BIG DATA ANALYSIS USING PARTITION TECHNIQUE

ABSTRACT

Analyzing a Big data is a challenging task because of its characteristics and presence of data in large amount. For large scale data analysis hadoop technology provides a key role. Aggregate queries executes on number of columns simultaneously and difficult for large amount of data .In this paper we are proposing FRAQ and Balanced Partition technique which gives better performance with help of PIG and generate a histogram for the respective partition. This Histogram is generating according to the specified column by user in interface. Histogram gives an effective result according to requested query and data set coming online.

Java

MULTI-FOCUS IMAGE FUSION BASED ON SPATIAL FREQUENCY IN DISCRETE COSINE TRANSFORM DOMAIN

ABSTRACT

Multi-focus image fusion in wireless visual sensor networks (WVSN) is a process of fusing two or more images to obtain a new one which contains a more accurate description of the scene than any of the individual source images. In this letter, we propose an efficient algorithm to fuse multi-focus images or videos using discrete cosine transform (DCT) based standards inWVSN. The spatial frequencies of the corresponding blocks from source images are calculated as the contrast criteria, and the blocks with the larger spatial frequencies compose the DCT presentation of the output image. Experiments on plenty of pairs ofmulti-focus images coded in Joint Photographic Experts Group (JPEG) standard are conducted to evaluate the fusion performance. The results show that our fusion method improves the quality of the output image visually and outperforms the previous DCT based techniques and the state-of-art methods in terms of the objective evaluation.

Java

A SKEWNESS ALGORITHM SCHEDULING APPROACH FOR THE ENERGETIC DISTRIBUTION OF RESOURCES FOR CLOUD COMPUTING ENVIRONMENT USING VIRTUAL MACHINES

ABSTRACT

Cloud computing is a new technology that uses the internet. At present cloud computing environment, the forecast approaches for (Virtual Machine) VM resources are mainly focus on the systems current state. In the resource management problems the Dynamic resource allocation problem is one of the most demanding problems. To present a better solution for solving the problem of dynamic resource allocation in a cloud computing environment, the proposed system demonstrates a skewness algorithm to determine the unevenness in the multi-dimensional resource utilization of a server. By reducing the skewness, we can combine different types of workloads and increase the overall utilization of server resources. By using virtualization technology, we can assign resources to the data center dynamically based on application demands and support green computing by optimizing the number of servers in use.

Java

ADAPTIVE QOS FOR MOBILE WEB SERVICES THROUGH CROSS-LAYER COMMUNICATION

ABSTRACT

Web services providers can use the WS-QoS framework to achieve QoS differentiation by integrating various aspects of distinct communication layers. An architecture based on the framework supports resource-constrained mobile devices, which will generate a large percentage of Web service requests in the future.

Java

BOUNDING THE ADVANTAGE OF MULTICAST NETWORK CODING IN GENERAL NETWORK MODELS

ABSTRACT

Network coding encourages information flow mixing in a network. It helps increase the throughput and reduce the cost of data transmission, especially for one-to-many multicast applications. An interesting problem is to understand and quantify the *coding advantage* and *cost advantage*, *i.e.*, the potential benefits of network coding, as compared to routing, interms of increasing throughput and reducing transmission cost, respectively. Two classic network models were considered in previous studies: directed networks and undirected networks. This work further focuses on two types of parameterized networks, including bidirected networks and hyper-networks, generalizing the directed and the undirected network models, respectively. Weprove upper- and lower-bounds on multicast coding advantage and cost advantage in these models.

Java

CIPHERXRAY: EXPOSING CRYPTOGRAPHIC OPERATIONS AND TRANSIENT SECRETS FROM MONITORED BINARY EXECUTION

ABSTRACT

Malwares are becoming increasingly stealthy, more and more malwares are using cryptographic algorithms (e.g., packing, encrypting C&C communication) to protect themselves from being analyzed. The use of cryptographic algorithms and truly transient cryptographic secrets inside the malware binary imposes a key obstacle to effective malware analysis and defense.To enable more effective malware analysis, forensics and reverse engineering, we have developed *CipherXRay* – a novel binary analysis framework that can automatically identify andrecover the cryptographic operations and transient secrets from the execution of potentially obfuscated binary executables. Based on the *avalanche effect* of cryptographic functions, CipherXRayis able to accurately pinpoint the boundary of cryptographic operation and recover truly transient cryptographic secrets that only exist in memory for one instant in between multiplenested cryptographic operations. CipherXRay can furtheridentify certain operation modes (e.g., ECB, CBC, CFB) of the identified block cipher and tell whether the identified block cipheroperation is encryption or decryption in certain cases. We have empirically validated CipherXRay with OpenSSL, popular password safe KeePassX, the ciphers used by malwareStuxnet, Kraken and Agobot, and a number of third party softwares with built-in compression and checksum. CipherXRay is able to identify various cryptographic operations and recover cryptographic secrets that exist in memory for only a few microseconds. Our results demonstrate that current software implementations of cryptographic algorithms hardly achieve any secrecy if their execution can be monitored.

GREEN NETWORKING WITH PACKET PROCESSING ENGINES:

MODELING AND OPTIMIZATION

ABSTRACT

With the aim of controlling power consumption in metro/transport and core networks, we consider energy-aware devices able to reduce their energy requirements, by adapting their performance. In particular, we focus on state-of-the-art packet processing engines, which generally represent the most energy-starving components of network devices, and which are often composed of a number of parallel pipelines to “divide and conquer” the incoming traffic load. Our goal is to control both the power configuration of pipelines, and the way to distribute traffic flows among them. We propose an analytical model to accurately represent the impact of green network technologies (i.e., low power idle and adaptive rate) on network- and energy aware performance indexes. The model has been validated with experimental results, performed by using energy-aware software routers loaded by real-world traffic traces. The achieved results demonstrate how the proposed model can effectively represent energy- and network-aware performance indexes. On this basis, we propose a constrained optimization policy, which seeks the best trade-off between power consumption and packet latency times. The procedure aims at dynamically adapting the energy aware device configuration to minimize energy consumption, while coping with incoming traffic volumes and meeting network performance constraints. In order to deeply understand the impact of such policy, a number of tests have been performed by using experimental data from software router architectures and real-world traffic traces.

