CODE	TITLE	YEAR	ABSTRACT
SPDN-01	Dynamic and Public Auditing with Fair Arbitration for Cloud Data	2016	Cloud users no longer physically possess their data, so how to ensure the integrity of their outsourced data becomes a challenging task. Recently proposed schemes such as “provable data possession” and “proofs of retrievability” are designed to address this problem, but they are designed to audit static archive data and therefore lack of data dynamics support. Moreover, threat models in these schemes usually assume an honest data owner and focus on detecting a dishonest cloud service provider despite the fact that clients may also misbehave. This paper proposes a public auditing scheme with data dynamics support and fairness arbitration of potential disputes. In particular, we design an index switcher to eliminate the limitation of index usage in tag computation in current schemes and achieve efficient handling of data dynamics. To address the fairness problem so that no party can misbehave without being detected, we further extend existing threat models and adopt signature exchange idea to design fair arbitration protocols, so that any possible dispute can be fairly settled. The security analysis shows our scheme is provably secure, and the performance evaluation demonstrates the overhead of data dynamics and dispute arbitration are reasonable.
SPDN-02	Enabling Cloud Storage Auditing with Verifiable Outsourcing of Key Updates	2016	Key-exposure resistance has always been an importantissue for in-depth cyber defence in many security applications. Recently, how to deal with the key exposure problem in the settings of cloud storage auditing has been proposed and studied. To address the challenge, existing solutions all require the client to update his secret keys in every time period, which may inevitably bring in new local burdens to the client, especially those with limited computation resources such as mobile phones. In this paper, we focus on how to make the key updates as transparent as possible for the client and propose a new paradigm called cloud storage auditing with

			verifiable outsourcing of key updates. In this paradigm, key updates can be safely outsourced to some authorized party, and thus the key-update burden on the client will be kept minimal. Specifically, we leverage the third party auditor (TPA) in many existing public auditing designs, let it play the role of authorized party in our case, and make it in charge of both the storage auditing and the secure key updates for key-exposure resistance. In our design, TPA only needs to hold an encrypted version of the client's secret key, while doing all these burdensome tasks on behalf of the client. The client only needs to download the encrypted secret key from the TPA when uploading new files to cloud. Besides, our design also equips the client with capability to further verify the validity of the encrypted secret keys provided by TPA. All these salient features are carefully designed to make the whole auditing procedure with key exposure resistance as transparent as possible for the client. We formalize the definition and the security model of this paradigm. The security proof and the performance simulation show that our detailed design instantiations are secure and efficient.
SPDN-03	Providing User Security Guarantees in Public Infrastructure Clouds	2016	The infrastructure cloud (IaaS) service model offers improved resource flexibility and availability, where tenants – insulated from the minutiae of hardware maintenance – rent computing resources to deploy and operate complex systems. Large-scale services running on IaaS platforms demonstrate the viability of this model; nevertheless, many organizations operating on sensitive data avoid migrating operations to IaaS platforms due to security concerns. In this paper, we describe a framework for data and operation security in IaaS, consisting of protocols for a trusted launch of virtual machines and domain-based storage protection. We continue with an extensive theoretical analysis with proofs about protocol resistance against attacks in the

			defined threat model. The protocols allow trust to be established by remotely attesting host platform configuration prior to launching guest virtual machines and ensure confidentiality of data in remote storage, with encryption keys maintained outside of the IaaS domain. Presented experimental results demonstrate the validity and efficiency of the proposed protocols. The framework prototype was implemented on a test bed operating a public electronic health record system, showing that the proposed protocols can be integrated into existing cloud environments.
SPDN-04	Service Usage Classification with Encrypted Internet Traffic in Mobile Messaging Apps	2016	The rapid adoption of mobile messaging Apps has enabled us to collect massive amount of encrypted Internet traffic of mobile messaging. The classification of this traffic into different types of in-App service usages can help for intelligent network management, such as managing network bandwidth budget and providing quality of services. Traditional approaches for classification of Internet traffic rely on packet inspection, such as parsing HTTP headers. However, messaging Apps are increasingly using secure protocols, such as HTTPS and SSL, to transmit data. This imposes significant challenges on the performances of service usage classification by packet inspection. To this end, in this paper, we investigate how to exploit encrypted Internet traffic for classifying in-App usages. Specifically, we develop a system, named CUMMA, for classifying service usages of mobile messaging Apps by jointly modeling user behavioral patterns, network traffic characteristics and temporal dependencies. Along this line, we first segment Internet traffic from traffic-flows into sessions with a number of dialogs in a hierarchical way. Also, we extract the discriminative features of traffic data from two perspectives: (i) packet length and (ii) time delay. Next, we learn a service usage predictor to classify these segmented dialogs

