

The 2nd International Conference on Integrated Information

The Design of a Spatio-Temporal Database to investigate on sex offenders

Paolino Di Felice*

*Dipartimento di Ingegneria Industriale, Informazione ed Economia,
Università di L'Aquila, Via Giovanni Gronchi n. 18 Nucleo Ind.le di Pile, Italy*

Abstract

Sex offenders have a high recidivism rate. Fight against recidivism is, therefore, a relevant issue. Recent laws call for the use of the GPS technology to monitor the whereabouts of sex offenders. In the paper, we design a Spatio-temporal database suitable to store the trip of criminals as moving points that is as time dependent geometries. This opens the frontier to a new generation of software applications much more effective than those currently used by several criminal investigation departments all over the world.

© 2013 The Authors. Published by Elsevier Ltd.

Selection and peer-review under responsibility of The 2nd International Conference on Integrated Information

Keywords: sex offender; criminal mobility; moving points; database

1. Introduction

Sex related crimes are a serious social problem because they cause devastating consequences for victims, families, and communities, and because of the high recidivism rates. For instance, a research by the Ministry of Justice of Japan reveals that 30% of repeat offenders were responsible for 60% of the crime committed in Japan from 1948 to 2006, [1]. The prevention of recidivism needs effective strategies to supervise sex offenders. Many jurisdictions have implemented residency restrictions. As quoted in [2], unfortunately “Residency restrictions have had little measurable crime reducing impact. Another measure against sex offenders is the so-called “registration and notification policy” that exists in many countries around the world since a long time. Unfortunately, even this policy did not reduce crime rates tangibly [3].

More recent laws call for the use of the global positioning satellite (GPS) technology to monitor the whereabouts of habitual sex offenders. This is the answer to the widespread social expectations, as it can be read

* Corresponding author: Paolino Di Felice. Tel.: +39-0862-434 418; fax: +39-0862-434 403.

E-mail address: paolino.difelice@univaq.it

in [4] which ends by reporting that “Magistrates and health professionals complain the lack of effective monitoring procedures for released offenders.” The adoption of GPS based monitoring policies, however, raises the critical issue of the violation of the personal privacy, a fundamental human right established in international laws and regulations. A debate about such concern is still on the carpet (e.g., [2]). The central point, of course, concerns *how* to treat the data about the movements of sex offenders.

This article adopts a Database Management System (DBMS) that supports data types and operators for the so-called moving points (in short, m-points) as meant in [5]. M-points are time dependent geometries. Two views on m-point data have been established in the past. The first one focuses on answering questions on the current position of m-points, and on their predicted temporal evolution in the (near) future. This approach is often called *tracking* [6]. A second approach represents (in a single database attribute) complete histories of m-points inside so-called *trajectory databases*, [5]. Of the two approaches, the second is better suited, in our opinion, to be used to study the movements of sex offenders mitigating, at the same time, the issue of privacy invasion even more pronounced in the case of tracking that implements a real-time control of people actions. Choice, the latter, that although decidedly “punitive” offers no certainty that it can produce beneficial effects on the reduction of the recidivism rates among sex offenders. At least, as mentioned, there are no studies that prove it.

The paper is structured as follows. Sec. 2 introduces the reference scenario about sex offenders and their movements. Sec. 3 concerns the design of a spatio-temporal database collecting the data about the problem and its implementation in the SECONDO DBMS [7]. Sec. 4 presents an example dataset used to feed the database. Sec. 5 concludes the paper and sketches the future work.

2. The application context

As mentioned above, the reference scenario of the study summarized in this paper concerns sex offenders whose movements are tracked (hereinafter, therefore, also called *subjects on probation*), sensible areas (schools, parks, public restrooms, train stations, ...) and movements of subjects on probation between sensible areas. The position of the sex offenders is acquired by an electronic device equipped with a GPS detector. Such data are transmitted to a server and filed for a certain number of years (according to the national and international laws in force).