Java

A HIGHLY SECURE VIDEO STEGANOGRAPHY USING HAMMING CODE (7, 4)

ABSTRACT

Due to the high speed of internet and advances in technology, people are becoming more worried about information being hacked by attackers. Recently, many algorithms of steganography and data hiding have been proposed. Steganography is a process of embedding the secret information inside the host medium (text, audio, image and video). Concurrently, many of the powerful steganographic analysis software programs have been provided to unauthorized users to retrieve the valuable secret information that was embedded in the carrier files. Some steganography algorithms can be easily detected by steganalytical detectors because of the lack of security and embedding efficiency.

In this paper, we propose a secure video steganography algorithm based on the principle of linear block code. Nine uncompressed video sequences are used as cover data and a binary image logo as a secret message. The pixels' positions of both cover videos and a secret message are randomly reordered by using a private key to improve the system's security. Then the secret message is encoded by applying Hamming code (7, 4) before the embedding process to make the message even more secure. The result of the encoded message will be added to random generated values by using XOR function. After these steps that make the message secure enough, it will be ready to be embedded into the cover video frames. In addition, the embedding area in each frame is randomly selected and it will be different from other frames to improve the steganography scheme's robustness. Furthermore, the algorithm has high embedding efficiency as demonstrated by the experimental results that we have obtained. Regarding the system's quality, the Pick Signal to Noise Ratio (PSNR) of stego videos are above 51 dB, which is close to the original video quality. The embedding payload is also acceptable, where in each video frame we can embed 16 Kbits and it can go up to 90 Kbits without noticeable degrading of the stego video's quality.

Java

MEET YOU – SOCIAL NETWORKING ON ANDROID

ABSTRACT

This paper aims to present a system that illustrates the social nature of a human being – the need to be always in touch with family and friends – taking into account facilities available on Android platform. The role of this application is to create a social network in which the users are being alerted when their friends are around. This gives them the possibility to set up a meeting or to avoid one. The users have the possibility to check in some locations and allow their friends to follow their activity. Taking into account the security of the users, we included in the facilities of the application an option which allows close friends or family to check the user's location based on a keyword text message. For this purpose, available Android location and messages services are used for finding an approximate location of a mobile phone running this program and then sharing it through Meet You or via SMS. Information is being displayed using default components provided by Android platform and also more complex elements including heterogeneous lists CWAC, Google Maps and augmented reality using Mixare Library.

Java

SCALABLE DISTRIBUTED SERVICE INTEGRITY ATTESTATION FOR SOFTWARE-AS-A-SERVICE CLOUDS

ABSTRACT

Software-as-a-service (SaaS) cloud systems enable application service providers to deliver their applications via massive cloud computing infrastructures. However, due to their sharing nature, SaaS clouds are vulnerable to malicious attacks. In this paper, we present IntTest, a scalable and effective service integrity attestation framework for SaaS clouds. IntTest provides a novel integrated attestation graph analysis scheme that can provide stronger attacker pinpointing power than previous schemes. Moreover, IntTest can automatically enhance result quality by replacing bad results produced by malicious attackers with good results produced by benign service providers. We have implemented a prototype of the IntTest system and tested it on a production cloud computing infrastructure using IBM System S stream processing applications. Our experimental results show that IntTest can achieve higher attacker pinpointing accuracy than existing approaches. IntTest does not require any special hardware or secure kernel support and imposes little performance impact to the application, which makes it practical for large-scale cloud systems.

Java

SEMANTIC LINK NETWORK-BASED MODEL FOR ORGANIZING MULTIMEDIA

BIG DATA

ABSTRACT

Recent research shows that multimedia resources *in the wild* are growing at a staggering rate. The rapid increase number of multimedia resources has brought an urgent need to develop intelligent methods to organize and process them. In this paper, the semantic link network model is used for organizing multimedia resources. A whole model for generating the association relation between multimedia resources using seman- tic link network model is proposed. The de_nitions, modules, and mechanisms of the semantic link network are used in the proposed method. The integration between the semantic link network and multimedia resources provides a new prospect for organizing them with their semantics. The tags and the surrounding texts of multimedia resources are used to measure their semantic association. The hierarchical semantic of multimedia resources is de_ned by their annotated tags and surrounding texts. The semantics of tags and surrounding texts are different in the proposed framework. The modules of semantic link network model are implemented to measure association relations. A real data set including 100 thousand images with social tags from Flickr is used in our experiments. Two evaluation methods, including clustering and retrieval, are performed, which shows the proposed method can measure the semantic relatedness between Flickr images accurately and robustly.

Java

STREAMING SOLUTIONS FOR FINE-GRAINED NETWORK

TRAFFIC MEASUREMENTS AND ANALYSIS

ABSTRACT

Online network traffic measurements and analysis is critical for detecting and preventing any real-time anomalies in the network. We propose, implement, and evaluate an online, adaptive measurement platform, which utilizes real-time traffic analysis results to refine subsequent traffic measurements.Central to our solution is the concept of Multi-Resolution Tiling (MRT), a heuristic approach that performs sequential analysis of traffic data to zoom into traffic subregions of interest. However, MRT is sensitive to transient traffic spikes. In this paper, we propose three novel traffic streaming algorithms that overcome the limitations of MRT and can cater to varying degrees of computational and storage budgets, detection latency, and accuracy of query response. We evaluate our streaming algorithms on a highly parallel and programmable hardware as well as a traditional software-based platforms. The algorithms demonstrate significant accuracy improvement over MRT in detecting anomalies consisting of synthetic hard-to-track elephant flows and global icebergs. Our proposed algorithms maintain the worst-case complexities of the MRT while incurring only a moderate increase in average resource utilization.