			into single-type usages or outliers. In addition, we design a clustering Hidden Markov Model (HMM) based method to detect mixed dialogs from outliers and decompose mixed dialogs into sub-dialogs of single-type usage. Indeed, CUMMA enables mobile analysts to identify service usages and analyze end-user in-App behaviors even for encrypted Internet traffic. Finally, the extensive experiments on real-world messaging data demonstrate the effectiveness and efficiency of the proposed method for service usage classification.
SPDN-05	**Text Mining the Contributors to Rail Accidents**	**2016**	**Rail accidents represent an important safety concern for the transportation industry in many countries. In the 11 years from 2001 to 2012, the U.S. had more than 40 000 rail accidents that cost more than $45 million. While most of the accidents during this period had very little cost, about 5200 had damages in excess of $141 500. To better understand the contributors to these extreme accidents, the Federal Railroad Administration has required the railroads involved in accidents to submit reports that contain both fixed field entries and narratives that describe the characteristics of the accident. While a number of studies have looked at the fixed fields, none have done an extensive analysis of the narratives. This paper describes the use of text mining with a combination of techniques to automatically discover accident characteristics that can inform a better understanding of the contributors to the accidents. The study evaluates the efficacy of text mining of accident narratives by assessing predictive performance for the costs of extreme accidents. The results show that predictive accuracy for accident costs significantly improves through the use of features found by text mining and predictive accuracy further improves through the use of modern ensemble methods. Importantly, this study also shows through case examples how the findings from text mining of the narratives**

			can improve understanding of the contributors to rail accidents in ways not possible through only fixed field analysis of the accident reports.
SPDN-06	**MMB*cloud*-tree: Authenticated Index for Verifiable Cloud Service Selection**	**2016**	**Cloud brokers have been recently introduced as an additional computational layer to facilitate cloud selection and service management tasks for cloud consumers. However, existing brokerage schemes on cloud service selection typically assume that brokers are completely trusted, and do not provide any guarantee over the correctness of the service recommendations. It is then possible for a compromised or dishonest broker to easily take advantage of the limited capabilities of the clients and provide incorrect or incomplete responses. To address this problem, we propose an innovative Cloud Service Selection Verification (CSSV) scheme and index structures (MMB*cloud*-tree) to enable cloud clients to detect misbehavior of the cloud brokers during the service selection process. We demonstrate correctness and efficiency of our approaches both theoretically and empirically.**
SPDN-07	**Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity Checking in Public Cloud**	**2016**	**More and more clients would like to store their data to PCS (public cloud servers) along with the rapid development of cloud computing. New security problems have to be solved in order to help more clients process their data in public cloud. When the client is restricted to access PCS, he will delegate its proxy to process his data and upload them. On the other hand, remote data integrity checking is also an important security problem in public cloud storage. It makes the clients check whether their outsourced data is kept intact without downloading the whole data. From the security problems, we propose a novel proxy-oriented data uploading and remote data integrity checking model in identity-based public key cryptography: IDPUIC (identity-based proxy-oriented data uploading and remote data integrity checking in public**

