
Interview with Yücel Saygin on Privacy in Spatiotemporal Data Analysis

Christine Körner



Yücel Saygin is an Associate Professor of Computer Science with the Faculty of Engineering and Natural Sciences at Sabanci University in Turkey. He received his B.S., M.S., and Ph.D. degrees from the Department of Computer Engineering at Bilkent University in 1994, 1996, and 2001, respectively. His main research interests include data mining and privacy preserving data management. Yücel Saygin has published in international journals like ACM Transactions on Database Systems, VLDB Journal, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Engineering Management, and in proceedings of international conferences. He co-chaired various conferences and workshops in the area of data mining and privacy preserving data management. He is currently the coordinator of the MODAP (Mobility, Data Mining, and Privacy) project funded by EU FP7 under the Future and Emerging Technologies Program.

KI: When did scientists become aware of privacy issues w.r.t. the analysis of mobility data and how has the research field developed since then?

Yücel Saygin: Data privacy in general was of concern for many decades. In fact, OECD has started the work on privacy in the early 1970s. In 1980, the OECD guidelines on the Protection of Privacy and Transborder Flows of Personal Data were adopted as an international consensus on how the personal information should be collected and managed. With data processing becoming an essential part of businesses, privacy issues gained even more importance. In 1980s, statisticians and the database community started to look into this problem, i.e. how to utilize the data for public interest while protecting the privacy of individuals. Statistical disclosure control techniques were developed to reduce the risk of disclosing sensitive information while publishing tabular data. On the other hand, database researchers developed techniques for answering aggregate queries over sensitive data leading to statistical databases. Privacy preserving data management really took off in early 2000s, and privacy preserving data mining and privacy preserving data publishing have become popular research fields. Then, the general research field of privacy preserving data management diffused into more specific application areas, and mobility data was one of them. Mobility data is basically the location data collected through various means, such as GPS, GSM networks, or Wi-Fi networks. There are two aspects: one is real time location data, and applies mostly to location based services, and the other is about historical data, where location is collected over time by service providers and needs to be utilized for decision making. The GeoPKDD project (Geographic Privacy-aware Knowledge Discovery and Delivery) funded by EU FP6 under the Future and Emerging Technologies Program

Address(es) of author(s) should be given

has initiated the mobility data mining research and looked into the problem of harvesting knowledge from mobility data. In the context of the GeoPKDD project, a consortium of 8 partners developed privacy aware techniques for collection, storage and analysis of mobility data. Privacy issues with respect to the analysis of mobility data have not been totally resolved yet, and I believe it is still an interesting research direction.

KI: What are the dangers of exposing location information respectively exposing information about where somebody is not?

Yücel Saygin: Releasing the information that someone was at a particular location or not may have serious consequences in terms of privacy. For that reason, location data has always been treated separately from other types of data. One can easily see that location data can be used to infer a lot about a person including religious beliefs (going regularly to a church or a mosque, or not going to a church or mosque), health information (being at a hospital specialized in cancer treatment), and sexual life (going to particular clubs) which are all considered sensitive information. Location data can actually be used to create an accurate profile of a person in combination with even very simple analysis tools. With sophisticated data mining tools, one can learn or infer a lot about a person by analyzing his/her location data collected over time.

KI: Which approaches exist for the safe handling of mobility information? What are their shortcomings?

Yücel Saygin: There are some privacy preserving publishing methods out there for mobility data, however, we still need to have provable measures that are also practical. By practical, I mean that after you anonymize the data, it must still be useful. Background information always makes the task of anonymization very difficult. In case of location data, this is even more serious since there are heaps of data sources that can be used as background information to recover sensitive information. Utility versus privacy is always a big challenge.

KI: What are the main research questions to be tackled in the next years?

Yücel Saygin: Anonymizing location data considering background information is still a research question. Nowadays geo-social networks are becoming popular such as foursquare. Facebook is also pushing its "Places" service. What happens when location data is combined with the rich multimedia data collected through social networks? This will be a challenge for the privacy preserving data management community. The idea of differential privacy proposed by Cynthia Dwork from Microsoft Research is very important and it needs to be applied to location data publishing.

KI: You are coordinator of the Coordination Action MODAP funded by the European Commission under FP7. What are the goals of MODAP and which achievements has it gained so far?

Yücel Saygin: The goal of MODAP is to create a platform for technical as well as non-technical people who are interested in mobility data mining and the privacy issues in mobility data analysis. We want to enlarge the GeoPKDD community into a more interdisciplinary one involving social scientists and legal experts. We want to create awareness in the society, and disseminate the results of the precursor project GeoPKDD at a wider scale. However, MODAP is not all about privacy but it is about data mining as well. Therefore, we are organizing activities to support privacy preserving data management research. Within the MODAP community we have a body called Privacy Observatory which involves legal people and computer scientists. Within the Privacy Observatory, we have already started to see interdisciplinary research efforts which are promising. We have organized training activities like the first MODAP summer school, and we are having the second one in the summer of 2012. We have organized a conference and various workshops. There are a couple of new research projects initiated by the MODAP community.

KI: From your experiences with the Privacy Observatory, how deeply are national authorities aware of privacy issues w.r.t. mobility data?

Yücel Saygin: I know Dr. Roberto Lattanzi from the Italian Data Protection Authority who created an extremely positive impression on me about the data protection authority in Italy. They really follow the technology and Dr. Roberto Lattanzi contributed a lot to the MODAP community with his experience in the field. I hope that the other data protection authorities are like that. I should also say that Turkey is in the process of establishing a data protection authority, and I hope that we will follow the good example of Italy. The MODAP community is specialized on data mining, and data mining is one of the concerns in terms of privacy especially in the context of profiling. Therefore, I see a lot of opportunities for collaboration with the data protection authorities in Europe and the MODAP community.

KI: How do you teach your students to be aware of privacy risks, e.g. in social networks?

Yücel Saygin: The best way to teach students is by projects where they can do research and understand the matters themselves. At my university, I give research projects to 2nd year students (not necessarily computer science students). For example, last semester, we asked the students to first look into the privacy issues raised by new technologies like social networks or

smart phones. Then we asked them to do a survey questionnaire to have an idea of how much society is aware of this issue. I was also surprised when reading some of the articles they found on the Internet, describing the influence of social networks in hiring, and also on the credibility of some professions. I guess students learn a lot doing such projects. I also give seminars at various universities I visit, and it is nice to see that students are very much interested in this topic though they work on completely different areas.

KI: Thank you very much for the interview!