Java

OCCT: A ONE-CLASS CLUSTERING TREE FOR IMPLEMENTING ONE-TO-MANY DATA LINKAGE

ABSTRACT

One-to-many data linkage is an essential task in many domains, yet only a handful of prior publications have addressed this issue. Furthermore, while traditionally data linkage is performed among entities of the same type, it is extremely necessary to develop linkage techniques that link between matching entities of different types as well. In this paper, we propose a new one-to-many data linkage method that links between entities of different natures. The proposed method is based on a one-class clustering tree (OCCT) that characterizes the entities that should be linked together. The tree is built such that it is easy to understand and transform into association rules, i.e., the inner nodes consist only of features describing the first set of entities, while the leaves of the tree represent features of their matching entities from the second data set. We propose four splitting criteria and two different pruning methods which can be used for inducing the OCCT. The method was evaluated using data sets from three different domains. The results affirm the effectiveness of the proposed method and show that the OCCT yields better performance in terms of precision and recall (in most cases it is statistically significant) when compared to a C4.5 decision tree-based linkage method.

Java

A NOVEL JOINT DATA-HIDING AND COMPRESSION SCHEME BASED ON SMVQ AND IMAGE INPAINTING

ABSTRACT

There are also a number of works on data hiding in the encrypted domain. The reversible data hiding in encrypted image is investigated in. Most of the work on reversible data hiding focuses on the data embedding/extracting on the plain spatial domain. A content owner encrypts the original image using an encryption key, and a data-hider can embed additional data into the encrypted image using a data-hiding key though he does not know the original content. With an encrypted image containing additional data, a receiver may first decrypt it according to the encryption key, and then extract the embedded data and recover the original image according to the data-hiding key. In a novel scheme for separable reversible data hiding in encrypted image. In the proposed scheme, the original image is encrypted using an encryption key and the additional data are embedded into the encrypted image using a data-hiding key.If a receiver has the data-hiding key, he can extract the additional data though he does not know the image content. If the receiver has the encryption key, he can decrypt the received data to obtain an image similar to the original one, but cannot extract the additional data. If the receiver has both the data-hiding key and the encryption key, he can extract the additional data and recover the original content without any error by exploiting the spatial correlation in natural image when the amount of additional data is not too large.

Java

AUTOMATED GRAPH REGULARIZED PROJECTIVE NONNEGATIVE MATRIX FACTORIZATION FOR DOCUMENT CLUSTERING

ABSTRACT

The class imbalance problem is encountered in a large number of practical applications of machine learning and data mining, for example, information retrieval and filtering, and the detection of credit card fraud. It has been widely realized that this imbalance raises issues that are either nonexistent or less severe compared to balanced class cases and often results in a classifier's suboptimal performance. This is even more true when the imbalanced data are also high dimensional. In such cases, feature selection methods are critical to achieve optimal performance. In this paper, we propose a new feature selection method, Feature Assessment by Sliding Thresholds (FAST), which is based on the area under a ROC curve generated by moving the decision boundary of a single feature classifier with thresholds placed using an even-bin distribution. FAST is compared to two commonly-used feature selection methods, correlation coefficient and RELevance In Estimating Features (RELIEF), for imbalanced data classification. The experimental results obtained on text mining, mass spectrometry, and microarray data sets showed that the proposed method outperformed both RELIEF and correlation methods on skewed data sets and was comparable on balanced data sets; when small number of features is preferred, the classification performance of the proposed method was significantly improved compared to correlation and RELIEF-based methods.

Java

BEHAVIORAL MALWARE DETECTION IN DELAY TOLERANT NETWORKS

ABSTRACT

The delay-tolerant-network (DTN) model is becoming a viable communication alternative to the traditional infrastructural model for modern mobile consumer electronics equipped with short-range communication technologies such as Bluetooth, NFC, and Wi-Fi Direct. Proximity malware is a class of malware that exploits the opportunistic contacts and distributed nature of DTNs for propagation. Behavioral characterization of malware is an effective alternative to pattern matching in detecting malware, especially when dealing with polymorphic or obfuscated malware. In this paper, we first propose a general behavioral characterization of proximity malware which based on naive Bayesian model, which has been successfully applied in non-DTN settings such as filtering email spams and detecting botnets. We identify two unique challenges for extending Bayesian malware detection to DTNs ("insufficient evidence versus evidence collection risk" and "filtering false evidence sequentially and distributedly"), and propose a simple yet effective method, look ahead, to address the challenges. Furthermore, we propose two extensions to look ahead, dogmatic filtering, and adaptive look ahead, to address the challenge of "malicious nodes sharing false evidence." Real mobile network traces are used to verify the effectiveness of the proposed methods.

Java

HIERARCHICAL STRING CUTS: A TRANSLATION, ROTATION, SCALE, AND MIRROR INVARIANT DESCRIPTOR FOR FAST SHAPE RETRIEVAL

ABSTRACT

This paper presents a novel approach for both fast and accurately retrieving similar shapes. A hierarchical string cuts (HSCs) method is proposed to partition a shape into multiple level curve segments of different lengths from a point moving around the contour to describe the shape gradually and completely from the global information to the finest details. At each hierarchical level, the curve segments are cut by strings to extract features that characterize the geometric and distribution properties in that particular level of details. The translation, rotation, scale, and mirror invariant HSC descriptor enables a fast metric-based matching to achieve the desired high accuracy. Encouraging experimental results on four databases demonstrated that the proposed method can consistently achieve higher (or similar) retrieval accuracies than the state-of-the-art benchmarks with a more than 120 times faster speed. This may suggest a new way of developing shape retrieval techniques in which a high accuracy can be achieved by a fast metric matching algorithm without using the time-consuming correspondence optimization strategy.

