Newsletter

Vol 1, Issue 6, Sep. - Oct. 2010

In this issue: Short Report of Summer School - Amendment of Turkish Constitution
After Summer School - Upcoming Events

www.modap.org

Editorial

Nuray AKMERIC, Editor in Chief

Welcome to the sixth issue of MODAP Newsletter. With this issue we will be publishing the last issue of Volume 1. Starting from November/2010, our Newsletter will be issued monthly that means we need your contribution much more than the first year of the project.

In this summer, there were two important events related with MODAP; the first one was MODAP Summer School that took place in Rhodes, Greece at the end of August. Our November issue of the Newsletter will be published as "Special Issue of MODAP Summer School", until then please find the short report written by Yannis Theodoridis, Organizing Chair of MODAP Summer School, in the next column.

The second important event was the referendum held in Turkey in September for the Amendment of the Turkish Constitution. With this Amendment, most articles of the Constitution regarding the Privacy of Personal Life have been changed. In this issue, we looked closer to these changes in the article written by Derya BAKSI, Attorney-at-Law (Istanbul Bar Association).

During the MODAP Summer School, MODAP Coordinator Yucel Saygin noticed a very interesting project. I am giving the following paragraph to him for his short introduction related with the matter.

"During the summer school, participants formed groups where each group was asked to do a small project. All the groups worked enthusiastically and presented their work to the participants. One of the project groups reported remarkable results that could be of interest to others, therefore I asked the group representative Alex Kotsifakos to write a short report on their project. I should especially thank Prof. Bert-Jaap Koops (from Tilburg University, Law School) who proposed this project. In this issue Prof. Koops gives a summary of the project and Alex explains his groups work and the results they obtained, I am sure you will enjoy reading this."

Happy reading...

Short Scientific Report on the 1st Summer School on Mobility, Data Mining, and Privacy

Yannis Theodoridis, University Piraeus, Organizing Chair of MODAP Summer School

The school targeted at doctoral students and young postdocs as well as a small number of senior researchers working on database management, machine learning and data mining, GIS, statistics, law informatics, etc. This was the first doctoral school ever on the 'hot' intersection of three domains: modelling and management of moving object databases (Mobility), data analysis and knowledge discovery from mobility data (Data Mining), and privacy aspects that raise when processing human mobility (Privacy). As such, it is expected to open new horizons in the field, and this is ensured by the quality of the invited speakers. Namely, the summer school included talks by Dr. Fosca Giannotti (ISTI/CNR), Prof. Harvey Miller (Utah U.), Drs. Natalia & Gennady Andrienko (Fraunnhofer), Prof. Christian Wietfeld (TU Dortmund), Prof. Dino Pedreschi (U. Pisa), Dr. Nikos Pelekis (U. Piraeus), Prof. Dimitrios Gunopulos (U. Athens), Dr. Roberto Trasarti (ISTI/CNR), Prof. Bert-Jaap Koops (Tilburg U.), Prof. Marta Gonzales (MIT), and Prof. Yucel Saygin (Sabanci U.) under the coordination of the program chair, Prof. Bart Kuijpers (Hasselt U.).

The participants of the summer school were selected after a single-blind evaluation procedure coordinated by a Steering Committee and based on the CVs and motivation letters by the applicants. The final list of participants included 64 researchers from 16 countries (from Europe, Asia, and America). To close this short report, here is the feedback from one of the participants sent to the organizers just after the school: "***I had a fantastic time, and we were all looked after extremely well, the talks were really varied and interesting and I have come away having learned more than I could have managed by myself in a year!***"

Further information about the event appears in http://mss2010.modap.org/.

