Research Publications:

Risk-Aware Management of Virtual Resources in Access Controlled Service-Oriented Cloud Datacenters [IEEE Transactions on Cloud Computing 2018]

For economic benefits and efficient management of resources, organizations are increasingly moving towards the paradigm of “cloud computing” by which they are allowed on-demand delivery of hardware, software and data as services. However, there are many security challenges which are particularly exacerbated by the multitenancy and virtualization features of cloud computing that allow sharing of resources among potentially untrusted tenants in access controlled cloud datacenters. This can result in increased risk of data leakage. To address this risk vulnerability, we propose an efficient risk-aware virtual resource assignment mechanism for clouds multitenant environment. In particular, we introduce the notion of sensitivity in datacenters and the objective is to minimize the risk of data leakage. In addition, the risk should not exceed in high sensitivity datacenters in comparison to low sensitivity datacenters. We present three assignment heuristics and compare their relative performance.

A Privacy Mechanism for Access Controlled Graph Data [IEEE TDSC 2018]

There has been significant interest in the development of anonymization schemes for publishing graph data. However, privacy is a major concern in dealing with graph data. In this paper, an integrated framework for ensuring privacy in the presence of an authorization mechanism is proposed. Access control mechanisms provide additional safeguard against data breaches and ensure that only authorized information is available to end-users based on their assigned roles. The integrated framework highlights a tradeoff between privacy and authorized privileges. To attain a pre-specified privacy level, access privileges might need to be relaxed. For the proposed framework, we formulate the k-anonymous Bi-objective Graph Partitioning (k-BGP) problem and provide its hardness results. Heuristics solutions are developed to solve the constraint problem. The framework provides an anonymous view based on the target class of role-based workloads for graph data. The proposed heuristics are empirically evaluated and a detailed security analysis of the framework in terms of risk associated with re-identification attack is conducted.

Architectures for Detecting Real-time Multiple Multi-stage Network Attacks Using Hidden Markov Model

With the growing Cyber threats, the need to develop high assurance Cyber systems is becoming increasingly important. The objective of this paper is to address the challenges of modeling and detecting sophisticated and diversified network attacks. Using one of the important statistical machine learning (ML) techniques, Hidden Markov Models (HMM), we develop two architectures that can detect and track in real-time the progress of these organized attacks. These architectures are based on developing a database of HMM templates and exhibit varying performance and complexity. For performance evaluation, in the presence of multiple multi-stage attack scenarios, various metrics are proposed which include (1) attack risk probability, (2) detection error rate, and (3) the number of correctly detected stages. Extensive simulation experiments are used based on the DARPA2000 dataset to demonstrate the efficacy of the proposed architectures.

Efficient and Scalable Integrity Verification of Data and Query Results for Graph Databases [IEEE Transactions on Knowledge and Data Engineering 2017]

Graphs are used for representing and understanding objects and their relationships for numerous applications such as social networks, Semantic Webs, and biological networks. Integrity assurance of data and query results for graph databases is an essential security requirement. In this paper, we propose two efficient integrity verification schemes—HMACs for graphs (gHMAC) for two-party data sharing, and redactable HMACs for graphs (rgHMAC) for third-party data sharing, such as a cloud-based graph database service. We compute one HMAC value for both the schemes and two other verification objects for rgHMAC scheme that are shared with the verifier. We show that the proposed schemes are provably secure with respect to integrity attacks on the structure and/ or content of graphs and query results. The proposed schemes have linear complexity in terms of the number of vertices and edges in the graphs, which is shown to be optimal. Our experimental results corroborate that the proposed HMAC-based schemes for graphs are highly efficient as compared to the digital signature-based schemes—computation of HMAC tags is about 10 times faster than the computation of digital signatures.

Composability Verification of Multi-Service Workflows in a Policy-Driven Cloud Computing Environment [IEEE Transactions on Dependable and Secure Computing 2017]

The emergence of cloud computing infrastructure and Semantic Web technologies has created unprecedented opportunities for composing large-scale business processes and workflow-based applications that span multiple organizational domains. A key challenge related to composition of such multi-organizational business processes and workflows is posed by the security and access control policies of the underlying organizational domains. In this paper, we propose a framework for verifying secure composability of distributed workflows in an autonomous multi-domain environment. The objective of workflow composability verification is to ensure that all the users or processes executing the designated workflow tasks conform to the time-dependent security policy specifications of all collaborating domains. A key aspect of such verification is to determine the time-dependent schedulability of distributed workflows, assumed to be invoked on a recurrent basis. We use a two-step approach for verifying secure workflow composability. In the first step, a distributed workflow is decomposed into domain-specific projected workflows and is verified for conformance with the respective domain's security and access control policy. In the second step, the cross-domain dependencies amongst the workflow tasks performed by different collaborating domains are verified.