			cloud). We give the formal definition, system model and security model. Then, a concrete ID-PUIC protocol is designed by using the bilinear pairings. The proposed ID-PUIC protocol is provably secure based on the hardness of CDH (computational Diffie-Hellman) problem. Our ID-PUIC protocol is also efficient and flexible. Based on the original client's authorization, the proposed ID-PUIC protocol can realize private remote data integrity checking, delegated remote data integrity checking and public remote data integrity checking.
SPDN-08	**Fine-grained Two-factor Access Control for Web-based Cloud Computing Services**	**2016**	**In this paper, we introduce a new fine-grained two-factor authentication (2FA) access control system for web-based cloud computing services. Specifically, in our proposed 2FA access control system, an attribute-based access control mechanism is implemented with the necessity of both user secret key and a lightweight security device. As a user cannot access the system if s/he does not hold both, the mechanism can enhance the security of the system, especially in those scenarios where many users share the same computer for web-based cloud services. In addition, attribute-based control in the system also enables the cloud server to restrict the access to those users with the same set of attributes while preserving user privacy, i.e., the cloud server only knows that the user fulfills the required predicate, but has no idea on the exact identity of the user. Finally, we also carry out a simulation to demonstrate the practicability of our proposed 2FA system.**
SPDN-09	**Cloud workflow scheduling with deadlines and time slot availability**	**2016**	**Allocating service capacities in cloud computing is based on the assumption that they are unlimited and can be used at any time. However, available service capacities change with workload and cannot satisfy users' requests at any time from the cloud provider's perspective because cloud services can be shared by multiple tasks. Cloud service providers provide available time slots for new user's requests based on available**

			capacities. In this paper, we consider workflow scheduling with deadline and time slot availability in cloud computing. An iterated heuristic framework is presented for the problem under study which mainly consists of initial solution construction, improvement, and perturbation. Three initial solution construction strategies, two greedy- and fair-based improvement strategies and a perturbation strategy are proposed. Different strategies in the three phases result in several heuristics. Experimental results show that different initial solution and improvement strategies have different effects on solution qualities.
SPDN-10	**Publicly Verifiable Inner Product Evaluation** **over Outsourced Data Streams under Multiple** **Keys**	**2016**	**Uploading data streams to a resource-rich cloud server for inner product evaluation, an essential building block in many popular stream applications (e.g., statistical monitoring), is appealing to many companies and individuals. On the other hand, verifying the result of the remote computation plays a crucial role in addressing the issue of trust. Since the outsourced data collection likely comes from multiple data sources, it is desired for the system to be able to pinpoint the originator of errors by allotting each data source a unique secret key, which requires the inner product verification to be performed under any two parties' different keys. However, the present solutions either depend on a single key assumption or powerful yet practicallyinefficient fully homomorphic cryptosystems. In this paper, we focus on the more challenging multi-key scenario where data streams are uploaded by multiple data sources with distinct keys. We first present a novel homomorphic verifiable tag technique to publicly verify the outsourced inner product computation on the dynamic data streams, and then extend it to support the verification of matrix product computation. We prove the security of our scheme in the random oracle model. Moreover, the experimental result also shows the practicability of our design.**

SPDN-11	**Inverted Linear Quadtree: Efficient Top K Spatial Keyword Search**	**2016**	**With advances in geo-positioning technologies and geo-location services, there are a rapidly growing amount of *spatio-textual* objects collected in many applications such as location based services and social networks, in which an object is described by its spatial location and a set of keywords (terms). Consequently, the study of spatial keyword search which explores both location and textual description of the objects has attracted great attention from the commercial organizations and research communities. In the paper, we study the problem of top *k* spatial keyword search (TOPK-SK), which is fundamental in the spatial keyword queries. Given a set of *spatio-textual* objects, a query location and a set of query keywords, the top *k* spatial keyword search retrieves the closest *k* objects each of which contains all keywords in the query. Based on the inverted index and the linear quadtree, we propose a novel index structure, called inverted linear quadtree (IL-Quadtree), which is carefully designed to exploit both spatial and keyword based pruning techniques to effectively reduce the search space. An efficient algorithm is then developed to tackle top *k* spatial keyword search. In addition, we show that the IL-Quadtree technique can also be applied to improve the performance of other spatial keyword queries such as the direction-aware top *k* spatial keyword search and the *spatiotextual* ranking query. Comprehensive experiments on real and synthetic data clearly demonstrate the efficiency of our methods.**
SPDN-12	**Securing SIFT: Privacy-preserving Outsourcing Computation of Feature Extractions over Encrypted Image Data**	**2016**	**Advances in cloud computing have greatly motivated data owners to outsource their huge amount of personal multimedia data and/or computationally expensive tasks onto the cloud by leveraging its abundant resources for cost saving and flexibility. Despite the tremendous benefits, the outsourced multimedia data and its originated applications may reveal the data**

