing a new event dissemination strategy to strengthen the weak subscription confidentiality, and 3) thorough analysis of different attacks on subscription confidentiality. The overall approach provides fine-grained key management and the cost for encryption, decryption, and routing is in the order of subscribed attributes. Moreover, the evaluations show that providing security is affordable w.r.t. 1) throughput of the proposed cryptographic primitives, and 2) delays incurred during the construction of the publish/subscribe overlay and the event dissemination.

## 30. Shared Authority Based Privacy-Preserving Authentication Protocol in Cloud Computing.

### Synopsis:

Cloud computing is an emerging data interactive paradigm to realize users' data remotely stored in an online cloud server. Cloud services provide great conveniences for the users to enjoy the on-demand cloud applications without considering the local infrastructure limitations. During the data accessing, different users may be in a collaborative relationship, and thus data sharing becomes significant to achieve productive benefits. The existing security solutions mainly focus on the authentication to realize that a user's privative data cannot be illegally accessed, but neglect a subtle privacy issue during a user challenging the cloud server to request other users for data sharing. The challenged access request itself may reveal the user's privacy no matter whether or not it can obtain the data access permissions. In this paper, we propose a shared authority based privacy-preserving authentication protocol (SAPA) to address above privacy issue for cloud storage. In the SAPA, 1) shared access authority is achieved by anonymous access request matching mechanism with security and privacy considerations (e.g., authentication, data anonymity, user privacy, and forward security); 2) attribute based access control is adopted to realize that the user can only access its own data fields; 3) proxy re-encryption is applied to provide data sharing among the multiple users. Meanwhile, universal composability (UC) model is established to prove that the SAPA theoretically has the design correctness. It indicates that the proposed protocol is attractive for multi-user collaborative cloud applications.

## 31. SocialTube: P2P-assisted video sharing in online social networks.

### Synopsis:

Video sharing has been an increasingly popular application in online social networks (OSNs). However, its sustainable development is severely hindered by the intrinsic limit of the client/server architecture deployed in current OSN video systems, which is not only costly in terms of server bandwidth and storage but also not scalable. The peer-assisted Video-on-Demand (VOD) technique, in which participating peers assist the server in delivering video content has been proposed recently. Unfortunately, videos can only be disseminated through friends in OSNs. Therefore, current VOD works that explore clustering nodes with similar interests or close location for high performance are suboptimal, if not entirely inapplicable, in OSNs. Based on our long-term real-world measurement of over 1,000,000 users and 2,500 videos in Facebook, we propose SocialTube, a novel peer-assisted video sharing system that explores social relationship, interest similarity, and physical location between peers in OSNs. Specifically, SocialTube incorporates three algorithms: a social network (SN)-based P2P overlay construction algorithm, a SN-based chunk prefetch algorithm, and a buffer management algorithm. The trace driven based simulation results show that SocialTube can improve the quality of user experience and system scalability over current P2P VOD techniques.

## 32. Optimal Multiserver Configuration for Profit Maximization in Cloud Computing.

### Synopsis:

As cloud computing becomes more and more popular, understanding the economics of cloud computing becomes critically important. To maximize the profit, a service provider should understand both service charges and business costs, and how they are determined by the characteristics of the applications and the configuration of a multiserver system. The problem of optimal multiserver configuration for profit maximization in a cloud computing environment is studied. Our pricing model takes such factors into considerations as the amount of a service, the workload of an application environment, the configuration of a multiserver system, the service-level agreement, the satisfaction of a consumer, the quality of a service, the penalty of a low-quality service, the cost of renting, the cost of energy consumption, and a service provider's margin and profit. Our approach is to treat a multiserver system as an M/M/m queuing model, such that our optimization problem can be formulated and solved analytically. Two server speed and power consumption models are considered, namely, the idle-speed model and the constant-speed model. The probability density function of the waiting time of a newly arrived service request is derived. The expected service charge to a service request is calculated. The expected net business gain in one unit of time is obtained. Numerical calculations of the optimal server size and the optimal server speed are demonstrated.

## 33. Priority-Based Consolidation of Parallel Workloads in the Cloud.

### Synopsis:

The cloud computing paradigm is attracting an increased number of complex applications to run in remote data centers. Many complex applications require parallel processing capabilities. Parallel applications of certain nature often show a decreasing utilization of CPU resources as parallelism grows, mainly because of the communication and synchronization among parallel processes. It is challenging but important for a data center to achieve a certain level of utilization of its nodes while maintaining the level of responsiveness of parallel jobs. Existing parallel scheduling mechanisms normally take responsiveness as the top priority and need nontrivial effort to make them work for data centers in the cloud era. In this paper, we propose a priority-based method to consolidate parallel workloads in the cloud. We leverage virtualization technologies to partition the computing capacity of each node into two tiers, the foreground virtual machine (VM) tier (with high CPU priority) and the background VM tier (with low CPU priority). We provide scheduling algorithms for parallel jobs to make efficient use of the two tier VMs to improve the responsiveness of these jobs. Our extensive experiments show that our parallel scheduling algorithm significantly outperforms commonly used algorithms such as extensible argonne scheduling system in a data center setting. The method is practical and effective for consolidating parallel workload in data centers.