In the paper we use the following notations.

- $\mathcal{C} = \{C_1, C_2, \dots, C_c\}$: the known sexual crimes,
- $\mathcal{P} = \{P_1, P_2, \dots, P_p\}$: the pending sexual crimes,
- $\mathcal{S} = \{S_1, S_2, \dots, S_s\}$: the subjects on probation,
- $\mathcal{A} = \{A_1, A_2, \dots, A_a\}$: the sensible areas,
- $\mathcal{T} = \{T_1, T_2, \dots, T_t\}$: the movements.

In turn, $T_i = \{\langle P, t \rangle \mid P \text{ is a point described by a pair of coordinates } \langle \text{lat}, \text{long} \rangle \text{ and } t \text{ is the time stamp of the acquisition of } P\}$, where $i=1, 2, \dots, |\mathcal{T}|$. The elements of T_i are temporally ordered. T_i expresses the trip of a subject in \mathcal{S} between two sensible areas in \mathcal{A} .

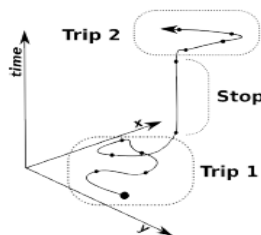


Figure 1. The whole GPS trace of a subject on probation made up of intermediary trips

An explicit comment can explain how it is possible to have data about the movements of the subjects on probation from a sensible area to another one. The premise is that the acquisition of the position of the generic subject is h24, for all the days of the year. The daily mobility of each subject can be summarized by a set of “single” trips that he/she performed during the day (Fig. 1). Therefore, this continuous stream of information contains different trips made by the subject. In order to distinguish between them, we need to detect when he/she stopped for a while in a place. This point in the stream will correspond to the end of a trip and the beginning of the next one.

In this way splitting the “whole history of the movements” of a subject into single trips becomes trivial. It arises, therefore, the need to process the movements of a subject in order to cut them into “elementary” units. For the purposes of this study, we assume that the single trip goes from the exit of a sensible area to the entrance into another one. Commonly, the stop points are identified knowing that they keep the same spatial position for a certain amount of time quantified by a predefined temporal threshold (e.g., [8]). In our study, vice versa, it will be appropriate to adopt a more rough schematization consisting in ignoring the movements that a sex offender performed inside a sensible area. Therefore, all the time spent within the area is as if he/she had been stopped.

Another specificity of the reference scenario of our study concerns the fact that the sex offender did not necessarily enter into the sensible areas. Consider, for example, a school with a fence and locked gate during the hours of teaching. Likely, he/she preferred get close enough to the building and remain under observation in order to spy on potential victims without the need to enter and without risking being seen. To deal with this aspect, it is sufficient to apply a “buffer” to the polygon that defines the geometry of the building. If the location of the sex offender falls within the buffer, to the purpose of our study it is as if he/she was within the school.

3. The solution

Aim of this section is the design of a spatio-temporal database suitable to collect the data of the application context of Sec. 2.

3.1. The conceptual schema of the database

Fig. 2 shows a possible conceptual schema of the database, composed of 5 entities and 4 relationships ([9] is a good textbook about databases conceptual design). The schema originates from the design choice of modeling the subjects’ trips as *m-points*. Each trip links two sensible areas.