Java

IMAGE QUALITY ASSESSMENT FOR FAKE BIOMETRIC DETECTION:

APPLICATION TO IRIS, FINGERPRINT, AND FACE RECOGNITION

ABSTRACT

To ensure the actual presence of a real legitimate trait in contrast to a fake self-manufactured synthetic or reconstructed sample is a significant problem in biometric authentication, which requires the development of new and efficient protection measures. In this paper, we present a novel software-based fake detection method that can be used in multiple biometric systems to detect different types of fraudulent access attempts. The objective of the proposed system is to enhance the security of biometric recognition frameworks, by adding liveness assessment in a fast, user-friendly, and non-intrusive manner, through the use of image quality assessment. The proposed approach presents a very low degree of complexity, which makes it suitable for real-time applications, using 25 general image quality features extracted from one image (i.e., the same acquired for authentication purposes) to distinguish between legitimate and impostor samples. The experimental results, obtained on publicly available data sets of fingerprint, iris, and 2D face, show that the proposed method is highly competitive compared with other state-of-the-art approaches and that the analysis of the general image quality of real biometric samples reveals highly valuable information that may be very efficiently used to discriminate them from fake traits.

Java

IRIS IMAGE CLASSIFICATION BASED ON HIERARCHICAL

Visual Codebook

ABSTRACT

Iris recognition as a reliable method for personal identification has been well-studied with the objective to assign the class label of each iris image to a unique subject. In contrast, iris image classification aims to classify an iris image to an application specific category, e.g. iris liveness detection (classification of genuine and fake iris images), race classification (e.g. classification of iris images of Asian and non-Asian subjects), coarse-to-fine iris identification (classification of all iris images in the central database into multiple categories). This paper proposes a general framework for iris image classification based on texture analysis. A novel texture pattern representation method called Hierarchical Visual Codebook (HVC) is proposed to encode the texture primitives of iris images. The proposed HVC method is an integration of two existing Bag-of-Words models, namely Vocabulary Tree (VT), and Locality-constrained Linear Coding (LLC). The HVC adopts a coarse-to-fine visual coding strategy and takes advantages of both VT and LLC for accurate and sparse representation of iris texture. Extensive experimental results demonstrate that the proposed iris image classification method achieves state-of-the-art performance for iris liveness detection, race classification, and coarse-to-fine iris identification. A comprehensive fake iris image database simulating four types of iris spoof attacks is developed as the benchmark for research of iris liveness detection.

Java

STREAMING SOLUTIONS FOR FINE-GRAINED NETWORK

TRAFFIC MEASUREMENTS AND ANALYSIS

ABSTRACT

Online network traffic measurements and analysis is critical for detecting and preventing any real-time anomalies in the network. We propose, implement, and evaluate an online, adaptive measurement platform, which utilizes real-time traffic analysis results to refine subsequent traffic measurements.Central to our solution is the concept of Multi-Resolution Tiling (MRT), a heuristic approach that performs sequential analysis of traffic data to zoom into traffic subregions of interest. However, MRT is sensitive to transient traffic spikes. In this paper, we propose three novel traffic streaming algorithms that overcome the limitations of MRT and can cater to varying degrees of computational and storage budgets, detection latency, and accuracy of query response. We evaluate our streaming algorithms on a highly parallel and programmable hardware as well as a traditional software-based platforms. The algorithms demonstrate significant accuracy improvement over MRT in detecting anomalies consisting of synthetic hard-to-track elephant flows and global icebergs. Our proposed algorithms maintain the worst-case complexities of the MRT while incurring only a moderate increase in average resource utilization.

Java

SUPPORTING PRIVACY PROTECTION IN PERSONALIZED WEB SEARCH

ABSTRACT

The proposed personalized web search engine is an innovative approach for personalizing web search results. By mining content and based concepts for user profiling, it utilizes both the content and Privacy preserving preferences to personalize search results for a user. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial.

Nowadays Internet is widely used by users to satisfy various information needs. However, ambiguous query/topic submitted to search engine doesn't satisfy user information needs, because different users may have different information needs on diverse aspects upon submission of same query/topic to search engine. So discovering different user search goals becomes complicated. The evaluation and depiction of user search goals can be very useful in improving search engine relevance and user knowledge. This paper proposes a novel approach for inferring user search goals by analyzing user query logs from various search engines. The proposed approach is used to discover different user search goals for a query by clustering the user feedback sessions. Feedback sessions are constructed from click-through logs of various search engines. The method first generates pseudo-documents to better represent feedback sessions for clustering. Finally, clustering pseudo-documents to discover different user search goals and depict them with some keywords. Then these user search goals are used to restructure the web search results.

Java

TRANSFORMATION-BASED MONETARY COST OPTIMIZATIONS FOR WORKFLOWS IN THE CLOUD

ABSTRACT

The recent emergence of public cloud offerings, surge computing -outsourcing tasks from an internal data center to a cloud provider in times of heavy load- has become more accessible to a wide range of consumers. Deciding which workloads to outsource to what cloud provider in such a setting, however, is far from trivial. The objective of this decision is to maximize the utilization of the internal data center and to minimize the cost of running the outsourced tasks in the cloud, while fulfilling the applications' quality of service constraints. We examine this optimization problem in a multi-provider hybrid cloud setting with deadline-constrained and preemptible but non-provider-migratable workloads that are characterized by memory, CPU and data transmission requirements. Linear programming is a general technique to tackle such an optimization problem.