			owner’s private information, such as the personal identity, locations or even financial profiles. This observation has recently aroused new research interest on privacy-preserving computations over outsourced multimedia data. In this paper, we propose an effective and practical privacy-preserving computation outsourcing protocol for the prevailing scale-invariant feature transform (SIFT) over massive encrypted image data. We first show that previous solutions to this problem have either efficiency/security or practicality issues, and none can well preserve the important characteristics of the original SIFT in terms of distinctiveness and robustness. We then present a new scheme design that achieves efficiency and security requirements simultaneously with the preservation of its key characteristics, by randomly splitting the original image data, designing two novel efficient protocols for secure multiplication and comparison, and carefully distributing the feature extraction computations onto two independent cloud servers. We both carefully analyze and extensively evaluate the security and effectiveness of our design. The results show that our solution is practically secure, outperforms the state-of-theart, and performs comparably to the original SIFT in terms of various characteristics, including rotation invariance, image scale invariance, robust matching across affine distortion, addition of noise and change in 3D viewpoint and illumination.
SPDN-13	**A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data**	**2016**	**Abstract—Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which**

			simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF_IDF model are combined in the index construction and query generation. We construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results . Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.
SPDN-14		**2016**	Search over encrypted data is a critically important enabling technique in cloud computing, where encryption-beforeoutsourcing is a fundamental solution to protecting user data privacy in the untrusted cloud server environment. Many secure search schemes have been focusing on the single-contributor scenario, where the outsourced dataset or the secure searchable index of the dataset are encrypted and managed by a single owner, typically based on symmetric cryptography. In this paper, we focus on a different yet more challenging scenario where the outsourced dataset can be contributed from multiple owners and are searchable by multiple users, i.e. multi-user multicontributor case. Inspired by attribute-based encryption (ABE), we present the first attribute-based keyword search scheme with efficient user revocation (ABKS-UR) that enables scalable fine-grained (i.e. file-level) search authorization. Our scheme allows multiple owners to encrypt and outsource

			their data to the cloud server independently. Users can generate their own search capabilities without relying on an always online trusted authority. Fine-grained search authorization is also implemented by the owner-enforced access policy on the index of each file. Further, by incorporating proxy re-encryption and lazy reencryption techniques, we are able to delegate heavy system update workload during user revocation to the resourceful semitrusted cloud server. We formalize the security definition and prove the proposed ABKS-UR scheme selectively secure against chosen-keyword attack. Finally, performance evaluation shows the efficiency of our scheme.
SPDN-15	**A Low-Cost Low-Power Ring Oscillator-based Truly Random Number Generator for Encryption on Smart Cards**	**2016**	**The design of a low-cost low-power ring oscillator-based truly random number generator (TRNG) macro-cell, suitable to be integrated in smart cards, is presented. The oscillator sampling technique is exploited and a tetrahedral oscillator with large jitter has been employed to realize the TRNG. Techniques to improve the statistical quality of the ring oscillator-based TRNGs' bit sequences have been presented and verified by simulation and measurement. Post digital processor is added to further enhance the randomness of the output bits. Fabricated in HHNEC 0.13 □m standard CMOS process, the proposed TRNG has an area as low as 0.005 mm2. Powered by a single 1.8 V supply voltage, the TRNG has a power consumption of 40 □W. Bit rate of the TRNG after post processing is 100 kb/s. The proposed TRNG has been made into an IP and successfully applied in an SD card for encryption application. The proposed TRNG has passed the NIST tests and Diehard tests.**
SPDN-16	**A recommendation system based on hierarchical clustering of an article-level citation network**	**2016**	**The scholarly literature is expanding at a rate that necessitates intelligent algorithms for search and navigation.For the most part, the problem of delivering scholarly articles has been solved. If one knows the title of an article, locating it requires little effort and, paywalls permitting, acquiring a digital copy**