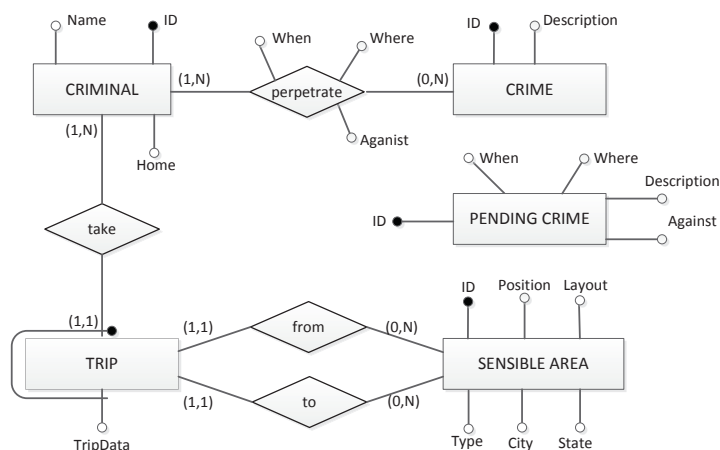


Figure 2. The conceptual schema of the database

Few words have to be spent about unsolved crimes (entity PendingCrime). For them we know the place and the time of the event, but we ignore the guilty. Therefore, such an entity has to be kept unrelated from the entity Criminal. Later on, we will return on this issue.

3.2. Mapping the conceptual schema into the logical one

The relational mapping of the conceptual schema of Fig.2 is affected by the DBMS generation one decides to refer to. At present it exists a twofold alternative: either adopt a relational DBMS which features a spatial extension (e.g.: PostgreSQL/PostGIS) (*Alternative 1*) or employ one featuring a spatio-temporal extension (e.g.: SECONDO) (*Alternative 2*). We refer to the second alternative because it offers many advantages, as discussed briefly below.

Alternative 1 vs. Alternative 2

The comparison will be focused on the:

- E enabling technologies,
- E elegance of the implementation,
- E feasibility of spatio-temporal analysis, and
- E performances.

Enabling technologies About Alternative 1, there is a large choice of a suitable DBMS (e.g.: IBM DB2-SE, Oracle/Spatial, PostgreSQL/PostGIS, ...) vice versa, in connection with Alternative 2, the choice is limited to SECONDO (still under development).

Elegance of the implementation In Alternative 1 the route followed by a subject to go from an area to another one has to be split in two parts (i.e., in two “attributes” in relational terms): the geometric component and the temporal one, that because of the lack of a suitable built-in data type in the DBMS. In detail, the trajectory geometry can be modeled as a line string (with linear interpolation between points), while the time stamps of the sampling points have to be collected in an array. In Alternative 2, instead, “the history” of each movement is encapsulated in a single attribute.

Spatio-temporal analysis Carry out a spatio-temporal analysis in the “conceptual framework” defined by Alternative 1 is critical because of the lack of specific operators at the DBMS level. To enhance the expressive power of the SQL of those DBMSs we are forced to implement ad hoc operators. The HERMES project, aiming to extend the Oracle DBMS, is an excellent example, [10]. Similar motivations underline work [11] where authors report about two operators (implemented as user defined functions on top of PostgreSQL/PostGIS) suitable to compute the spatio-temporal intersection of pairs of uncertain m-points’ trajectories.

This difficulty disappears working within the conceptual framework defined by Alternative 2. In fact, SECONDO features a rich bag of spatio-temporal operators suitable to meet the needs of spatio-temporal analysis raised by the application context we refer to in this paper.

Performances As with most of the problems about m-points, the challenge is the efficient retrieval of the information from amongst millions of data points. The solution is indexing. In the current version of SECONDO, the trajectories of m-points can be indexed by a spatio-temporal (2D + time) R-tree that guarantees good performances even for very complex operations (e.g., [12]).

The mapping of the conceptual schema of Fig. 2 produces the following SECONDO database:

```
let crime = [const rel(tuple ([IDCrime: string, Description: string ])) value ()]
let pendingCrime = [const rel(tuple ([IDPC: string, When: instant, Where: point,
    Description: string, Against: string])) value ()]
```

```

let criminal = [const rel(tuple ([IDCriminal: string, Name: string, Home: point]))
  value ()]
let sensibleArea = [const rel(tuple ([IDSA: string, Type: string, City: string,
  State: string, Position: point, Layout: region ])) value ()]
trip (IDCriminal: string, TripData: mpoint, From: string, To: string)
let perpetrate = [const rel(tuple ([IDCriminal: string, IDCcrime: string,
  When: instant, Where: point, Against: string ])) value ()]

```

Tables above will store the values from the sets \mathcal{C} , \mathcal{P} , \mathcal{S} , \mathcal{A} , and \mathcal{T} , in sequence.