At present, it is however unclear whether this technique is suitable for the problem at hand and what the performance implications of its use are. We therefore analyze and propose a binary integer program formulation of the scheduling problem and evaluate the computational costs of this technique with respect to the problem's key parameters. We found out that this approach results in a tractable solution for scheduling applications in the public cloud, but that the same method becomes much less feasible in a hybrid cloud setting due to very high solve time variances.

Java

A NEW ALGORITHM FOR INFERRING USER SEARCH GOALS WITH FEEDBACK SESSIONS

ABSTRACT

For a broad-topic and ambiguous query, different users may have different search goals when they submit it to a search engine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. In this paper, we propose a novel approach to infer user search goals by analyzing search engine query logs. First, we propose a framework to discover different user search goals for a query by clustering the proposed feedback sessions. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of users. Second, we propose a novel approach to generate pseudo-documents to better represent the feedback sessions for clustering. Finally, we propose a new criterion "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of our proposed methods.

Java

A LOAD BALANCING MODEL BASED ON CLOUD PARTITIONING FOR THE PUBLIC CLOUD

ABSTRACT

Load balancing in the cloud computing environment has an important impact on the performance. Good load balancing makes cloud computing more efficient and improves user satisfaction. This article introduces a better load balance model for the public cloud based on the cloud partitioning concept with a switch mechanism to choose different strategies for different situations. The algorithm applies the game theory to the load balancing strategy to improve the efficiency in the public cloud environment. Key words: load balancing model; public cloud; cloud partition; game theory

Java

A NEW ALGORITHM FOR INFERRING USER SEARCH GOALS WITH FEEDBACK SESSIONS

ABSTRACT

For a broad-topic and ambiguous query, different users may have different search goals when they submit it to a search engine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. In this paper, we propose a novel approach to infer user search goals by analyzing search engine query logs. First, we propose a framework to discover different user search goals for a query by clustering the proposed feedback sessions. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of users. Second, we propose a novel approach to generate pseudo-documents to better represent the feedback sessions for clustering. Finally, we propose a new criterion "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of our proposed methods.

Java

CONTENT SHARING OVER SMARTPHONE-BASED DELAY-TOLERANT NETWORKS

ABSTRACT

With the growing number of smartphone users, peer-to-peer ad hoc content sharing is expected to occur more often. Thus, new content sharing mechanisms should be developed as traditional data delivery schemes are not efficient for content sharing due to the sporadic connectivity between smartphones. To accomplish data delivery in such challenging environments, researchers have proposed the use of store-carry-forward protocols, in which a node stores a message and carries it until a forwarding opportunity arises through an encounter with other nodes. Most previous works in this field have focused on the prediction of whether two nodes would encounter each other, without considering the place and time of the encounter. In this paper, we propose discover-predict-deliver as an efficient content sharing scheme for delay-tolerant smartphone networks. In our proposed scheme, contents are shared using the mobility information of individuals. Specifically, our approach employs a mobility learning algorithm to identify places indoors and outdoors. A hidden Markov model is used to predict an individual's future mobility information. Evaluation based on real traces indicates that with the proposed approach, 87 percent of contents can be correctly discovered and delivered within 2 hours when the content is available only in 30 percent of nodes in the network. We implement a sample application on commercial smartphones, and we validate its efficiency to analyze the practical feasibility of the content sharing application. Our system approximately results in a 2 percent CPU overhead and reduces the battery lifetime of a smartphone by 15 percent at most.

Java

DCIM: DISTRIBUTED CACHE INVALIDATION METHOD FOR MAINTAINING CACHE CONSISTENCY IN WIRELESS MOBILE NETWORKS

ABSTRACT

This paper proposes distributed cache invalidation mechanism (DCIM), a client-based cache consistency scheme that is implemented on top of a previously proposed architecture for caching data items in mobile ad hoc networks (MANETs), namely COACS, where special nodes cache the queries and the addresses of the nodes that store the responses to these queries. We have also previously proposed a server-based consistency scheme, named SSUM, whereas in this paper, we introduce DCIM that is totally client-based. DCIM is a pull-based algorithm that implements adaptive time to live (TTL), piggybacking, and prefetching, and provides near strong consistency capabilities. Cached data items are assigned adaptive TTL values that correspond to their update rates at the data source, where items with expired TTL values are grouped in validation requests to the data source to refresh them, whereas unexpired ones but with high request rates are prefetched from the server. In this paper, DCIM is analyzed to assess the delay and bandwidth gains (or costs) when compared to polling every time and push-based schemes. DCIM was also implemented using ns2, and compared against client-based and server-based schemes to assess its performance experimentally. The consistency ratio, delay, and overhead traffic are reported versus several variables, where DCIM showed to be superior when compared to the other systems.

Java

DISTRIBUTED PROCESSING OF PROBABILISTIC TOP-K QUERIES IN WIRELESS SENSOR NETWORKS

ABSTRACT

In this paper, we introduce the notion of sufficient set and necessary set for distributed processing of probabilistic top-k queries in cluster-based wireless sensor networks. These two concepts have very nice properties that can facilitate localized data pruning in clusters. Accordingly, we develop a suite of algorithms, namely, sufficient set-based (SSB), necessary set-based (NSB), and boundary-based (BB), for intercluster query processing with bounded rounds of communications. Moreover, in responding to dynamic changes of data distribution in the network, we develop an adaptive algorithm that dynamically switches among the three proposed algorithms to minimize the transmission cost. We show the applicability of sufficient set and necessary set to wireless sensor networks with both two-tier hierarchical and tree-structured network topologies. Experimental results show that the proposed algorithms reduce data transmissions significantly and incur only small constant rounds of data communications. The experimental results also demonstrate the superiority of the adaptive algorithm, which achieves a near-optimal performance under various conditions.