## 34. QoS Ranking Prediction for Cloud Services.

### Synopsis:

Cloud computing is becoming popular. Building high-quality cloud applications is a critical research problem. QoS rankings provide valuable information for making optimal cloud service selection from a set of functionally equivalent service candidates. To obtain QoS values, real-world invocations on the service candidates are usually required. To avoid the time-consuming and expensive real-world service invocations, this paper proposes a QoS ranking prediction framework for cloud services by taking advantage of the past service usage experiences of other consumers. Our proposed framework requires no additional invocations of cloud services when making QoS ranking prediction. Two personalized QoS ranking prediction approaches are proposed to predict the QoS rankings directly. Comprehensive experiments are conducted employing real-world QoS data, including 300 distributed users and 500 real-world web services all over the world. The experimental results show that our approaches outperform other competing approaches.

## 35. An Ontology-Based Hybrid Approach to Activity Modeling for Smart Homes.

### Synopsis:

Activity models play a critical role for activity recognition and assistance in ambient assisted living. Existing approaches to activity modeling suffer from a number of problems, e.g., cold-start, model reusability, and incompleteness. In an effort to address these problems, we introduce an ontology-based hybrid approach to activity modeling that combines domain knowledge based model specification and data-driven model learning. Central to the approach is an iterative process that begins with “seed” activity models created by ontological engineering. The “seed” models are deployed, and subsequently evolved through incremental activity discovery and model update. While our previous work has detailed ontological activity modeling and activity recognition, this paper focuses on the systematic hybrid approach and associated methods and inference rules for learning new activities and user activity profiles. The approach has been implemented in a feature-rich assistive living system. Analysis of the experiments conducted has been undertaken in an effort to test and evaluate the activity learning algorithms and associated mechanisms.

## 36. A Highly Scalable Key Pre-Distribution Scheme for Wireless Sensor Networks.

### Synopsis:

Given the sensitivity of the potential WSN applications and because of resource limitations, key management emerges as a challenging issue for WSNs. One of the main concerns when designing a key management scheme is the network scalability. Indeed, the protocol should support a large number of nodes to enable a large scale deployment of the network. In this paper, we propose a new scalable key management scheme for WSNs which provides a good secure connectivity coverage. For this purpose, we make use of the unital design theory. We show that the basic mapping from unitals to key pre-distribution allows us to achieve high network scalability. Nonetheless, this naive mapping does not guarantee a high key sharing probability. Therefore, we propose an enhanced unital-based key pre-distribution scheme providing high network scalability and good key sharing probability approximately lower bounded by 1-e-1 ≈ 0.632. We conduct approximate analysis and simulations and compare our solution to those of existing methods for different criteria such as storage overhead, network scalability, network connectivity, average secure path length and network resiliency. Our results show that the proposed approach enhances the network scalability while providing high secure connectivity coverage and overall improved performance. Moreover, for an equal network size, our solution reduces significantly the storage overhead compared to those of existing solutions.

## 37. A Novel Anti phishing framework based on visual cryptography.

### Synopsis:

With the advent of internet, various online attacks has been increased and among them the most popular attack is phishing. Phishing is an attempt by an individual or a group to get personal confidential information such as passwords, credit card information from unsuspecting victims for identity theft, financial gain and other fraudulent activities. Fake websites which appear very similar to the original ones are being hosted to achieve this. In this paper we have proposed a new approach named as "A Novel Anti-phishing framework based on visual cryptography "to solve the problem of phishing. Here an image based authentication using Visual Cryptography is implemented. The use of visual cryptography is explored to preserve the privacy of an image captcha by decomposing the original image captcha into two shares (known as sheets) that are stored in separate database servers(one with user and one with server) such that the original image captcha can be revealed only when both are simultaneously available; the individual sheet images do not reveal the identity of the original image captcha. Once the original image captcha is revealed to the user it can be used as the password. Using this website cross verifies its identity and proves that it is a genuine website before the end users.

## 38. Computing Localized Power-Efficient Data Aggregation Trees for Sensor Networks.

### Synopsis:

We propose localized, self organizing, robust, and energy-efficient data aggregation tree approaches for sensor networks, which we call Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs). They are based on topologies, such as LMST and RNG, that can approximate minimum spanning tree and can be efficiently computed using only position or distance information of one-hop neighbors. The actual routing tree is constructed over these topologies. We also consider different parent selection strategies while constructing a routing tree. We compare each topology and parent selection strategy and conclude that the best among them is the shortest path strategy over LMST structure. Our solution also involves route maintenance procedures that will be executed when a sensor node fails or a new node is added to the network. The proposed solution is also adapted to consider the remaining power levels of nodes in order to increase the network lifetime. Our simulation results show that by using our power-aware localized approach, we can almost have the same performance of a centralized solution in terms of network lifetime, and close to 90 percent of an upper bound derived here.

## 39. Behavior Rule Specification-Based Intrusion Detection for Safety Critical Medical Cyber Physical Systems.

### Synopsis:

We propose and analyze a behavior-rule specification-based technique for intrusion detection of medical devices embedded in a medical cyber physical system (MCPS) in which the patient's safety is of the utmost importance. We propose a methodology to transform behavior rules to a state machine, so that a device that is being monitored for its behavior can easily be checked against the transformed state machine for deviation from its behavior specification. Using vital sign monitor medical devices as an example, we demonstrate that our intrusion detection technique can effectively trade false positives off for a high detection probability to cope with more sophisticated and hidden attackers to support ultra safe and secure MCPS applications. Moreover, through a comparative analysis, we demonstrate that our behavior-rule specification-based IDS technique outperforms two existing anomaly-based techniques for detecting abnormal patient behaviors in pervasive healthcare applications.

