University POLITEHNICA of Bucharest

Faculty of Automatic Control and Computers

Department of Computer Science and Engineering

Abstract

Doctoral Thesis

Diff-Anonym Algorithm to Provide Privacy for Patients' Location based Services

Based on Big Data in Cloud

Author: **Imad Al - Tameemi**

Scientific Supervisor: **Prof. dr. ing. Nicolae Ţăpuş**

Bucharest 2019

Abstract

The use of mobile phones, computers, cloud computing, etc. in recent years have generated zero trust. Privacy of information has been overlooked in the fast development of embedded sensors, systems, techniques, processing power, storage capacities of devices and networks. All over the world, computers, mobile phones, cloud computing and other technologies are used by over 7 billion people. All these technologies are connected and are able to collect and exchange information. An individual has the right to select the personal information shared. Location information is an important part of the private information related to an individual. It is also the main focus of the present study, which mainly considers the concerns of disclosing location history data to third parties, according to user preferences. The increasing usage of GSNs, or Geosocial Networks as Twitter, Instagram or Facebook, has led to public sharing of the user's location, through "check-in" services of the networks. Based on this location information, the actions of the user can be retraced, and even future location can be anticipated. However, the risk of unauthorized disclosure is high related to sensitive locations, such as hospitals, home address, or banks' location. Data breaches are possible to occur on the data storage servers, considered non-trusted servers, used for collecting private information. Storage servers also present the risk of single point servers, in which case an attack on a single device can compromise the private information stored on this device. In case of private information disclosure, the home address of a user could be uncovered, the user could receive unrequested advertisement or even medical conditions of the individual could be uncovered. As a result, these experiences could be very unpleasant, or could even generate complications. People who have access to location information belong to different social groups, such as colleagues, friends or family. Part of our interest is also the information type shared with these groups. An individual might choose to share certain location information only with some circles, such as nightlife check-ins should not be visible to the family group. The levels of data privacy are modifying the business models, from social insurance to retail and farming. Information related to all activities is collected at very high rates, which opens possibilities to adjust the practices and procedures in order to increase efficiency. In a commercial environment, selling and distributing an item after its development, is expected. Big retailers are able to establish

the purchase needs of the potential customers based on client information. The use of mobile phones, computers, cloud computing, etc. in recent years have generated zero trust. A new concept has been created, named participatory sensing, in which individuals are able to exchange through technologies they use, information sensed within their surrounding environment. Privacy methods and algorithms are not sufficient for protecting the individual's privacy. There are numerous methods available, such as technologies and systems that enable constant behaviour, actions and preferences analysis, alongside data mining, analysing and cataloguing techniques. Government agencies, business or organizations that collect and analyse data, such as search queries, telephone call logs, or users' locations information, have been under privacy attacks. Users have started to distrust these entities, which led to avoiding the use of numerous applications or submitting fake data. Users feel generally more comfortable when less information are requested by the applications, or they are optional. Another possibility is to allow users to obscure information, if supported by the system. This dissertation also presents the most important elements that should be preserved in terms of location information or personal data. The general opinion of the users is that, by hiding their personal information, privacy attacks can be avoided. They are much less concerned about their location information. The first chapter of the dissertation presents an overview of the background of the thesis topics and the main concerns of customer privacy for location-based services in big data. The second chapter presents the literature review on the topics in general. In addition, we explain the results of the research considered and examine the research that follow solutions for general data disclosure versus location information disclosure in terms of customer privacy impact. In the third chapter, the effect on customer privacy of disclosing general information versus disclosing location information is considered. The analysis is conducted in two stages – create and publish an online questionnaire and analyse the data collected using this survey. The purpose of the analysis was to identify the effect of disclosing location information compared to the effect of disclosing general information. Based on the results, we concluded that any data disclosure has negative effects. The fourth chapter includes a presentation of the proposed method that ensures privacy protection for normal and big data. The implementation of the Diff-Anonym algorithm is illustrated in chapter four with two examples of applications based on normal and on big data. In the fifth chapter, the methods proposed are combined. The aim is to ensure data privacy of the patients' location for location-based services used in services integrated in health-emergency alert modules.