Java

DYNAMIC CONTROL OF CODING FOR PROGRESSIVE PACKET ARRIVALS IN DTNS

ABSTRACT

In Delay Tolerant Networks (DTNs) the core challenge is to cope with lack of persistent connectivity and yet be able to deliver messages from source to destination. In particular, routing schemes that leverage relays' memory and mobility are a customary solution in order to improve message delivery delay. When large files need to be transferred from source to destination, not all packets may be available at the source prior to the first transmission. This motivates us to study general packet arrivals at the source, derive performance analysis of replication based routing policies and study their optimization under two hop routing. In particular, we determine the conditions for optimality in terms of probability of successful delivery and mean delay and we devise optimal policies, so-called *piecewise-threshold policies*. We account for linear block-codes and rate less random linear coding to efficiently generate redundancy, as well as for an energy constraint in the optimization. We numerically assess the higher efficiency of piecewise-threshold policies compared with other policies by developing heuristic optimization of the thresholds for all flavors of coding considered.

Java

DYNAMIC RESOURCE ALLOCATION USING VIRTUAL MACHINES FOR CLOUD COMPUTING ENVIRONMENT

ABSTRACT

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 multi-dimensional 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

Java

MULTIPARTY ACCESS CONTROL FOR ONLINE SOCIAL NETWORKS: MODEL AND MECHANISMS

ABSTRACT

Online social networks (OSNs) have experienced tremendous growth in recent years and become a de facto portal for hundreds of millions of Internet users. These OSNs offer attractive means for digital social interactions and information sharing, but also raise a number of security and privacy issues. While OSNs allow users to restrict access to shared data, they currently do not provide any mechanism to enforce privacy concerns over data associated with multiple users. To this end, we propose an approach toenable the protection of shared data associated with multiple users in OSNs. We formulate an access control model to capture the essence of multiparty authorization requirements, along with a multiparty policy specification scheme and a policy enforcement mechanism. Besides, we present a logical representation of our access control model that allows us to leverage the features of existing logic solvers to perform various analysis tasks on our model. We also discuss a proof-of-concept prototype of our approach as part of an application in Facebook and provide usability study and system evaluation of our method.

Java

NICE: NETWORK INTRUSION DETECTION AND COUNTERMEASURE SELECTION IN VIRTUAL NETWORK SYSTEMS

ABSTRACT

Cloud security is one of most important issues that has attracted a lot of research and development effort in past few years. Particularly, attackers can explore vulnerabilities of a cloud system and compromise virtual machines to deploy further large-scaleDistributed Denial-of-Service (DDoS). DDoS attacks usually involve early stage actions such as multi-step exploitation, low frequency vulnerability scanning, and compromising identified vulnerable virtual machines as zombies, and finally DDoS attacks through the compromised zombies. Within the cloud system, especially the Infrastructure-as-a-Service (IaaS) clouds, the detection of zombie exploration attacks is extremely difficult. This is because cloud users may install vulnerable applications on their virtual machines. To prevent vulnerable virtual machines from being compromised in the cloud, we propose a multi-phase distributed vulnerability detection, measurement, and countermeasure selection mechanism called NICE, which is built on attack graph based analytical models and reconfigurable virtual network-based countermeasures. The proposed framework leverages OpenFlow network programming APIs to build a monitor and control plane over distributed programmable virtual switches in order to significantly improve attack detection and mitigate attack consequences. The system and security evaluations demonstrate the efficiency and effectiveness of the proposed solution.

Java

ON THE NODE CLONE DETECTION IN WIRELESS SENSOR NETWORKS

ABSTRACT

Wireless sensor networks are vulnerable to the node clone, and several distributed protocols have been proposed to detect this attack. However, they require too strong assumptions to be practical for large-scale, randomly deployed sensor networks. In this paper, we propose two novel node clone detection protocols with different tradeoffs on network conditions and performance. The first one is based on a distributed hash table (DHT), by which a fully decentralized, key-based caching and checking system is constructed to catch cloned nodes effectively. The protocol performance on efficient storage consumption and high security level is theoretically deducted through a probability model, and the resulting equations, with necessary adjustments for real application, are supported by the simulations. Although the DHT-based protocol incurs similar communication cost as previous approaches, it may be considered a little high for some scenarios. To address this concern, our second distributed detection protocol, named randomly directed exploration, presents good communication performance for dense sensor networks, by a probabilistic directed forwarding technique along with random initial direction and border determination. The simulation results uphold the protocol design and show its efficiency on communication overhead and satisfactory detection probability.

Java

PARTICIPATORY PRIVACY ENABLING PRIVACY IN PARTICIPATORY SENSING

ABSTRACT

Participatory sensing is an emerging computing paradigm that enables the distributed collection of data by self-selected participants. It allows the increasing number of mobile phone users to share local knowledge acquired by their sensor-equipped devices (e.g., to monitor temperature, pollution level, or consumer pricing information). While research initiatives and prototypes proliferate, their real-world impact is often bounded to comprehensive user participation. If users have no incentive, or feel that their privacy might be endangered, it is likely that they will not participate. In this article, we focus on privacy protection in participatory sensing and introduce a suitable privacy-enhanced infrastructure. First, we provide a set of definitions of privacy requirements for both data producers (i.e., users providing sensed information) and consumers (i.e., applications accessing the data). Then we propose an efficient solution designed for mobile phone users, which incurs very low overhead. Finally, we discuss a number of open problems and possible research directions.