			has become trivial.However, the navigational aspect of scientific search – finding relevant, influential articles that one does not know exist – is in its early development. In this paper, we introduce Eigenfactor Recommends – a citation-based method for improving scholarly navigation. The algorithm uses the hierarchical structure of scientific knowledge, making possible multiple scales of relevance for different users. We implement the method and generate more than 300 million recommendations from more than 35 million articles from various bibliographic databases including the AMiner dataset. We find little overlap with co-citation, another well-known citation recommender, which indicates potential complementarity. In an online A-B comparison using SSRN, we find that our approach performs as well as co-citation, but this new approach offers much larger recommendation coverage. We make the code and recommendations freely available at babel.eigenfactor.org and provide an API for implementing and comparing the recommendations on their own platforms.
SPDN-17	**Dual-Server Public-Key Encryption with Keyword Search for Secure Cloud Storage**	**2016**	Searchable encryption is of increasing interest for protecting the data privacy in secure searchable cloud storage. In this work, we investigate the security of a well-known cryptographic primitive, namely Public Key Encryption with Keyword Search (PEKS) which is very useful in many applications of cloud storage. Unfortunately, it has been shown that the traditional PEKS framework suffers from an inherent insecurity called inside Keyword Guessing Attack (KGA) launched by the malicious server. To address this security vulnerability, we propose a new PEKS framework named Dual-Server Public Key Encryption with Keyword Search (DS-PEKS). As another main contribution, we define a new variant of the Smooth Projective Hash Functions (SPHFs) referred to as linear and homomorphic SPHF (LH-SPHF). We then

			show a generic construction of secure DS-PEKS from LH-SPHF. To illustrate the feasibility of our new framework, we provide an efficient instantiation of the general framework from a DDH-based LH-SPHF and show that it can achieve the strong security against inside KGA.
SPDN-18	**Efficient Group Key Transfer Protocol for WSNs**	**2016**	**Special designs are needed for cryptographic schemes in wireless sensor networks (WSNs). This is because sensor nodes are limited in memory storage and computational power. The existing group key transfer protocols for WSNs using classical secret sharing require that a *t*-degree interpolating polynomial be computed in order to encrypt and decrypt the secret group key. This approach is too computationally intensive. In this paper, we propose a new group key transfer protocol using a linear secret sharing scheme (LSSS) and factoring assumption. The proposed protocol can resist potential attacks and also significantly reduce the computation complexity of the system while maintaining low communication cost. Such a scheme is desirable for secure group communications in wireless sensor networks (WSNs), where portable devices or sensors need to reduce their computation as much as possible due to battery power limitations.**
SPDN-19	**Demand-Aware Centralized Traffic Scheduling in Wireless LANs**	**2016**	***Abstract*—A heavy deployment of IEEE 802.11 Wireless LANs and limited number of orthogonal channels make lots of Access Points (APs) overlap their interference regions, which greatly increases interferences between APs and stations. In order to cope with the performance degradation caused by the interferences, we propose CO-FI, a centralized Wi-Fi architecture that effectively coordinates downlink transmissions by APs and improves network performance in terms of throughput and endto- end delay. CO-FI adaptively allocates time slots for APs and stations based on both traffic demands on the stations and a conflict graph that represents interference relationships among**

			the devices. The scheme allows APs in exposed node relationship to use the channel simultaneously by setting the same backoff time. It also effectively avoids downlink conflicts created by hidden node and non-hidden/non-exposed node, by allocating non-overlapping time slots to interfering stations. To implement these adaptive traffic schedules, we design CoMAC, a hybrid MAC protocol at APs. Our evaluation results show that when APs are densely deployed and the network is highly loaded, the scheme achieves 3-5 times more throughput gain than Centaur, a state-of-the-art scheme while its end-to-end delays are 10-90% lower than those of Centaur and CSMA/CA.
SPDN-20	**Unsupervised Feature Selection for Text Classification via Word Embedding**	**2016**	**The key of big text documents data analysis is to classify those text documents. To classify those text documents, it is necessary to represent those text documents as vectors which is vector space model (*VSM*). A powerful vector space model should remain the classification information with dimensions as little as possible. To achieve that, it is important to select most effective features for text classification. Unlike the supervised selection method which utilizes the category information in the training data, we propose an unsupervised feature selection method. Our method requires no category information which makes our method has more application scenarios as the labeled data is expensive and inaccurate. Unlike other unsupervised methods, our method utilizes word embedding to find the words with similar semantic meaning. The word embedding maps the words into vectors and remains the semantic relationships between words. We select the most representative word on behalf of the words with similar semantic meaning because it is redundant to include all those words as features. We demonstrate on Reuters-21578 dataset, that our method outperforms other methods. Especially, our method has great advantage when select limited features.**