The **pendingCrime** table requires an additional consideration that complements those already made in Sec. 3.1. Such a table contains a tuple for each unsolved crime. During the life of the database, each time a crime is attributed to a sex offender, then an ad hoc script that migrates the corresponding values of the attribute **Where**, **When**, and **Against** (from **pendingCrime**) into a tuple of the **perpetrate** table has to be executed; tuple to be completed with the values of the attributes: **IDCriminal** and **IDCrime**. Immediately afterwards it will be necessary to delete from **pendingCrime** the tuple migrated into **perpetrate**.

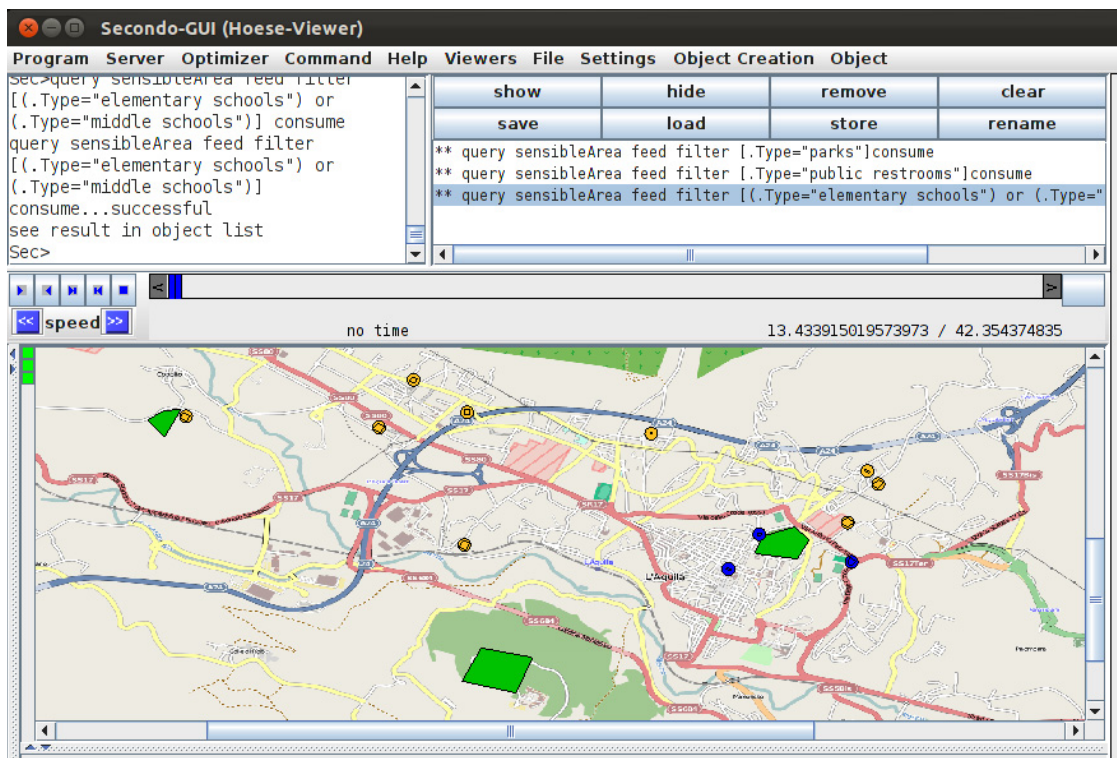


Figure 3. The map of the sensible areas in the dataset shown by SECONDO. The output is generated by the three queries shown in the top-right window of the figure, according to the selected visual style for the returned spatial objects: green areas represent parks, yellow (blue) circles represent schools (public restrooms).