Java

PRIVACY-PRESERVING PUBLIC AUDITING FOR SECURE CLOUD STORAGE

ABSTRACT

Using cloud storage, users can remotely store their data and enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources, without the burden of local data storage and maintenance. However, the fact that users no longer have physical possession of the outsourced data makes the data integrity protection in cloud computing a formidable task, especially for users with constrained computing resources. Moreover, users should be able to just use the cloud storage as if it is local, without worrying about the need to verify its integrity. Thus, enabling public auditability for cloud storage is of critical importance so that users can resort to a third-party auditor (TPA) to check the integrity of outsourced data and be worry free. To securely introduce an effective TPA, the auditing process should bring in no new vulnerabilities toward user data privacy, and introduce no additional online burden to user. In this paper, we propose a secure cloud storage system supporting privacy-preserving public auditing. We further extend our result to enable the TPA to perform audits for multiple users simultaneously and efficiently. Extensive security and performance analysis show the proposed schemes are provably secure and highly efficient. Our preliminary experiment conducted on Amazon EC2 instance further demonstrates the fast performance of the design.

Java

PROTECTING SENSITIVE LABELS IN SOCIAL NETWORK DATA ANONYMIZATION

ABSTRACT

Privacy is one of the major concerns when publishing or sharing social network data for social science research and business analysis. Recently, researchers have developed privacy models similar to k-anonymity to prevent node reidentification through structure information. However, even when these privacy models are enforced, an attacker may still be able to infer one's private information if a group of nodes largely share the same sensitive labels (i.e., attributes). In other words, the label-node relationship is not well protected by pure structure anonymization methods. Furthermore, existing approaches, which rely on edge editing or node clustering, may significantly alter key graph properties. In this paper, we define a k-degree-l-diversity anonymity model that considers the protection of structural information as well as sensitive labels of individuals. We further propose a novel anonymization methodology based on adding noise nodes. We develop a new algorithm by adding noise nodes into the original graphwith the consideration of introducing the least distortion to graph properties. Most importantly, we provide a rigorous analysis of the theoretical bounds on the number of noise nodes added and their impacts on an important graph property. We conduct extensive experiments to evaluate the effectiveness of the proposed technique.

Java

SECURITY ANALYSIS OF A SINGLE SIGN-ON MECHANISM FOR DISTRIBUTED COMPUTER NETWORKS

ABSTRACT

Single sign-on (SSO) is a new authentication mechanism that enables a legal user with a single credential to be authenticated by multiple service providers in a distributed computer network. Recently, Chang and Lee proposed a new SSO scheme and claimed its security by providing well-organized security arguments. In this paper, however, we demonstrative that their scheme is actually insecure as it fails to meet credential privacy and soundness of authentication. Specifically, we present two impersonation attacks. The first attack allows a malicious service provider, who has successfully communicated with a legal user twice, to recover the user's credential and then to impersonate the user to access resources and services offered by other service providers. In another attack, an outsider without any credential may be able to enjoy network services freely by impersonating any legal user or a nonexistent user. We identify the flaws in their security arguments to explain why attacks are possible against their SSO scheme. Our attacks also apply to another SSO scheme proposed by Hsu and Chuang, which inspired the design of the Chang–Lee scheme. Moreover, by employing an efficient verifiable encryption of RSA signatures proposed by Ateniese, we propose an improvement for repairing the Chang–Lee scheme. We promote the formal study of the soundness of authentication as one open problem.

Java

TW-K-MEANS: AUTOMATED TWO-LEVEL VARIABLE WEIGHTING CLUSTERING ALGORITHM FOR MULTIVIEW DATA

ABSTRACT

This system proposes TW-K-means, an automated two-level variable weighting clustering algorithm for multi view data, which can simultaneously compute weights for views and individual variables. In this algorithm, a view weight is assigned to each view to identify the compactness of the view and a variable weight is also assigned to each variable in the view to identify the importance of the variable. Both view weights and variable weights are used in the distance function to determine the clusters of objects. In the new algorithm, two additional steps are added to the iterative k-means clustering process to automatically compute the view weights and the variable weights. We used two real-life data sets to investigate the properties of two types of weights in TW-k-means and investigated the difference between the weights of TW-k-means and the weights of the individual variable weighting method. The experiments have revealed the convergence property of the view weights in TW-k-means. We compared TW-k-means with five clustering algorithms on three real-life data sets and the results have shown that the TW-k-means algorithm significantly outperformed the other five clustering algorithms in four evaluation indices.

Java

ABACUS: AN AUCTION-BASED APPROACH TO CLOUD SERVICE DIFFERENTIATION

ABSTRACT

The emergence of the cloud computing paradigm has greatly enabled innovative service models, such as Platform as a Service (PaaS), and distributed computing frameworks, such as Map Reduce. However, most existing cloud systems fail to distinguish users with different preferences, or jobs of different natures. Consequently, they are unable to provide *service differentiation,* leading to inefficient allocations of cloud resources. Moreover, contentions on the resources exacerbate this inefficiency, when prioritizing crucial jobs is necessary, but impossible. Motivated by this, we propose *Abacus,* a generic resource management framework addressing this problem. Abacus interacts with users through an *auction mechanism,* which allows users to specify their priorities using *budgets,* and job characteristics via *utility functions*. Based on this information, Abacus computes the optimal allocation and scheduling of resources. Meanwhile, the auction mechanism in Abacus possesses important properties including incentive compatibility (i.e., the users' best strategy is to simply bid their true budgets and job utilities) and monotonicity (i.e., users are motivated to increase their budgets in order to receive better services). In addition, when the user is unclear about her utility function, Abacus automatically learns this function based on statistics of her previous jobs. An extensive set of experiments, running on Hadoop, demonstrate the high performance and other desirable properties of Abacus.