Intelligent Shelter Allotment for Emergency Evacuation Planning: A Case Study of Makkah [IEEE Intelligent Systems 2015]

The emergence of cloud computing infrastructure and Semantic Web technologies has created unprecedented opportunities for composing large-scale business processes and workflow-based applications that span multiple organizational domains. A key challenge related to composition of such multi-organizational business processes and workflows is posed by the security and access control policies of the underlying organizational domains. In this paper, we propose a framework for verifying secure composability of distributed workflows in an autonomous multi-domain environment. The objective of workflow composability verification is to ensure that all the users or processes executing the designated workflow tasks conform to the time-dependent security policy specifications of all collaborating domains. A key aspect of such verification is to determine the time-dependent schedulability of distributed workflows, assumed to be invoked on a recurrent basis. We use a two-step approach for verifying secure workflow composability. In the first step, a distributed workflow is decomposed into domain-specific projected workflows and is verified for conformance with the respective domain's security and access control policy. In the second step, the cross-domain dependencies amongst the workflow tasks performed by different collaborating domains are verified.

A Knowledge Driven Agent-Based Semantic Model for Epidemic Surveillance [International Journal of Semantic Computing 2015]

In this paper we propose a probabilistic approach to synthesize an agent-based heterogeneous semantic model depicting population interaction and analyzing the spatio-temporal dynamics of an airborne epidemic, such as influenza, in a metropolitan area. The methodology is generic in nature and can generate a baseline population for cities for which detailed population summary tables are not available. The joint probabilities of population demographics are estimated using the International Public Use Microsimulation Data (IPUMS) sample set. Agents are assigned various activities based on several characteristics. The agent-based model for the city of Lahore, Pakistan is synthesized and a rule based disease spread model of influenza is simulated. The simulation results are visualized to produce semantic analysis for the spatio-temporal dynamics of the epidemic. The results show that the proposed model can be used by officials and medical experts to simulate an outbreak.

Precision-Bounded Access Control Using Sliding-Window Query Views for Privacy-Preserving Data Streams [IEEE Transactions on Knowledge and Data Engineering 2015]

Access control mechanisms and Privacy Protection Mechanisms (PPM) have been proposed for data streams. The access control for a data stream allows roles access to tuples satisfying an authorized predicate sliding-window query. Sharing the sensitive stream data without PPM can compromise the privacy. The PPM meets privacy requirements, e.g., k-anonymity or l-diversity by generalization of stream data. Imprecision introduced by generalization can be reduced by delaying the publishing of stream data. However, the delay in sharing the stream tuples to achieve better accuracy can lead to false-negatives if the tuples are held by PPM while the query predicate is evaluated. Administrator of an access control policy defines the imprecision bound for each query. The challenge for PPM is to optimize the delay in publishing of stream data so that the imprecision bound for the maximum number of queries is satisfied. We formulate the precision-bounded access control for privacy-preserving data streams problem, present the hardness results, provide an anonymization algorithm, and conduct experimental evaluation of the proposed algorithm. Experiments demonstrate that the proposed heuristic provides better precision for a given data stream access control policy as compared to the minimum or maximum delay heuristics proposed in existing literature.

A Framework for Composition and Enforcement of Privacy-Aware and Context-Driven Authorization Mechanism for Multimedia Big Data [IEEE Transactions on Multimedia 2015]

The proliferation of multimedia big data for dissemination and sharing of massive amounts of information raises important security and privacy concerns. One such concern is the composition and enforcement of privacy policies in order to securely manage access of multimedia big data. Several researchers have pointed out that for proper enforcement of privacy policies, the privacy requirements should be captured in access control systems. In this paper, we propose a hybrid approach where privacy requirements are captured in an access control system and present a framework for composition and enforcement of privacy policies. The focus is to allow a user, not a system or security administrator to compose conflict free policies for their online multimedia data. An additional requirement is that such a policy be context-aware. We also present a methodology for verifying the privacy policy in order to ensure correctness and logical consistency. The verification process is also used to ensure that sensitive security requirements are not violated when privacy rules are enforced. A prototype, named Intelligent Privacy Manager (iPM), has been implemented for sharing of multimedia big data in a secure and private manner.