## 40. Mobi-Sync: Efficient Time Synchronization for Mobile Underwater Sensor Networks.

### Synopsis:

Time synchronization is an important requirement for many services provided by distributed networks. A lot of time synchronization protocols have been proposed for terrestrial Wireless Sensor Networks (WSNs). However, none of them can be directly applied to Underwater Sensor Networks (UWSNs). A synchronization algorithm for UWSNs must consider additional factors such as long propagation delays from the use of acoustic communication and sensor node mobility. These unique challenges make the accuracy of synchronization procedures for UWSNs even more critical. Time synchronization solutions specifically designed for UWSNs are needed to satisfy these new requirements. This paper proposes Mobi-Sync, a novel time synchronization scheme for mobile underwater sensor networks. Mobi-Sync distinguishes itself from previous approaches for terrestrial WSN by considering spatial correlation among the mobility patterns of neighboring UWSNs nodes. This enables Mobi-Sync to accurately estimate the long dynamic propagation delays. Simulation results show that Mobi-Sync outperforms existing schemes in both accuracy and energy efficiency.

## 41. SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency.

### Synopsis:

With the pervasiveness of smart phones and the advance of wireless body sensor networks (BSNs), mobile Healthcare (m-Healthcare), which extends the operation of Healthcare provider into a pervasive environment for better health monitoring, has attracted considerable interest recently. However, the flourish of m-Healthcare still faces many challenges including information security and privacy preservation. In this paper, we propose a secure and privacy-preserving opportunistic computing framework, called SPOC, for m-Healthcare emergency. With SPOC, smart phone resources including computing power and energy can be opportunistically gathered to process the computing-intensive personal health information (PHI) during m-Healthcare emergency with minimal privacy disclosure. In specific, to leverage the PHI privacy disclosure and the high reliability of PHI process and transmission in m-Healthcare emergency, we introduce an efficient user-centric privacy access control in SPOC framework, which is based on an attribute-based access control and a new privacy-preserving scalar product computation (PPSPC) technique, and allows a medical user to decide who can participate in the opportunistic computing to assist in processing his overwhelming PHI data. Detailed security analysis shows that the proposed SPOC framework can efficiently achieve user-centric privacy access control in m-Healthcare emergency. In addition, performance evaluations via extensive simulations demonstrate the SPOC's effectiveness in term of providing high-reliable-PHI process and transmission while minimizing the privacy disclosure during m-Healthcare emergency.

## 42. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing.

### Synopsis:

Cloud Computing has been envisioned as the next-generation architecture of IT Enterprise. It moves the application software and databases to the centralized large data centers, where the management of the data and services may not be fully trustworthy. This unique paradigm brings about many new security challenges, which have not been well understood. This work studies the problem of ensuring the integrity of data storage in Cloud Computing. In particular, we consider the task of allowing a third party auditor (TPA), on behalf of the cloud client, to verify the integrity of the dynamic data stored in the cloud. The introduction of TPA eliminates the involvement of the client through the auditing of whether his data stored in the cloud are indeed intact, which can be important in achieving economies of scale for Cloud Computing. The support for data dynamics via the most general forms of data operation, such as block modification, insertion, and deletion, is also a significant step toward practicality, since services in Cloud Computing are not limited to archive or backup data only. While prior works on ensuring remote data integrity often lacks the support of either public auditability or dynamic data operations, this paper achieves both. We first identify the difficulties and potential security problems of direct extensions with fully dynamic data updates from prior works and then show how to construct an elegant verification scheme for the seamless integration of these two salient features in our protocol design. In particular, to achieve efficient data dynamics, we improve the existing proof of storage models by manipulating the classic Merkle Hash Tree construction for block tag authentication. To support efficient handling of multiple auditing tasks, we further explore the technique of bilinear aggregate signature to extend our main result into a multiuser setting, where TPA can perform multiple auditing tasks simultaneously. Extensive security and performance analysis show that the proposed schemes are highly efficient and provably secure.

## 43.COMIC: Cost Optimization for Internet Content Multihoming.

### Synopsis:

Content service is a type of Internet cloud service that provides end-users plentiful contents. To ensure high performance for content delivering, content service utilizes a technology known as content multihoming: contents are generated from multiplegeographically distributed data centers and delivered by multiple distributed content distribution networks (CDNs). The electricity costs for data centers and the usage costs for CDNs are major contributors to the contents service cost. As electricity prices vary across data centers and usage costs vary across CDNs, scheduling data centers and CDNs has a tremendous consequence for optimizing content service cost. In this paper, we propose a novel framework named Cost Optimization for Internet Content Multihoming (COMIC). COMIC dynamically balances end-users' loads among data centers and CDNs so as to minimize the content service cost. Using real-lifeelectricity prices and CDN traces, the experiments demonstrate that COMIC effectively reduces the content service cost by more than 20 percent.