Java

DISTRIBUTED PROCESSING OF PROBABILISTIC TOP-K QUERIES IN WIRELESS SENSOR NETWORKS

ABSTRACT

In this paper, we introduce the notion of sufficient set and necessary set for distributed processing of probabilistic top-k queries in cluster-based wireless sensor networks. These two concepts have very nice properties that can facilitate localized data pruning in clusters. Accordingly, we develop a suite of algorithms, namely, sufficient set-based (SSB), necessary set-based (NSB), and boundary-based (BB), for intercluster query processing with bounded rounds of communications. Moreover, in responding to dynamic changes of data distribution in the network, we develop an adaptive algorithm that dynamically switches among the three proposed algorithms to minimize the transmission cost. We show the applicability of sufficient set and necessary set to wireless sensor networks with both two-tier hierarchical and tree-structured network topologies. Experimental results show that the proposed algorithms reduce data transmissions significantly and incur only small constant rounds of data communications. The experimental results also demonstrate the superiority of the adaptive algorithm, which achieves a near-optimal performance under various conditions.

Java

OPTIMIZING CLOUD RESOURCES FOR DELIVERING IPTV SERVICES THROUGH VIRTUALIZATION

ABSTRACT

Virtualized cloud-based services can take advantage of statistical multiplexing across applications to yield significant cost savings to the operator. However, achieving similar benefits with real-time services can be a challenge. In this paper, we seek to lower a provider's costs of real-time IPTV services through a virtualized IPTV architecture and through intelligent time shifting of service delivery. We take advantage of the differences in the deadlines associated with Live TV versus Video-on-Demand (VoD) to effectively multiplex these services. We provide a generalized framework for computing the amount of resources needed to support multiple services, without missing the deadline for any service. We construct the problem as an optimization formulation that uses a generic cost function.

We consider multiple forms for the cost function (e.g., maximum, convex and concave functions) to reflect the different pricing options. The solution to this formulation gives the number of servers needed at different time instants to support these services. We implement a simple mechanism for time-shifting scheduled jobs in a simulator and study the reduction in server load using real traces from an operational IPTV network. Our results show that we are able to reduce the load by _ 24% (compared to a possible _ 31%). We also show that there are interesting open problems in designing mechanisms that allow time-shifting of load in such environments.

Java

SPOC: A SECURE AND PRIVACY-PRESERVING OPPORTUNISTIC COMPUTING FRAMEWORK FOR MOBILE-HEALTHCARE EMERGENCY

ABSTRACT

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.

Java

WARNINGBIRD: A NEAR REAL-TIME DETECTION SYSTEM FOR SUSPICIOUS URLS IN TWITTER STREAM

ABSTRACT

Twitter is prone to malicious tweets containing URLs for spam, phishing, and malware distribution. Conventional Twitter spam detection schemes utilize account features such as the ratio of tweets containing URLs and the account creation date, or relation features in the Twitter graph. These detection schemes are ineffective against feature fabrications or consume much time and resources. Conventional suspicious URL detection schemes utilize several features including lexical features of URLs, URL redirection, HTML content, and dynamic behavior. However, evading techniques such as time-based evasion and crawler evasion exist. In this paper, we propose WARNINGBIRD, a suspicious URL detection system for Twitter. Our system investigates correlations of URL redirect chains extracted from several tweets. Because attackers have limited resources and usually reuse them, their URL redirect chains frequently share the same URLs. We develop methods to discover correlated URL redirect chains using the frequently shared URLs and to determine their suspiciousness. We collect numerous tweets from the Twitter public timeline and build a statistical classifier using them. Evaluation results show that our classifier accurately and efficiently detects suspicious URLs. We also present

WARNINGBIRD as a near real-time system for classifying suspicious URLs in the Twitter stream.

Java

A SURVEY OF INTRUSION DETECTION SYSTEMS IN WIRELESS SENSOR NETWORKS

ABSTRACT

Wireless Sensor Networking is one of the most promising technologies that have applications ranging from health care to tactical military. Although Wireless Sensor Networks (WSNs) have appealing features (e.g., low installation cost, unattended network operation), due to the lack of a physical line of defense (i.e., there are no gateways or switches to monitor the information flow), the security of such networks is a big concern, especially for the applications where confidentiality has prime importance. Therefore, in order to operate WSNs in a secure way, any kind of intrusions should be detected before attackers can harm the network (i.e., sensor nodes) and/or information destination (i.e., data sink or base station). In this article, a survey of the state-of-the-art in Intrusion Detection Systems (IDSs) that are proposed for WSNs is presented. Firstly, detailed information about IDSs is provided. Secondly, a brief survey of IDSs proposed for Mobile Ad-Hoc Networks (MANETs) is presented and applicability of those systems to WSNs are discussed. Thirdly, IDSs proposed for WSNs are presented. This is followed by the analysis and comparison of each scheme along with their advantages and disadvantages. Finally, guidelines on IDSs that are potentially applicable to WSNs are provided. Our survey is concluded by highlighting open research issues in the field.

Java

MINING THE SITUATION: SPATIOTEMPORAL TRAFFIC PREDICTION WITH BIG DATA

ABSTRACT

With the vast availability of traffic sensors from which traffic information can be derived, a lot of research effort has been devoted to developing traffic prediction techniques, which in turn improve route navigation, traffic regulation, urban area planning etc. One key challenge in traffic prediction is how much to rely on prediction models that are constructed using historical data in real-time traffic situations, which may differ from that of the historical data and change over time. In this paper, we propose a novel online framework that could learn from the current traffic situation (or context) in real-time and predict the future traffic by matching the current situation to the most effective prediction model trained using historical data. As real-time traffic arrives, the traffic context space is adaptively partitioned in order to efficiently estimate the effectiveness of each base predictor in different situations. We obtain and prove both short-term and long-term performance guarantees (bounds) for our online algorithm. The proposed algorithm also works effectively in scenarios where the true labels (i.e. realized traffic) are missing or become available with delay. Using the proposed framework, the context dimension that is the most relevant to traffic prediction can also be revealed, which can further reduce the implementation complexity as well as inform traffic policy making. Our experiments with real-world data in reallife conditions show that the proposed approach significantly outperforms existing solutions.

Java