SPDN-21	*Opinion mining from student feedback data using supervised learning algorithms*	2016	This paper explores opinion mining using supervised learning algorithms to find the polarity of the student feedback based on pre-defined features of teaching and learning. The study conducted involves the application of a combination of machine learning and natural language processing techniques on student feedback data gathered from module evaluation survey results of Middle East College, Oman. In addition to providing a step by step explanation of the process of implementation of opinion mining from student comments using the open source data analytics tool Rapid Miner, this paper also presents a comparative performance study of the algorithms like SVM, Naïve Bayes, K Nearest Neighbor and Neural Network classifier. The data set extracted from the survey is subjected to data preprocessing which is then used to train the algorithms for binomial classification. The trained models are also capable of predicting the polarity of the student comments based on extracted features like examination, teaching etc. The results are compared to find the better performance with respect to various evaluation criteria for the different algorithms.
SPDN-22	Privacy-Preserving and Regular Language Search over Encrypted Cloud Data	2016	Using cloud-based storage service, users can remotely store their data to clouds but also enjoy the high quality data retrieval services, without the tedious and cumbersome local data storage and maintenance. However, the sole storage service cannot satisfy all desirable requirements of users. Over the last decade, privacy-preserving search over encrypted cloud data has been a meaningful and practical research topic for outsourced data security. The fact of remote cloud storage service that users cannot have full physical possession of their data makes the privacy data search a formidable mission. A naive solution is to delegate a trusted party to access the stored data and fulfill a search task. This, nevertheless, does not scale well in

			practice as the fully data access may easily yield harm for user privacy. To securely introduce an effective solution, we should guarantee the privacy of search contents, i.e. what a user wants to search, and return results, i.e. what a server returns to the user. Furthermore, we also need to guarantee privacy for the outsourced data, and bring no additional local search burden to user. In this paper, we design a novel privacy-preserving functional encryption based search mechanism over encrypted cloud data. A major advantage of our new primitive compared to the existing public key based search systems is that it supports an extreme expressive search mode, regular language search. Our security and performance analysis show that the proposed system is provably secure and more efficient than some searchable systems with high expressiveness.
SPDN-23	**KSF-OABE: Outsourced Attribute-Based Encryption with Keyword Search Function for Cloud Storage**	**2016**	**Abstract—Cloud computing becomes increasingly popular for data owners to outsource their data to public cloud servers while allowing intendeddata users to retrieve these data stored in cloud. This kind of computing model brings challenges to the security and privacy of data stored in cloud. Attribute-based encryption (ABE) technology has beenusedto design fine-grained access control system, which provides one good method to solve the security issuesin cloud setting. However, the computation cost and ciphertext size in most ABE schemes grow with the complexity of the access policy. Outsourced ABE(OABE) with fine-grained access control system can largely reduce the computation cost for users who want to access encrypted data stored in cloud by outsourcing the heavy computation to cloud service provider (CSP). However, as the amountof encrypted files stored in cloud is becomingvery huge, which will hinder efficient query processing. To deal withabove problem, we presenta new cryptographic primitive called attribute-based encryption**

			scheme with outsourcing key-issuing and outsourcing decryption, which can implement keyword search function (KSF-OABE). The proposed KSF-OABE scheme isprovedsecure against chosen-plaintext attack (CPA). CSP performs partial decryption task delegated by data user without knowing anything about the plaintext. Moreover,the CSP can perform encrypted keyword search without knowing anything about the keywords embeddedin trapdoor.
SPDN-24	**Ensuring Replication-based Data Integrity and Availability in Multicloud Storage**	**2016**	**With the growing popularity of cloud storage service, how to ensure the integrity and availability of outsourced data has become a critical problem. To solve this, several remote data checking protocols are proposed to prove the data possession and retrievability. However, these methods either incur large overhead or are unable to repair the remote corruptions. In this paper, we propose the multi-replicas based data integrity verification and recovery (*MRVR*) scheme to realize public auditing and rapid recovery in multi-cloud environment. The protocols of replication-based verification and recovery are proposed based on the techniques of homomorphism, bilinear map and the binary aggregation tree (*BAT*). Through security analysis, *MRVR* is proved to be secure and privacy preserving with high data recoverability and availability. Finally, we analyze the system performance from both computation and communication overheads, and compare the algorithm efficiencies in the courses of verification and recovery with relevant schemes.**
SPDN-25	**Cost-aware Cloud Storage Service Allocation for Distributed Data Gathering**	**2016**	**In today cyber-infrastructures, large datasets are produced in real-time by different sources geographically distributed. These data must be acquired and preserved for further use in knowledge extraction. In the context of multi-cloud environments, the cost-efficient storage service selection is a challenge. There are plenty of Cloud storage providers offering multiple options so, it is**