## 44. Privacy Preserving Back-Propagation Neural Network Learning Made Practical with Cloud Computing.

Synopsis:

To improve the accuracy of learning result, in practice multiple parties may collaborate through conducting joint Back-Propagation neural network learning on the union of their respective data sets. During this process no party wants to disclose her/his private data to others. Existing schemes supporting this kind of collaborative learning are either limited in the way of data partition or just consider two parties. There lacks a solution that allows two or more parties, each with an arbitrarily partitioned data set, to collaboratively conduct the learning. This paper solves this open problem by utilizing the power of cloud computing. In our proposed scheme, each party encrypts his/her private data locally and uploads the ciphertexts into the cloud. The cloud then executes most of the operations pertaining to the learning algorithms over ciphertexts without knowing the original private data. By securely offloading the expensive operations to the cloud, we keep the computation and communication costs on each party minimal and independent to the number of participants. To support flexible operations over ciphertexts, we adopt and tailor the BGN "doubly homomorphic" encryption algorithm for the multiparty setting. Numerical analysis and experiments on commodity cloud show that our scheme is secure, efficient, and accurate.

## 45. Balancing the Trade-Offs between Query Delay and Data Availability in MANETs.

Synopsis:

In mobile ad hoc networks (MANETs), nodes move freely and link/node failures are common, which leads to frequent network partitions. When a network partition occurs, mobile nodes in one partition are not able to access data hosted by nodes in other partitions, and hence significantly degrade the performance of data access. To deal with this problem, we apply data replication techniques. Existing data replication solutions in both wired or wireless networks aim at either reducing the query delay or improving the data availability, but not both. As both metrics are important for mobile nodes, we propose schemes to balance the trade-offs between data availability and query delay under different system settings and requirements. Extensive simulation results show that the proposed schemes can achieve a balance between these two metrics and provide satisfying system performance.

## 46. Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud.

Synopsis:

In recent years ad hoc parallel data processing has emerged to be one of the killer applications for Infrastructure-as-a-Service (IaaS) clouds. Major Cloud computing companies have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. However, the processing frameworks which are currently used have been designed for static, homogeneous cluster setups and disregard the particular nature of a cloud. Consequently, the allocated compute resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost. In this paper, we discuss the opportunities and challenges for efficient parallel data processing in clouds and present our research project Nephele. Nephele is the first data processing framework to explicitly exploit the dynamic resource allocation offered by today's IaaS clouds for both, task scheduling and execution. Particular tasks of a processing job can be assigned to different types of virtual machines which are automatically instantiated and terminated during the job execution. Based on this new framework, we perform extended evaluations of MapReduce-inspired processing jobs on an IaaS cloud system and compare the results to the popular data processing framework Hadoop.

## 47. Exploiting Rateless Codes in Cloud Storage Systems.

Synopsis:

Block-level cloud storage (BLCS) offers to users and applications the access to persistent block storage devices (virtual disks) that can be directly accessed and used as if they were raw physical disks. In this paper we devise ENIGMA, an architecture for the back-end of BLCS systems able to provide adequate levels of access and transfer performance, availability, integrity, and confidentiality, for the data it stores. ENIGMA exploits LT rateless codes to store fragments of sectors on storage nodes organized in clusters. We quantitatively evaluate how the various ENIGMA system parameters affect the performance, availability, integrity, and confidentiality of virtual disks. These evaluations are carried out by using both analytical modeling (for availability, integrity, and confidentiality) and discrete event simulation (for performance), and by considering a set of realistic operational scenarios. Our results indicate that it is possible to simultaneously achieve all the objectives set forth for BLCS systems by using ENIGMA, and that a careful choice of the various system parameters is crucial to achieve a good compromise among them. Moreover, they also show that LT coding-based BLCS systems outperform traditional BLCS systems in all the aspects mentioned before.

## 48. Secure and Efficient Data Transmission for Cluster-Based Wireless Sensor Networks.

Synopsis:

Secure data transmission is a critical issue for wireless sensor networks (WSNs). Clustering is an effective and practical way to enhance the system performance of WSNs. In this paper, we study a secure data transmission for cluster-based WSNs (CWSNs), where the clusters are formed dynamically and periodically. We propose two secure and efficient data transmission (SET) protocols for CWSNs, called SET-IBS and SET-IBOOS, by using the identity-based digital signature (IBS) scheme and the identity-based online/offline digital signature (IBOOS) scheme, respectively. In SET-IBS, security relies on the hardness of the Diffie-Hellman problem in the pairing domain. SET-IBOOS further reduces the computational overhead for protocol security, which is crucial for WSNs, while its security relies on the hardness of the discrete logarithm problem. We show the feasibility of the SET-IBS and SET-IBOOS protocols with respect to the security requirements and security analysis against various attacks. The calculations and simulations are provided to illustrate the efficiency of the proposed protocols. The results show that the proposed protocols have better performance than the existing secure protocols for CWSNs, in terms of security overhead and energy consumption.

## 49. BECAN: A Bandwidth-Efficient Cooperative Authentication Scheme for Filtering Injected False Data in Wireless Sensor Networks.

Synopsis:

Injecting false data attack is a well known serious threat to wireless sensor network, for which an adversary reports bogus information to sink causing error decision at upper level and energy waste in en-route nodes. In this paper, we propose a novel bandwidth-efficient cooperative authentication (BECAN) scheme for filtering injected false data. Based on the random graph characteristics of sensor node deployment and the cooperative bit-compressed authentication technique, the proposed BECAN scheme can save energy by early detecting and filtering the majority of injected false data with minor extra overheads at the en-route nodes. In addition, only a very small fraction of injected false data needs to be checked by the sink, which thus largely reduces the burden of the sink. Both theoretical and simulation results are given to demonstrate the effectiveness of the proposed scheme in terms of high filtering probability and energy saving.