Upcoming Events / Announcements

MODAP Workshop on "Privacy-aware Analysis of GSM Data"
November 26, 2010, Fraunhofer Center Schloss Birlinghoven, Sankt Augustin, Germany
http://ws-gsm2010.modap.org

New Book! - Ubiquitous Knowledge Discovery
Due on October 11, 2010
http://www.springer.com/computer/ai/book/978-3-642-16391-3

SPRINGL 2010 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS
November 2, 2010, San Jose, CA, USA

Amendment of the Turkish Republic Constitution and Its Impacts over Privacy

Derya BAKSI, Attorney-at-Law (Istanbul Bar Association), Tarlan – Baksı Law Office

In Turkey one of the recent hot topics is the referendum, the reason of which is the 'Amendment of the Turkish Constitution'. As a result of the referendum held in Turkey on September 12th, 2010; the amendments of the Turkish Constitution have been accepted by the public. With these amendments; most articles of the Constitution, including the article 20 regarding the Privacy of Personal Life, have been changed. As per the amendment made to the article 20 of the Constitution, which is effective as a consequence of the referendum, a new paragraph has been added to this article.

According to this additional paragraph; everybody has a right to request the protection of personal data. This right covers the possibility to reach and to get information about personal data, to request the amendment or deletion of this data. Everybody would learn for which purposes these data are being used and regarding this article, personal data can only be used in cases prescribed by laws or with the owner's explicit consent.

There are a lot of discussions about this additional paragraph whereby some people are for this amendment but some people are against it. According to the ones who are for this change, protection and privacy of personal data is a crucial matter and must absolutely take part in the Constitution. To get information, to reach all kinds of personal data is really one of the most important requirements of the democratic system. For citizens, it is a great privilege to learn for which purposes these data are being used or will be used.

On the contrary, others who are against this change claim that in the current Turkish Penal Code, there are provisions about the protection of personal data thus no special provision is required in the Constitution. Furthermore, for almost four years, 'The Draft Law for Protection of Personal Data' has been waiting to be enacted at the sub-committee of The Grand National Assembly of Turkey. With the perspective of legal systematic, this additional paragraph should be regulated in the Codes, but not in the Constitution. It is also pointed out that this additional regulation is not enough, not clear and has many vague sides. For example, there is not an independent institution to supervise the protection of personal data. It is not clear which institution will give the data and which institution will keep the data and who will supervise these institutions. Those are some topics that should be set forth in detail in a special code.

Most of the European countries have their own regulations about protection and privacy of personal data since 1970s. Therefore when it comes to Turkey, it is an important step to have such a regulation about the protection and privacy of personal data in the Constitution. Nevertheless, what has to be done is to clarify the details, regulate the vague parts in a special code and to implement these rules in the daily life. These rights must not only stay as a regulation in the Constitution but also should be assisted with an effective legal system in practice. With the help of a uniform legal approach, the vague sides of this regulation should be clarified in favor of the benefits of the Turkish citizens.

Right after the Summer School

Prof. Bert-Jaap Koops, Tilburg University, Law School

In the summer school, I presented legal and ethical issues of data mining mobility data, particularly focusing on privacy and personal data. To try and make students think about what privacy means, and how broad the legal concept of 'personal data' (i.e., data that can be traced, directly or indirectly, to a unique individual) is, I gave the students advance homework to write down their own location data for three days, and wrote an exercise to analyse these data. It was interesting to see that almost all students, when I asked them to turn in their forms, indicated they felt somewhat uncomfortable to provide the data, even if they were pseudonymized, suggesting that they felt the data were easily identifiable (at least for their fellow students and the teacher…) and potentially embarrassing. That is precisely what privacy is about.

Unfortunately I could not attend the presentation of the exercise, having to take a very early flight the last day, but I was very happy to read a report from the students presenting the results of the exercise. I found it very informative, and hope that it makes researchers studying mobility data aware of how vulnerable people can be when location data are stored, processed, and re-used, particularly in other contexts such as family life and law enforcement. It shows the importance of having privacy and data-protection laws to govern the use of data in the database age.