			crucial to select the best solution in terms of cost and quality of service that meet customers requirements. Due to its multi-objective nature, the process of optimal service selection becomes a difficult problem. In this paper, we study the multi-objective optimization problem for storage service selection. We start from a real world case scenario and build our mathematical model for the optimization problem. Then we propose an aggregated linear programming technique to find a near optimal solution for the service selection problem.
SPDN-26	An effective Soft Multiple-Assignments strategies for enhancing the accuracy of the Content-Based Image Retrieval Systems	2016	The multiple-assignments approach alleviates the quantization error and enhances the accuracy of the Content- Based Image Retrieval (CBIR) systems. It aims to hard assign each feature vector to k-nearest visual words. However, during the matching step, the k-nearest visual words are used independently and ignore the significant of the best visual word. In this paper, we present our CBIR system which encapsulates several approaches such Hamming embedding, soft-assignment, multiple-assignments and graph fusion. We particularly focus on the multiple-assignments strategy. We propose an efficient soft multiple-assignments strategy to highlight the best k-nearest visual word. To this end, we explore the SOM topology which proved its performance in so doing. Moreover, we use graph fusion approach to fuse multi features ranking lists. Extensive experiments are conducted on Holiday and Ukbench public datasets. The experimental results are promising and outperform the state-of-the-art CBIR systems. In fact, we have reached a mAP = 85.6 on Holidays dataset and a KS score of 3.87 on Ukbench dataset.
SPDN-27	Mining and Visualizing Associations of Concepts on a Large-scale Unstructured Data	2016	We investigate the problem of finding unknown associations between ‘concepts’ in a given text corpus. A ‘concept’ is an entity, which is referred to by a phrase or multiple phrases (in case an entity has several names

			or synonyms , e.g. ”illness”, ”disease”), while an ‘association’ is a relationship defined in a particular domain. The pairwise associations computation poses major challenges when the size of concepts to be investigated is large. In this paper, we propose a new framework for computing and visualizing the association between concepts using the generic association measures and the public knowledge available in data sources such as Wikipedia. Core to our methodology are the pruning techniques employed to filter the pairs of concepts which are very unlikely to be associated. Moreover, we study the performance of different correlation measures in finding both ‘direct’ and ‘indirect’ associations between concepts. The indirect association analysis extends the utility of the proposed framework to a broader class of association mining applications including the interesting area of new hypothesis generation. The extracted associations in our framework are demonstrated using the Neo4J graph database, which not only provides a user-friendly visualization interface for observing the associations between concepts in a big graph, but also supports more advanced analytics such as ‘community detection’, ‘centrality analysis’, and SQL-like querying of the concept graph. We evaluate the performance of our proposed methodology on two datasets, including a ‘gene-disease’ association dataset from the DisGeNET public database, and ‘disease-symptom-treatment’ associations extracted manually from the “MedicineNet.net” website. Our results show that the proposed association measure is capable of finding 80% of true gene-disease associations with a false positive rate of 15%.
SPDN-28	**A recommendation system based on hierarchical clustering of an article-level citation network**	**2016**	**he scholarly literature is expanding at a rate that necessitates intelligent algorithms for search and navigation.For the most part, the problem of delivering scholarly articles has been solved. If one knows the title of an**

