Fraud, Anonymization and Privacy in the Internet: A Data-oriented Perspective (Keynote Address, COMAD 2009)

Amr El Abbadi
Department of Computer Science
University of California, Santa Barbara
amr@cs.ucsb.edu

Abstract

Data is everywhere, and manifests itself in various formats. Data can be publicly available and can be privately owned. Data can be persistent on a server or ephemeral in a data stream. Society depends on data and hence the security, privacy and reliability of the data are critical in diverse ways. In this talk I will touch of varying security aspects of data in different contexts that arise in todays internet applications.

In Internet advertising networks, Internet advertisers pay Internet publishers to display advertisements on their Web sites and drive traffic to the advertisers from surfer's clicks. Internet advertising is crucial for the thriving of the entire Internet, since it allows producers to advertise their products, and hence contributes to the well being of e-commerce. Some publishers, however, are dishonest, and use automation to generate traffic to defraud the advertisers. We describe the advertising network model, and discuss the issue of fraud that is an integral problem in such a setting. Our approach builds on recently developed data streaming summarization techniques.

The increasing popularity of social networks has initiated a fertile research area in information extraction and data mining. Although such analysis can facilitate better understanding of sociological, behavioral, and other interesting phenomena, there is growing concern about personal privacy being breached, thereby requiring effective anonymization techniques. In this context we will describe methods for anonymizing the relevant contents of the social network graph while preserving critical properties of the graph. Our approach uses linear programming to ensure robust edge weight anonymization.

The advent of cloud computing, as well as the ubiquitous availability of internet information service providers has resulted in different settings where users are concerned about the privacy of their data. This may arise when private data is outsourced to a service provider, and the client wishes to retrieve some

of this private data in an efficient privacy preserving manner, or when a client wishes to retrieve publicly available data in way that does not reveal the specific interests of the client to the service provider. We will explore different privacy preserving approaches in various data retrieval settings.