Remarks about the query syntax.

The keyword **query** starts queries written in the SECONDO language. The **feed** operation reads a relation (e.g., **sensibleArea**) from disk and puts its tuples into a stream as they are. The **consume** operation collects a tuple stream into a persistent relation suitable to be kept and indexed. The **filter** operation is equivalent to the **WHERE** clause of SQL as can be understood by looking at the query below that reformulates the first one above: **SELECT * FROM sensibleArea WHERE Type = "parks"**.

An explicit remark about the attributes **from** and **to** of table **trip** is required, because there are reasons to cancel them and others to keep them. In case the scenario we refer to in this study would be implemented, then the information about the initial and final point of each trip would be possible to be inferred by calculating the intersection between the m-point and the geometry of the sensible areas (hence attributes **from** and **to** could be canceled). Vice versa, working in the abstract setting of Sec. 2 the movements between pairs of sensible areas of each subject on probation are generated for our purposes, so it is quite reasonable to keep track of them explicitly in the **trip** table. There is, however, a further motivation that suggests keeping attributes **from** and **to** even in the real scenario and it is related to performances. In fact, all the times we will need to know the extreme areas of a trip, it will be sufficient to carry out a query that makes a look-up of the **trip** table, vice versa those data have to be computed “from scratch”.

4. An example dataset

To make preliminary experiments, we refer to a synthetic dataset where the movements of sex offenders have been generated by a Java program which receives as input the sensible area of departure and arrival, the date and start time of the trip and returns a text file that describes the journey. The example dataset is small, but significant enough for the purposes of the study. It consists of 20 subjects and 20 sensible areas (located in the town of L’Aquila and nearby, centre of Italy – Fig.3, and composed of: 7 elementary schools, 6 middle schools, 4 public restrooms, and 3 parks).

The database on which the tests were conducted also contains the data of 30 trajectories corresponding to as many trips of the 20 subjects on probation between pairs of sensible areas. These movements are shown by the graph of Fig. 4, where each node is labeled with the code of a sensible area, while the arcs are labeled with the code of the offender that moved between the extreme nodes.

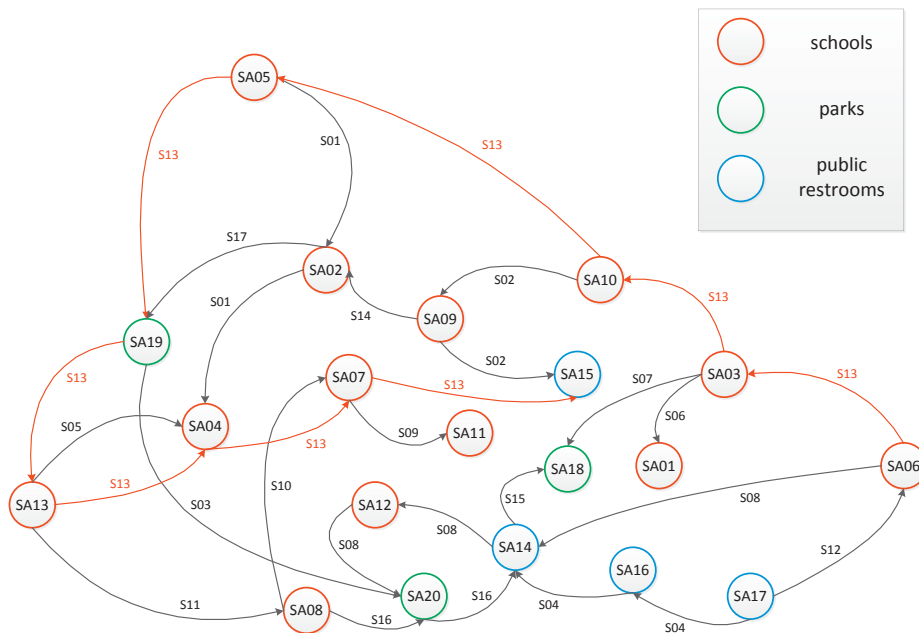


Figure 4. The graph of the movements of the 20 subjects on probation being part of the example dataset

To learn how to load tracking data collected by a GPS device into an attribute (**tripData** in our case) of a relation (**trip**) as a **mpoint** data type, the reader interested may refer to [13].