## 50. Rumor Riding: Anonymizing Unstructured Peer-to-Peer Systems.

Synopsis:

Although anonymizing Peer-to-Peer (P2P) systems often incurs extra costs in terms of transfer efficiency, many systems try to mask the identities of their users for privacy considerations. Existing anonymity approaches are mainly path-based: peers have to pre-construct an anonymous path before transmission. The overhead of maintaining and updating such paths is significantly high. In this paper, we propose Rumor Riding (RR), a lightweight mutual anonymity protocol for decentralized P2P systems. RR employs a random walk scheme which frees initiating peers from the heavy load of path construction. Compared with previous RSA-based anonymity approaches, RR also takes advantage of lower cryptographic overhead by mainly utilizing a symmetric cryptographic algorithm to achieve anonymity. We demonstrate the effectiveness of this design through trace-driven simulations. The analytical and experimental results show that RR is more efficient than existing protocols. We also discuss our early implementation experiences with the RR prototype.

## 51. Exploiting Service Similarity for Privacy in Location-Based Search Queries.

Synopsis:

Location-based applications utilize the positioning capabilities of a mobile device to determine the current location of a user, and customize query results to include neighboring points of interests. However, location knowledge is often perceived as personal information. One of the immediate issues hindering the wide acceptance of location-based applications is the lack of appropriate methodologies that offer fine grain privacy controls to a user without vastly affecting the usability of the service. While a number of privacy-preserving models and algorithms have taken shape in the past few years, there is an almost universal need to specify one's privacy requirement without understanding its implications on the service quality. In this paper, we propose a user-centric location-based service architecture where a user can observe the impact of location inaccuracy on the service accuracy before deciding the geo-coordinates to use in a query. We construct a local search application based on this architecture and demonstrate how meaningful information can be exchanged between the user and the service provider to allow the inference of contours depicting the change in query results across a geographic area. Results indicate the possibility of large default privacy regions (areas of no change in result set) in such applications.

## 52. SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency.

Synopsis:

With the pervasiveness of smart phones and the advance of wireless body sensor networks (BSNs), mobile Healthcare (m-Healthcare), which extends the operation of Healthcare provider into a pervasive environment for better health monitoring, has attracted considerable interest recently. However, the flourish of m-Healthcare still faces many challenges including information security and privacy preservation. In this paper, we propose a secure and privacy-preserving opportunistic computing framework, called SPOC, for m-Healthcare emergency. With SPOC, smart phone resources including computing power and energy can be opportunistically gathered to process the computing-intensive personal health information (PHI) during m-Healthcare emergency with minimal privacy disclosure. In specific, to leverage the PHI privacy disclosure and the high reliability of PHI process and transmission in m-Healthcare emergency, we introduce an efficient user-centric privacy access control in SPOC framework, which is based on an attribute-based access control and a new privacy-preserving scalar product computation (PPSPC) technique, and allows a medical user to decide who can participate in the opportunistic computing to assist in processing his overwhelming PHI data. Detailed security analysis shows that the proposed SPOC framework can efficiently achieve user-centric privacy access control in m-Healthcare emergency. In addition, performance evaluations via extensive simulations demonstrate the SPOC's effectiveness in term of providing high-reliable-PHI process and transmission while minimizing the privacy disclosure during m-Healthcare emergency.

## 53. BloomCast: Efficient and Effective Full-Text Retrieval in Unstructured P2P Networks.

### Synopsis:

Efficient and effective full-text retrieval in unstructured peer-to-peer networks remains a challenge in the research community. First, it is difficult, if not impossible, for unstructured P2P systems to effectively locate items with guaranteed recall. Second, existing schemes to improve search success rate often rely on replicating a large number of item replicas across the wide area network, incurring a large amount of communication and storage costs. In this paper, we propose BloomCast, an efficient and effective full-text retrieval scheme, in unstructured P2P networks. By leveraging a hybrid P2P protocol, BloomCast replicates the items uniformly at random across the P2P networks, achieving a guaranteed recall at a communication cost of O(√N), where N is the size of the network. Furthermore, by casting Bloom Filters instead of the raw documents across the network, BloomCast significantly reduces the communication and storage costs for replication. We demonstrate the power of BloomCast design through both mathematical proof and comprehensive simulations based on the query logs from a major commercial search engine and NIST TREC WT10G data collection. Results show that BloomCast achieves an average query recall of 91 percent, which outperforms the existing WP algorithm by 18 percent, while BloomCast greatly reduces the search latency for query processing by 57 percent.

## 54. Throughput and Delay Analysis for Convergecast with MIMO in Wireless Networks.

### Synopsis:

This paper investigates throughput and delay based on a traffic pattern, called convergecast, where each of the n nodes in the network acts as a destination with k randomly chosen sources corresponding to it. Adopting Multiple-Input-Multiple-Output (MIMO) technology, we devise two many-to-one cooperative schemes under convergecast for both static and mobile ad hoc networks (MANETs), respectively. We call them Convergimo Schemes. In static networks, our Convergimo scheme highly utilizes hierarchical cooperation MIMO transmission. This feature overcomes the bottleneck which hinders convergecast traffic from yielding ideal performance in traditional ad hoc network, by turning the originally interfering signals into interference-resistant ones. It helps to achieve an aggregate throughput up to $\Omega(n1-\epsilon)$ for any $\epsilon$ >;0. In the mobile ad hoc case, our Convergimo scheme characterizes on joint transmission from multiple nodes to multiple receivers. With optimal network division where the number of nodes per cell is constantly bounded, the achievable per-node throughput can reach $\Theta(1)$ with the corresponding delay reduced to $\Theta(k)$. The gain comes from the strong and intelligent cooperation between nodes in our scheme, along with the maximum number of concurrent active cells and the shortest waiting time before transmission for each node within a cell. This increases the chances for each destination to receive the data it needs with minimum overhead on extra transmission. Moreover, our converge-based analysis well unifies and generalizes previous work since the results derived from convergecast in our schemes can also cover other traffic patterns. Last but not the least, our schemes are of interest not only from a theoretical perspective but also provide useful theoretical guidelines to future design of MIMO schemes in wireless networks.

