Access Control and Privacy Policy
Challenges in Big Data

Position Paper
NSF Workshop on Big Data Security and Privacy

Ram Krishnan
University of Texas at San Antonio
Ram.Krishnan@utsa.edu

The proliferation of new technologies and applications have enabled collection of data at a scale that was unimaginable a decade ago. Sophisticated analytics techniques are being developed to analyze such vast amount of data to discover meaningful patterns and gain valuable insights to facilitate decision-making. A salient aspect is that data are being used in unplanned ways, and for purposes that were unforeseen at the time of collection. This paper discusses a modest set of 2 major challenges relating to access control and privacy policy specification and enforcement in big data applications. It is strongly emphasized that if these challenges are addressed to a large degree, it can alleviate many of the concerns that arise in big data applications.

Challenge #1: Fine-Grained Access Control to Large-Scale Disparate Data

Big data is often characterized to have a variety of data including structured (e.g. relational), semi-structured (e.g. JSON or XML formatted) and unstructured data (e.g. images and text files). Traditional approaches such as access control lists or role-based access control for specifying access control policies can soon become cumbersome and unmanageable due to different granularities and the scale of data in big data applications. Access control is one of the most powerful and fundamental ways of risk mitigation in any application. Interestingly, despite security concerns raised by big data applications, there is minimal research in this area. We emphasize that insights on administrative and operational access control aspects of big data is an essential pillar to secure future applications. Conceptually, attribute based access control (ABAC) seems to provide the necessary flexibility in dealing with scale. We envision that a foundational theory of attribute-based access control needs to be developed for successful adoption of ABAC in big data applications.
Challenge #2: Practical Data Usage and Purpose-Specific Privacy Policy Specification and Compliance:

Abstractly speaking, we have consumers or end-users that provide data, and service providers that often times monetize the collected data in order to provide a service. The state-of-the-art in consumer privacy compliance is via notice and consent. The notice is often hidden in legalese and it is the burden of the consumers to “take it” or “leave it”. To provide any level of privacy guarantee to an average consumer, the above system must be re-evaluated. It is critical to communicate how the collected data will be used and for what purpose. What are the ways to communicate privacy policies effectively to end-users? Can the notion of purpose and use be codified, and can a taxonomy be created (e.g. DHS threat level)? How can such privacy policies be translated faithfully into enforceable policies? How can complex big data applications comply faithfully (or approximately) to stated privacy policies?

Closing Remark: Consumers see value in the services provided by service providers, and hence are willing to share their data. The lack of visibility into access control and privacy policies on large-scale data is the major root cause for the concerns in data sharing and analytics practices of service providers. The 2 challenges presented in this paper are in line with these concerns. Given the prevalence of data collection and a level of forced opaqueness in the sharing and analysis practices of service providers, developing practical techniques that address these challenges can enable service providers to adopt these techniques. Contrary to the popular thought of losing customers, this can indeed catalyze new business models to attract more.