Risk-Aware Virtual Resource Management for Multitenant Cloud Datacenters [IEEE Cloud Computing 2015]

The cloud computing platform-as-a-service (PaaS) paradigm allows application developers to deploy big data applications in the cloud. These applications can be found in the areas of healthcare, e-government, science, and business.1 PaaS cloud providers can host customer data stores on premise and outsource the computation to virtual resources from multiple infrastructure-as-a-service (IaaS) cloud providers. These virtual resources can be hosted by multitenant public cloud providers such as the Amazon Elastic Compute Cloud (EC2). The sheer size of big data poses serious security challenges for these applications. The backend data store can use an access control mechanism to isolate and enforce controlled data sharing.2 However, when the data is transferred from the backend data store to application logic, it can be leaked through virtual resource vulnerabilities. In a multitenant environment, untrusted tenants can exploit these vulnerabilities, increasing the data leakage risk.

Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data [IEEE Transactions on Knowledge and Data Engineering 2014]

Access control mechanisms protect sensitive information from unauthorized users. However, when sensitive information is shared and a Privacy Protection Mechanism (PPM) is not in place, an authorized user can still compromise the privacy of a person leading to identity disclosure. A PPM can use suppression and generalization of relational data to anonymize and satisfy privacy requirements, e.g., k-anonymity and l-diversity, against identity and attribute disclosure. However, privacy is achieved at the cost of precision of authorized information. In this paper, we propose an accuracy-constrained privacy-preserving access control framework. The access control policies define selection predicates available to roles while the privacy requirement is to satisfy the k-anonymity or l-diversity. An additional constraint that needs to be satisfied by the PPM is the imprecision bound for each selection predicate. The techniques for workload-aware anonymization for selection predicates have been discussed in the literature. However, to the best of our knowledge, the problem of satisfying the accuracy constraints for multiple roles has not been studied before. In our formulation of the aforementioned problem, we propose heuristics for anonymization algorithms and show empirically that the proposed approach satisfies imprecision bounds for more permissions and has lower total imprecision than the current state of the art.

Security of Graph Data: Hashing Schemes and Definitions [ACM CODASPY 2014]

Use of graph-structured data models is on the rise – in graph databases, in representing biological and healthcare data as well as geographical data. In order to secure graph structured data, and develop cryptographically secure schemes for graph databases, it is essential to formally define and develop suitable collision resistant one-way hashing schemes and show them they are efficient. The widely used Merkle hash technique is not suitable as it is, because graphs may be directed acyclic ones or cyclic ones. In this paper, we are addressing this problem. Our contributions are: (1) define the practical and formal security model of hashing schemes for graphs, (2) define the formal security model of perfectly secure hashing schemes, (3) describe constructions of hashing and perfectly secure hashing of graphs, and (4) performance results for the constructions. Our constructions use graph traversal techniques, and are highly efficient for hashing, redaction, and verification of hashes graphs. We have implemented the proposed schemes, and our performance analysis on both real and synthetic graph data sets support our claims.

Policy-driven High Assurance Cyber Infrastructure-Based Systems [IEEE HASE 2014]

The objective of this paper is to present major challenges and a framework for modeling and managing contextaware policy-driven Cyber Infrastructure-Based Systems (CIBS). With the growing reliance on Cyber technology providing solutions for a broad range of CIBS applications, comes the high assurance challenges in terms of reliability, trustworthiness and vulnerabilities. The paper proposes a development framework to allow dynamic reconfigurability of CIBS components under various contexts to achieve a desired degree of assurance.

A Distributed Access Control Architecture for Cloud Computing [IEEE Software 2012]

The growing popularity of cloud computing draws attention to its security challenges, which are particularly exacerbated due to resource sharing. Cloud computing's multitenancy and virtualization features pose unique security and access control challenges due to sharing of physical resources among potential untrusted tenants, resulting in an increased risk of side-channel attacks. Additionally, the interference of multitenancy computation can result in unauthorized information flow. Heterogeneity of services in cloud computing environments demands varying degrees of granularity in access control mechanisms. Therefore, an inadequate or unreliable authorization mechanism can significantly increase the risk of unauthorized use of cloud resources and services. In addition to preventing such attacks, a fine-grained authorization mechanism can assist in implementing standard security measures. Such access control challenges and the complexities associated with their management call for a sophisticated security architecture that not only adequately captures access management requirements but also ensures secure interoperation across multiple clouds. We present a distributed access control architecture for multitenant and virtualized environments. The design of this architecture is based on the principles from security management and software engineering. From a security management perspective, the goal is to meet cloud users' access control requirements. From a software engineering perspective, the goal is to generate detailed specifications of such requirements.