			article, locating it requires little effort and, paywalls permitting, acquiring a digital copy has become trivial.However, the navigational aspect of scientific search – finding relevant, influential articles that one does not know exist – is in its early development. In this paper, we introduce Eigenfactor Recommends – a citation-based method for improving scholarly navigation. The algorithm uses the hierarchical structure of scientific knowledge, making possible multiple scales of relevance for different users. We implement the method and generate more than 300 million recommendations from more than 35 million articles from various bibliographic databases including the AMiner dataset. We find little overlap with co-citation, another well-known citation recommender, which indicates potential complementarity. In an online A-B comparison using SSRN, we find that our approach performs as well as co-citation, but this new approach offers much larger recommendation coverage. We make the code and recommendations freely available at babel.eigenfactor.org and provide an API for others to use for implementing and comparing the recommendations on their own platforms.
SPDN-29	**Prefix-adaptive and Time-sensitive Personalized Query Auto Completion?**	**2016**	**Query auto completion (QAC) methods recommend queries to search engine users when they start entering a query. Current QAC methods mostly rank query completions based on their past popularity, i.e., on the number of times they have previously been submitted as a query. However, query popularity changes over time and may vary drastically across users. Accordingly, the ranking of query completions should be adjusted. Previous time-sensitive and user-specific QAC methods have been developed separately, yielding significant improvements over methods that are neither time-sensitive nor personalized. We propose a hybrid QAC method that is both time-sensitive and**

			personalized. We extend it to handle long-tail prefixes, which we achieve by assigning optimal weights to the contribution from time-sensitivity and personalization. Using real-world search log datasets, we return top N query suggestions ranked by predicted popularity as estimated from popularity trends and cyclic popularity behavior; we rerank them by integrating similarities to a user's previous queries (both in the current session and in previous sessions). Our method outperforms state-of-the-art time-sensitive QAC baselines, achieving total improvements of between 3% and 7% in terms of mean reciprocal rank (MRR). After optimizing the weights, our extended model achieves MRR improvements of between 4% and 8%.
SPDN-30	**A Proposed Implementation Method of an Audio Steganography Technique**	**2016**	**Steganography is the art of science dealing with hiding secret data inside image, audio, video or text files. In audio steganography; secret message is embedded in the digital sound by slightly altering the binary sequence of the sound file. Existing audio steganography software deal with WAV, AU, and even MP3 sound files. Embedding secret messages in the digital sound is usually a more difficult process than embedding messages in other forms, such as digital images. Audio steganography uses different algorithms, but (LSB) least significant bit is applied in this paper. The quality of sound is depended on the size of the audio which the user selects and length of the message**
SPDN-31	**Design and Analysis of an Efficient Friend-to-Friend Content Dissemination System**	**2016**	**Opportunistic communication, off-loading and decentrlaized distribution have been proposed as a means of cost efficient disseminating content when users are geographically clustered into communities. Despite its promise, none of the proposed systems have not been widely adopted due to unbounded high content delivery latency, security and privacy concerns. This paper, presents a novel hybrid content storage and distribution system addressing the trust and privacy concerns of users, lowering the cost**

			of content distribution and storage, and shows how they can be combined uniquely to develop mobile social networking services. The system exploit the fact that users will trust their friends, and by replicating content on friends’ devices who are likely to consume that content it will be possible to disseminate it to other friends when connected to low cost networks. The paper provides a formal definition of this content replication problem, and show that it is NP hard. Then, it presents a community based greedy heuristic algorithm with novel dynamic centrality metrics that replicates the content on a minimum number of friends’ devices, to maximize availability. Then using both real world and synthetic datasets, the effectiveness of the proposed scheme is demonstrated. The practicality of the proposed system, is demonstrated through an implementation on Android smartphones
SPDN-32	Stochastic Content-Centric Multicast Scheduling for Cache-Enabled Heterogeneous Cellular Networks	2016	Caching at small base stations (SBSs) has demonstrated significant benefits in alleviating the backhaul requirement in heterogeneous cellular networks (HetNets). While many existing works focus on what contents to cache at each SBS, an equally important problem is what contents to deliver so as to satisfy dynamic user demands given the cache status. In this paper, we study optimal content delivery in cache-enabled HetNets by taking into account the inherent multicast capability of wireless medium. We consider stochastic content multicast scheduling to jointly minimize the average network delay and power costs under a multiple access constraint. We establish a content-centric request queue model and formulate this stochastic optimization problem as an infinite horizon average cost Markov decision process (MDP). By using *relative value iteration* and special properties of the request queue dynamics, we characterize some properties of the value function of the MDP. Based on these properties, we show that the optimal

			multicast scheduling policy is of threshold type. Then, we propose a structure-aware optimal algorithm to obtain the optimal policy. We also propose a lowcomplexity suboptimal policy, which possesses similar structural properties to the optimal policy, and develop a low-complexity algorithm to obtain this policy.