## 55. Link Quality Aware Code Dissemination in Wireless Sensor Networks.

### Synopsis:

Wireless reprogramming is a crucial technique for software deployment in wireless sensor networks (WSNs). Code dissemination is a basic building block to enable wireless reprogramming. We present ECD, an Efficient Code Dissemination protocol leveraging 1-hop link quality information based on the TinyOS platform. Compared to prior works, ECD has three salient features. First, it supports dynamically configurable packet sizes. By increasing the packet size for high PHY rate radios, it significantly improves the transmission efficiency. Second, it employs an accurate sender selection algorithm to mitigate transmission collisions and transmissions over poor links. Third, it employs a simple impact-based backoff timer design to shorten the time spent in coordinating multiple eligible senders so that the largest impact sender is most likely to transmit. We implement ECD based on TinyOS and evaluate its performance extensively via testbed experiments and simulations. Results show that ECD outperforms state-of-the-art protocols, Deluge and MNP, in terms of completion time and data traffic (e.g., about 20 percent less traffic and 20-30 percent shorter completion time compared to Deluge).

## 56. Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data.

### Synopsis:

Cloud computing economically enables the paradigm of data service outsourcing. However, to protect data privacy, sensitive cloud data have to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization service a very challenging task. Although traditional searchable encryption techniques allow users to securely search over encrypted data through keywords, they support only Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud. In this paper, we define and solve the problem of secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. Specifically, we explore the statistical measure approach, i.e., relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information. The resulting design is able to facilitate efficient server-side ranking without losing keyword privacy. Thorough analysis shows that our proposed solution enjoys “as-strong-as-possible” security guarantee compared to previous searchable encryption schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution.

## 57. In Cloud, Can Scientific Communities Benefit from the Economies of Scale?.

### Synopsis:

The basic idea behind cloud computing is that resource providers offer elastic resources to end users. In this paper, we intend to answer one key question to the success of cloud computing: in cloud, can small-to-medium scale scientific communities benefit from the economies of scale? Our research contributions are threefold: first, we propose an innovative public cloud usage model for small-to-medium scale scientific communities to utilize elastic resources on a public cloud site while maintaining their flexible system controls, i.e., create, activate, suspend, resume, deactivate, and destroy their high-level management entities-service management layers without knowing the details of management. Second, we design and implement an innovative system-DawningCloud, at the core of which are lightweight service management layers running on top of a common management service framework. The common management service framework of DawningCloud not only facilitates building lightweight service management layers for heterogeneous workloads, but also makes their management tasks simple. Third, we evaluate the systems comprehensively using both emulation and real experiments. We found that for four traces of two typical scientific workloads: High-Throughput Computing (HTC) and Many-Task Computing (MTC), DawningCloud saves the resource consumption maximally by 59.5 and 72.6 percent for HTC and MTC service providers, respectively, and saves the total resource consumption maximally by 54 percent for the resource provider with respect to the previous two public cloud solutions. To this end, we conclude that small-to-medium scale scientific communities indeed can benefit from the economies of scale of public clouds with the support of the enabling system.

## 58. QoS Aware Geographic Opportunistic Routing in Wireless Sensor Networks.

### Synopsis:

QoS routing is an important research issue in wireless sensor networks (WSNs), especially for mission-critical monitoring and surveillance systems which requires timely and reliable data delivery. Existing work exploits multipath routing to guarantee both reliability and delay QoS constraints in WSNs. However, the multipath routing approach suffers from a significant energy cost. In this work, we exploit the geographic opportunistic routing (GOR) for QoS provisioning with both end-to-end reliability and delay constraints in WSNs. Existing GOR protocols are not efficient for QoS provisioning in WSNs, in terms of the energy efficiency and computation delay at each hop. To improve the efficiency of QoS routing in WSNs, we define the problem of efficient GOR for multiconstrained QoS provisioning in WSNs, which can be formulated as a multiobjective multiconstraint optimization problem. Based on the analysis and observations of different routing metrics in GOR, we then propose an Efficient QoS-aware GOR (EQGOR) protocol for QoS provisioning in WSNs. EQGOR selects and prioritizes the forwarding candidate set in an efficient manner, which is suitable for WSNs in respect of energy efficiency, latency, and time complexity. We comprehensively evaluate EQGOR by comparing it with the multipath routing approach and other baseline protocols through ns-2 simulation and evaluate its time complexity through measurement on the MicaZ node. Evaluation results demonstrate the effectiveness of the GOR approach for QoS provisioning in WSNs. EQGOR significantly improves both the end-to-end energy efficiency and latency, and it is characterized by the low time complexity.