Project Details

Alex Kotsifakos, Ph.D. Candidate, Dept. of Informatics & Telecommunications, University of Athens

In our project we worked on privacy and identifiability in mobile data. We were given MODAP students' location data of three days and were asked to analyse these data. More specifically, before the summer school all the participants were given a form, in which they should, optionally, provide their location data, i.e. the city, address, and type of activity (e.g. hotel, residence, on road, bar, business) they were doing, for all the hours of the 21st, 24th, and 26th of August 2010. All the participants were identified by a pseudonym, a restaurant name and a number ranging from 0 to 9, e.g. "Rozalia9".

Based on the above data, we firstly had to answer the questions below. We have to note that for question a) we were also in knowledge of information about the participants (name, surname, country, institution, e-mail, junior or senior level, and short bio information).

a) Which persons can you identify
1. if you are a European LBS company?
2. if you are the FBI?
b) Suppose that you don't know the data come from the group of MODAP students; the reference group is the world population. What other data can you use to connect the data to specific individuals?
1. use publicly available information (e.g., surf the Internet to find out more information about the person or location);
2. which non-publicly available information could the European company or the FBI acquire to make identification possible?"

Due to the fact that there is a strong correlation between a) and b), we answered these questions together.

First of all, for the majority of participants we have the exact addresses of their residence and working place. For example, given the data in one specific form, we know that a person X is going regularly from residence Sant Joan de Vilatorrada, Coll Baix 56 to campus UAB and back.

Thus, if we are a European LBS company in order to find who the person X is, we can generally use publicly and/or non-publicly available information. Regarding the first type/source of information, we can use the Internet and rip the list of employees and students from the UAB website and then intersect the surname with the address using the website of "Infobel" until we get a unique hit, independently of whether we have information about the participants or not. As for the second type, two possible options for acquiring more information are the following: 1) send a leaflet to the address to be filled in by the customer, who has to send it back by (e-)mail, and 2) retrieve data from a collaborative third-party company, depending on the contract constraints between the companies.

If we are the FBI, aiming at identifying person X, we can use our local database and intersect it with easily accessible (since we are the FBI) non-publicly available information, such as IP logs, mobile data, people's history, and flight information.

Consequently, whether or not we have the participants' list information, we can use publicly and non-publicly available information and identify, sooner or later, all the individuals.

Next, in the context of our project, we were asked to do the following exercise. Firstly, we had to choose the location data from one or two persons. Secondly, we had to invent a story-line fitting the data which would be embarrassing or potentially harmful to the data subjects. Then, supposing that we are a European LBS company or the FBI, in two separate scenarios of exploiting the data given, an answer to the following question had to be given:

"What are the risks that the persons are actually harmed, when these data are interpreted by this story-line?"

Fitting the location data of two persons, we invented a story-line:

"Rozalia9" and "Portego7" who know each other went to a wedding on Saturday the 21st of August in Portaria (Greece), and "Rozalia9" lied to his wife saying that he would be in Athens.

Regarding the first scenario, being the European LBS company, assume that the provided service would be to inform people for their closest pizza restaurants, we would send "Rozalia9" a message for such a restaurant in Portaria. However, this message may harm the subject in case that he would forget to delete it from his mobile phone and his wife would look at it a few days later (for unsafely and curiosity reasons).

As for the second scenario, being the FBI (and to make matters worse), imagine that after the wedding there was a murder during the party time. We, as FBI, would have to interrogate all the guests. Therefore, we would visit the house of "Rozalia9", suppose that his wife would open the door, and we would say:

"FBI. We would like to ask you and your husband a few questions for the murder in Portaria. Your husband was in a room with "Portego7", close to the murder scene…"

The consequences of such a story would be definitely harmful for all the subjects involved in it.

To conclude, publishing "innocent" information, such as location, type of location, age, sex, may harm people. This is because these kinds of information can be combined with the use of publicly (in our case internet and telephone directory) and non-publicly available sources of information, leading to privacy leak, as the individuals may be easily identified. Finally, due to the fact that re-identification is possible because of massive public information, one must take special care when he or she releases bulk data, focusing on K-anonymity, L-diversity, and T-closeness. These technical-organisational measures will allow controlling the risks of misinterpretation and misuse of the data.