5. Conclusions and future work

A spatio-temporal database collecting data about the movements of sex offenders has been designed. The article proposed alternative solutions and compared their strengths and weaknesses. Our conclusions can be summarized as follows: at the conceptual level, the best organization of the database can be achieved by modeling the subjects' trips as m-points, while, at the implementation level, the SECONDO DBMS is today the most mature enabling technology.

When a database like the one designed in this paper is put into operation, the next step to focus on concerns the definition and the implementation of suitable investigative strategies that emphasize the strengths of such an innovative technological solution. In essence, the issue is to take advantage of the availability of the data about the movements of sex offenders, a kind of data that goes beyond the capabilities of current DBMSs. A preliminary step in such a direction may be found in [14].

References

- [1] Someda, K. (2009). An international comparative overview on the rehabilitation of offenders and effective measures for the prevention of recidivism. *Legal Medicine* 11, 82–85.
- [2] Payne, B. K., DeMichele, M. (2011). Sex offender policies: Considering unanticipated consequences of GPS sex offender monitoring. *Aggression and Violent Behavior*, 16, 177–187.
- [3] Lieb, R., Kemshall, H. T. (2011). Post-release controls for sex offenders in the U.S. and UK. *International Journal of Law and Psychiatry* 34, 226–232.
- [4] Kaplan, L. V. (2008). Sexual psychopathy, public policy, and the liberal state. *International Journal of Law and Psychiatry* 31, 172–188.
- [5] Güting, R.H. and Schneider, M. (2005). *Moving Objects Databases*. Morgan Kaufmann Publishers.
- [6] Sistla, A.P., Wolfson, O., Chamberlain, S., and Dao, S. (1997). Modeling and Querying Moving Objects. Proc. 13th Int. Conf. on Data Engineering (ICDE, Birmingham, U.K.), 422-432.
- [7] Güting R.H., Behr, T. and Düntgen, C. (2010). SECONDO: A platform for moving objects database research and for publishing and integrating research implementations. *IEEE Data Engineering Bulletin* 33:2, 56-63.
- [8] Trasarti, R., Pinelli, F., Nanni, M., and Giannotti, F. (2012). Individual mobility profiles: methods and application on vehicle sharing. 20th Italian Symposium on Advanced Databases Systems.
- [9] Elmasri, R., Navathe, S. B. (2011). *Fundamentals of Database Systems*, 6th Ed., Addison Wesley - ISBN-13: 978-0-136-08620-8.
- [10] Pelekis N., Frentzos, E., Giatrakos, N., Theodoridis, Y. (2012). "HERMES: A trajectory database engine for mobility-centric applications", *International Journal of Knowledge-based Organizations*, 2012 (in press).
- [11] Di Felice P., Orsini, M. (2011). Spatio-temporal intersection of trajectories under uncertainties. 19th Italian Symposium on Advanced Database Systems. 357-368.
- [12] Güting R.H., Behr, T. and Xu, J. (2010). Efficient k-nearest neighbor search on moving object trajectories. *The VLDB Journal* 19:5, 687-714.
- [13] Güting R.H. (2011). Example: Loading Tracks and Displaying Them With a Map Background. (<http://dna.fernuni-hagen.de/Secondo.html/files/SecondoTracks.pdf>, accessed on July 26-th, 2012).
- [14] Di Felice, P. (2012) Investigative strategies on top of a spatio-temporal database about sex offenders. Proc. 2nd Inter. Conf. on Integrated Information.