Secure Data Communications and Targeted Information Propagation

Keywords: Privacy, Access Control, User Profiling, Cloud Data Management, Data Leakage Detection, Encrypted Search, Blockchain

Problem Statement: My research aims to provide the following:

1. Data protection at rest and in transit, providing data leakage prevention and detection with threat assessment as well as enforcing role-based and attribute-based access control.
2. Encrypted search over encrypted data, with capabilities of performing data analysis over encrypted data.
3. Collaborative secure software development system which collects provenance data and provides its integrity
4. Targeted Information Propagation based on machine learning algorithms

In service-oriented architecture (SOA), services can communicate and share data among themselves. Services and associated data can be hosted by cloud platforms, which are vulnerable to large attack surface that could violate data privacy. My colleagues and I have designed a solution that provides data protection in transit and at rest. This solution also provides data leakage prevention and detection for multiple leakage scenarios that can be performed by an external attacker or a malicious insider. The prototype called "WAXEDPRUNE" (Web-based Access to Encrypted Data Processing in Untrusted Environments), implemented in collaboration with Northrop Grumman, MIT and W3C, was demonstrated at Northrop Grumman Tech Expo in 2016. The approach ensures that each service can access only those data subsets for which the service is authorized. Encrypted search and extensive data analysis over encrypted data records are supported as well. Northrop Grumman funded the projects I have been working in for 5 consecutive years, renewing the funding every year, starting from 2014. "WAXEDPRUNE" project received two awards. In April 2017, it was selected to be funded as #1 out of 21 research projects by Corporate Partners of Computer Science Department and CERIAS at Purdue University, including Northrop Grumman, Intel, Qualcomm, Raytheon, Eli Lilly. In March 2015, the research poster "PD3: Policy-based Distributed Data Dissemination", which is based on the predecessor of "WAXEDPRUNE", was selected by Corporate Partners as #1 out of 43 at 16th CERIAS Security Symposium.

Data protection method is based on using Active Bundles (AB). AB is a self-protecting structure that contains data in encrypted form, access control policies and a policy enforcement engine (Virtual Machine). Data are stored in AB in the form of key-value pairs with encrypted values. Each data subset is encrypted with its own symmetric key using a novel "on-the-fly" key generation scheme, based on the execution flow. As a use case, Electronic Health Record (EHR) of a patient can be stored as an AB. An example of a key-value pair stored in an AB, which represents EHR, is: *{"ab.patientID" : "***Enc***(0123456)"}.* AB provides data leakage prevention since access to unauthorized data will be denied by the AB kernel. An attempt to decrypt data made by an unauthorized service will be recorded by a trusted Central Monitor (CM). I address a challenging data leakage scenario when a service (insider) authorized for data d_i can get these data from an AB, store them locally and then send them without the AB behind the scene to an unauthorized service as plaintext[1]. I aim to help investigating the leakage and do forensics based on provenance records stored on CM each time a data request is served by the AB. To mitigate data leakage problem, I embed digital and visual watermarks into data.

To ensure *integrity, trust and immutability* of software and data (user data and attack event data), me and my colleagues designed blockchain-based technology[2], "Blockhub", for collaborative software development. Every access, transfer and update of data and software is recorded in the blockchain public ledger, can be verified any time in the future and cannot be erased or repudiated by invokers. "Blockhub" allows tracking and controlling what data or software modules are shared between entities across multiple security domains.

In my other research project, "Targeted Information Propagation", the data from heterogeneous sources are extracted, transformed, consolidated, cleaned and then are fed into the Knowledge Discovery Engine, which applies supervised and unsupervised learning algorithms to detect data patterns relevant to user's needs and to push the results to the relevant subjects or relevant roles.

[1] D. Ulybyshev, B. Bhargava, A. Alsalem, "Secure Data Exchange and Data Leakage Detection in Untrusted Cloud", Springer Journal on 1-st Intl Conf. on Applications of Computing and Communication Technologies (ICACCT), 2018, pp. 99-113

[2] D. Ulybyshev, M. Villarreal, B. Bhargava, G. Mani, S. Seaberg, P. Conoval, R. Pike, J. Kobes "Blockhub: Blockchain-based Software Development System for Untrusted Environments", IEEE CLOUD, San-Francisco, 2018