## 59. Privacy-preserving multi-keyword ranked search over encrypted cloud data.

### Synopsis:

With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching", i.e., as many matches as possible, to capture the relevance of data documents to the search query. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

## 60. Expert Discovery and Interactions in Mixed Service-Oriented Systems.

### Synopsis:

Web-based collaborations and processes have become essential in today's business environments. Such processes typically span interactions between people and services across globally distributed companies. Web services and SOA are the defacto technology to implement compositions of humans and services. The increasing complexity of compositions and the distribution of people and services require adaptive and context-aware interaction models. To support complex interaction scenarios, we introduce a mixed service-oriented system composed of both human-provided and Software-Based Services (SBSs) interacting to perform joint activities or to solve emerging problems. However, competencies of people evolve over time, thereby requiring approaches for the automated management of actor skills, reputation, and trust. Discovering the right actor in mixed service-oriented systems is challenging due to scale and temporary nature of collaborations. We present a novel approach addressing the need for flexible involvement of experts and knowledge workers in distributed collaborations. We argue that the automated inference of trust between members is a key factor for successful collaborations. Instead of following a security perspective on trust, we focus on dynamic trust in collaborative networks. We discuss Human-Provided Services (HPSs) and an approach for managing user preferences and network structures. HPS allows experts to offer their skills and capabilities as services that can be requested on demand. Our main contributions center around a context-sensitive trust-based algorithm called ExpertHITS inspired by the concept of hubs and authorities in web-based environments. ExpertHITS takes trust-relations and link properties in social networks into account to estimate the reputation of users.

## 61. Dynamic Authentication for Cross-Realm SOA-Based Business Processes.

### Synopsis:

Modern distributed applications are embedding an increasing degree of dynamism, from dynamic supply-chain management, enterprise federations, and virtual collaborations to dynamic resource acquisitions and service interactions across organizations. Such dynamism leads to new challenges in security and dependability. Collaborating services in a system with a Service-Oriented Architecture (SOA) may belong to different security realms but often need to be engaged dynamically at runtime. If their security realms do not have a direct cross-realm authentication relationship, it is technically difficult to enable any secure collaboration between the services. A potential solution to this would be to locate intermediate realms at runtime, which serve as an authentication path between the two separate realms. However, the process of generating an authentication path for two distributed services can be highly complicated. It could involve a large number of extra operations for credential conversion and require a long chain of invocations to intermediate services. In this paper, we address this problem by designing and implementing a new cross-realm authentication protocol for dynamic service interactions, based on the notion of service-oriented multiparty business sessions. Our protocol requires neither credential conversion nor establishment of any authentication path between the participating services in a business session. The correctness of the protocol is formally analyzed and proven, and an empirical study is performed using two production-quality Grid systems, Globus 4 and CROWN. The experimental results indicate that the proposed protocol and its implementation have a sound level of scalability and impose only a limited degree of performance overhead, which is for example comparable with those security-related overheads in Globus 4.

## 62. Recursive Linear and Differential Cryptanalysis of Ultralightweight Authentication Protocols.

### Synopsis:

Privacy is faced with serious challenges in the ubiquitous computing world. In order to handle this problem, some researchers in recent years have focused on design and analysis of privacy-friendly ultralightweight authentication protocols. Although the majority of these schemes have been broken to a greater or lesser extent, most of these attacks are based on ad-hoc methods that are not extensible to a large class of ultralightweight protocols. So this research area still suffers from the lack of structured cryptanalysis and evaluation methods. In this paper, we introduce new frameworks for full disclosure attacks on ultralightweight authentication protocols based on new concepts of recursive linear and recursive differential cryptanalysis. The recursive linear attack is passive, deterministic, and requires only a single authentication session, if it can be applied successfully. The recursive differential attack is more powerful and can be applied to the protocols on which the linear attack may not work. This attack is probabilistic, active in the sense that the attacker suffices only to block some specific messages, and requires a few authentication sessions. Having introduced these frameworks in a general view, we apply them on some well-known ultralightweight protocols. The first attack can retrieve all the secret data of Yeh and SLMAP authentication protocols and the second one can retrieve all the secret data of LMAP++, SASI, and David-Prasad authentication protocols.

## 63. Secure Deduplication with Efficient and Reliable Convergent Key Management.

### Synopsis:

Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. We first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey , a new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers. Security analysis demonstrates that Dekey is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments.

## 64. Query Access Assurance in Outsourced Databases.

### Synopsis:

Query execution assurance is an important concept in defeating lazy servers in the database as a service model. We show that extending query execution assurance to outsourced databases with multiple data owners is highly inefficient. To cope with lazy servers in the distributed setting, we propose query access assurance (Qaa) that focuses on IO-bound queries. The goal in Qaa is to enable clients to verify that the server has honestly accessed all records that are necessary to compute the correct query answer, thus eliminating the incentives for the server to be lazy if the query cost is dominated by the IO cost in accessing these records. We formalize this concept for distributed databases, and present two efficient schemes that achieve Qaa with high success probabilities. The first scheme is simple to implement and deploy, but may incur excessive server to client communication cost and verification cost at the client side, when the query selectivity or the database size increases. The second scheme is more involved, but successfully addresses the limitation of the first scheme. Our design employs a few number theory techniques. Extensive experiments demonstrate the efficiency, effectiveness, and usefulness of our schemes.