M3 : Stream Processing on Main-Memory MapReduce [IEEE ICDE 2012]

The continuous growth of social web applications along with the development of sensor capabilities in electronic devices is creating countless opportunities to analyze the enormous amounts of data that is continuously steaming from these applications and devices. To process large scale data on large scale computing clusters, MapReduce has been introduced as a framework for parallel computing. However, most of the current implementations of the MapReduce framework support only the execution of fixed-input jobs. Such restriction makes these implementations inapplicable for most streaming applications, in which queries are continuous in nature, and input data streams are continuously received at high arrival rates. In this demonstration, we showcase M3, a prototype implementation of the MapReduce framework in which continuous queries over streams of data can be efficiently answered. M3 extends Hadoop, the open source implementation of MapReduce, bypassing the Hadoop Distributed File System (HDFS) to support main-memory-only processing. Moreover, M3 supports continuous execution of the Map and Reduce phases where individual Mappers and Reducers never terminate.

A Framework for Verification and Optimal Reconfiguration of Event-driven Role Based Access Control Policies [ACM SACMAT 2012]

Role based access control (RBAC) is the de facto model used for advanced access control due to its inherent richness and flexibility. Despite its great success at modeling a variety of organizational needs, maintaining large complex policies is a challenging problem. Conflicts within policies can expose the underlying system to numerous vulnerabilities and security risks. Therefore, more comprehensive verification tools for RBAC need to be developed to enable effective access control. In this paper, we propose a verification framework for detection and resolution of inconsistencies and conflicts in policies modeled through event-driven RBAC, an important subset of generalized temporal RBAC applicable to many domains, such as SCADA systems. We define the conflict resolution problem and propose an integer programming based heuristic. The proposed approach is generic and can be tuned to a variety of optimality measures.

Specification and Verification of a Context-Based Access Control Framework for Cyber Physical Systems [CERIAS Tech Report 2011-19]

Cyber Physical Systems (CPS) are complex systems that operate in a dynamic environment where security characteristics of contexts are unique, and uniform access to secure resources anywhere anytime to mobile entities poses daunting challenges. To capture context parameters such as location and time in an access control policy for CPS, we propose a Generalized Spatio- Temporal RBAC (GST-RBAC) model. In this model spatial and temporal constraints are defined for role enabling, user-role assignment, role-permission assignment, role activation, separation of duty and role hierarchy. The inclusion of multiple types of constraints exposes the need of composing a policy which is verifiable for consistency. The second contribution in this paper is GST-RBAC policy specification and verification framework using light weight formal modeling language, Alloy. The analysis assists in consistency verification leading to conflict free composition of the actual policy for implementation for CPS.

The Similarity-aware Relational Database Set Operators [Information Systems Journal, 2016]

Finding data items in the proximity of a query item can be formulated as a similarity operation. Identifying these similarities in large datasets is an essential operation in several applications such as bioinformatics, pattern recognition, and data integration. To make a relational database management system similarity-aware, the core relational operators have to be extended. While similarity-awareness has been introduced in database engines for relational operators such as joins and group-by, little has been achieved for relational set operators, namely Intersection, Difference, and Union. In this paper, we propose to extend the semantics of relational set operators to take into account the similarity of values. We develop efficient query processing algorithms for evaluating them, and implement these operators inside an open-source database system, namely PostgreSQL. By extending several queries from the TPC-H benchmark to include predicates that involve similarity-based set operators, we perform extensive experiments that demonstrate up to three orders of magnitude speedup in performance over equivalent queries that only employ regular operators.

Tornado: A Distributed SpatioTextual Stream Processing System [VLDB 2015 Demo paper]

We have developed a prototype open-source system termed Tornado. Tornado is motivated by the needs for efficiently querying Big Spatial Data. Nowadays, most spatial data is also associated with text data. The widespread use of location-aware devices together with the increased popularity of micro-blogging applications (e.g., Twitter) has led to the creation of large streams of spatio-textual data. In order to serve real-time applications, the processing of these large-scale spatio-textual streams needs to be distributed. However, existing distributed stream processing systems (e.g., Spark and Storm) are not optimized for spatial/textual content. In this demonstration, we introduce Tornado, a distributed in-memory spatio-textual stream processing server that extends Storm. To efficiently process spatio-textual streams, Tornado introduces a spatio-textual indexing layer to the architecture of Storm. The indexing layer is adaptive, i.e., dynamically re-distributes the processing across the system according to changes in the data distribution and/or query workload. In addition to keywords, higher-level textual concepts are identified and are semantically matched against spatio-textual queries. Tornado provides data de-duplication and fusion to eliminate redundant textual data. We demonstrate a prototype of Tornado running against real Twitter streams, where the users can register continuous or snapshot spatio-textual queries using a map-assisted query interface.