## 65. Random4: An Application Specific Randomized Encryption Algorithm to Prevent SQL Injection.

### Synopsis:

Web Applications form an integral part of our day to day life. The number of attacks on websites and the compromise of many individuals' secure data are increasing at an alarming rate. With the advent of social networking and e-commerce, Web security attacks such as phishing and spamming have become quite common. The consequences of these attacks are ruthless. Hence, providing increased amount of security for the users and their data becomes essential. Most important vulnerability as described in top 10 web security issues by Open Web Application Security Project is SQL Injection Attack (SQLIA) [3]. This paper focuses on how the advantages of randomization can be employed to prevent SQL injection attacks in web based applications. SQL injection can be used for unauthorized access to a database to penetrate the application illegally, modify the database or even remove it. For a hacker to modify a database, details such as field and table names are required. So we try to propose a solution to the above problem by preventing it using an encryption algorithm based on randomization. It has better performance and provides increased security in comparison to the existing solutions. Also the time to crack the database takes more time when techniques such as dictionary and brute force attack are deployed. Our main aim is to provide increased security by developing a tool which prevents illegal access to the database.

## 66. REMO: Resource-Aware Application State Monitoring for Large-Scale Distributed Systems.

### Synopsis:

To observe, analyze and control large scale distributed systems and the applications hosted on them, there is an increasing need to continuously monitor performance attributes of distributed system and application states. This results in application state monitoring tasks that require fine-grained attribute information to be collected from relevant nodes efficiently. Existing approaches either treat multiple application state monitoring tasks independently and build ad-hoc monitoring trees for each task, or construct a single static monitoring tree for multiple tasks. We argue that a careful planning of multiple application state monitoring tasks by jointly considering multi-task optimization and node level resource constraints can provide significant gains in performance and scalability. In this paper, we present REMO, a REsource-aware application state MOnitoring system. REMO produces a forest of optimized monitoring trees through iterations of two phases, one phase exploring cost sharing opportunities via estimation and the other refining the monitoring plan through resource-sensitive tree construction. Our experimental results include those gathered by deploying REMO on a BlueGene/P rack running IBM's large-scale distributed streaming system - System S. Using REMO running over 200 monitoring tasks for an application deployed across 200 nodes results in a 35%-45% decrease in the percentage error of collected attributes compared to existing schemes.

## 67. Uniform Embedding for Efficient JPEG Steganography.

### Synopsis:

Steganography is the science and art of covert communication, which aims to hide the secret messages into a cover medium while achieving the least possible statistical detectability. To this end, the framework of minimal distortion embedding is widely adopted in the development of the steganographic system, in which a well designed distortion function is of vital importance. In this paper, a class of new distortion functions known as uniform embedding distortion function (UED) is presented for both side-informed and non side-informed secure JPEG steganography. By incorporating the syndrome trellis coding, the best codeword with minimal distortion for a given message is determined with UED, which, instead of random modification, tries to spread the embedding modification uniformly to quantized discrete cosine transform (DCT) coefficients of all possible magnitudes. In this way, less statistical detectability is achieved, owing to the reduction of the average changes of the first- and second-order statistics for DCT coefficients as a whole. The effectiveness of the proposed scheme is verified with evidence obtained from exhaustive experiments using popular steganalyzers with various feature sets on the BOSSbase database. Compared with prior arts, the proposed scheme gains favorable performance in terms of secure embedding capacity against steganalysis.

## 68. SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency.

### Synopsis:

With the pervasiveness of smart phones and the advance of wireless body sensor networks (BSNs), mobile Healthcare (m-Healthcare), which extends the operation of Healthcare provider into a pervasive environment for better health monitoring, has attracted considerable interest recently. However, the flourish of m-Healthcare still faces many challenges including information security and privacy preservation. In this paper, we propose a secure and privacy-preserving opportunistic computing framework, called SPOC, for m-Healthcare emergency. With SPOC, smart phone resources including computing power and energy can be opportunistically gathered to process the computing-intensive personal health information (PHI) during m-Healthcare emergency with minimal privacy disclosure. In specific, to leverage the PHI privacy disclosure and the high reliability of PHI process and transmission in m-Healthcare emergency, we introduce an efficient user-centric privacy access control in SPOC framework, which is based on an attribute-based access control and a new privacy-preserving scalar product computation (PPSPC) technique, and allows a medical user to decide who can participate in the opportunistic computing to assist in processing his overwhelming PHI data. Detailed security analysis shows that the proposed SPOC framework can efficiently achieve user-centric privacy access control in m-Healthcare emergency. In addition, performance evaluations via extensive simulations demonstrate the SPOC's effectiveness in term of providing high-reliable-PHI process and transmission while minimizing the privacy disclosure during m-Healthcare emergency.

## 69. Web Service Recommendation via Exploiting Location and QoS Information.

## Synopsis:

Web services are integrated software components for the support of interoperable machine-to-machine interaction over a network. Web services have been widely employed for building service-oriented applications in both industry and academia in recent years. The number of publicly available Web services is steadily increasing on the Internet. However, this proliferation makes it hard for a user to select a proper Web service among a large amount of service candidates. An inappropriate service selection may cause many problems (e.g., ill-suited performance) to the resulting applications. In this paper, we propose a novel collaborative filtering-based Web service recommender system to help users select services with optimal Quality-of-Service (QoS) performance. Our recommender system employs the location information and QoS values to cluster users and services, and makes personalized service recommendation for users based on the clustering results. Compared with existing service recommendation methods, our approach achieves considerable improvement on the recommendation accuracy. Comprehensive experiments are conducted involving more than 1.5 million QoS records of real-world Web services to demonstrate the effectiveness of our approach.