Inference attacks through mobility Markov chains

Sébastien Gambs
IRISA/INRIA - Université de Rennes 1
Campus Universitaire de Beaulieu, 35042 Rennes, France
Email: sgambs@irisa.fr

Marc-Olivier Killijian and Miguel Núñez del Prado Cortez
CNRS ; LAAS ;
7 avenue du Colonel Roche, F-31077 Toulouse, France
Université de Toulouse ; UPS, INSA, INP, ISAE ; LAAS ;
F-31077 Toulouse, France
Email: Marco.Killijian@laas.fr and mnpc@computer.org

***Abstract*—In this paper, we adress the modelization by an adversary of an individual mobility. The mobility model learnt from the individual record of position is represented as a Markov chain. We describe in the paper our techniques for building such mobility Markov chains and report on some case studies. These cases studies are based on real user's data we captured. They are analyzed to illustrate what can be learnt from these mobility models.**

I. Inference Attacks on Geolocated Data

An *inference attack* is an algorithm that takes as input some geolocated data D, possibly with some auxiliary information aux, and produces as output some additional knowledge. For example, an inference attack may consist in identifying the house or the place of work of an individual. The auxiliary information reflects any *a priori* knowledge that the adversary might have gathered (for instance through previous attacks and by accessing some public data source) and which may help him in conducting an inference attack. We propose to classify the inference attacks according to (at least) three dimensions such as the type of data it works on, the objective of the attack as well as the specific technique used.

Geolocated data. Nowadays, the rapid growth and development of geolocated applications has multiplied the potential sources of geolocated data. The geolocated data generated by these diverse applications varies in its exact form and content but it also shares some common characteristics. A *mobility trace* is characterized by: an identifier, a spatial coordinate, a timestamp and potentially additional information (speed, direction, etc.).

A *geolocated dataset* D is a dataset which contains mobility traces of individuals. Technically, this data may have been collected either by recording locally the movements of each geolocated system for a certain period of time, or centrally by a server which can track the location of these systems in real-time. A *trail of traces* is a collection of mobility traces that corresponds to the movements of an individual over some period of time. A geolocated dataset D is generally constituted by an ensemble of trails of traces from different individuals. The Crawdad project[1] is an example of a public repository giving access to geolocated datasets, which can be used for research purpose.

Objective of the attack. An adversary attacking some geolocated data may have various objectives ranging from identifying the home of the target to reconstructing his social network, through obtaining knowledge of his favourite jogging tracks. More precisely, the objective of an inference attack may be to:

- *Identify important places*, called *Points Of Interests* (POIs), which characterize the interests of an individual [2].
- *Predict the movement patterns of an individual* such as his past, present and future locations. According to some recent work [3], [4], our movements are easily predictable by nature. For instance, the authors of these papers have estimated to 93% the chance of correctly guessing the future location of a given individual after some training period on his mobility patterns.
- *Learn the semantics of the mobility behaviour of an individual* from the knowledge of his POIs and movement patterns. For instance, some mobility models such as *semantic trajectories* [5], [6] do not only represent the evolution of the movements of an individual over time but also attach a semantic label to the places visited.
- *Link the records of the same individual*, which can be contained in different geolocated datasets or in the same dataset, either anonymized or under different pseudonyms.
- *Discover social relations* between individuals by considering for instance that two individuals that are in contact during a non-negligible amount of time share some kind of social link (of course false positive may happen) [9].

In this paper, we consider that the adversary possesses a geolocated dataset of some previous mobility traces from its target. His objectives are first to establish the POIs for this individual, and second, to model the mobility between these POIs such as to be able to predict its current or future positions. It is worth noting that for space reasons we cannot review the state of the art but that there exist some related work, especially [3], [**?**], and [**?**].

II. Mobility Markov Chain

In this section, we introduce a novel form of mobility model that we coin as *mobility Markov chain* that can represent in a compact yet relatively precise way the mobility behaviour of an individual. Basically, a mobility Markov chain is a probabilistic automaton in which states represent POIs and transitions between states corresponds to a movement from one

[1] http://crawdad.cs.dartmouth.edu/

POI (i.e. state) to another POI. The automaton is probabilistic in the sense that a transition between one POI to another is not deterministic but rather that there are a probability distribution over the transitions leaving from the current POI representing the probability. We will describe in Section III an algorithm that can learn the mobility Markov chain of an individual from his trail of traces.

More formally, a mobility Markov chain is a transition system composed of:

- *A set of states* $P = \{p_1, \ldots, p_n\}$, in which each state p_i corresponds to a POI (or a set of POIs). These POIs may have been learned for instance by running a clustering algorithm on the trail of mobility traces from an individual or simply by collecting the locations that he has posted on a geolocated social network such as Foursquare or Gowalla. Each state (i.e. POI) is therefore associated with a physical location. In the mobility Markov chains considered in this paper, it will often happen that p_1 is the "home" of this individual and p_2 is his"work". Therefore, it is often possible to attach a *semantic label* to the states of the mobility Markov chain. The states are ordered by decreasing importance of the POIs they embody and the last state is often made of what we call the "unfrequent POIs", which are POIs that have been visited several times by an individual but not on a frequent basis.
- *A set of transitions*, $T = \{t_{1,1}, \ldots, t_{n,n}\}$, where each transition $t_{i,j}$ represents a movement from the state p_i to the state p_j. Each transition $t_{i,j}$ has a probability assigned to it that corresponds to the probability of moving from state p_i to state p_j. Sometimes an individual can move from one POI, go somewhere else (but not one of his usual POIs) and come back later to the same POI. For example, an individual might leave his house to go wash his car in a facility near his home and come back 30 minutes later. This type of behaviour is materialized in the mobility Markov chain by a transition from one state to itself. For instance, according to the previous example if p_1 is the home of the individual then the transition $t_{1,1}$ would be assigned a non-null probability. The sum of the probabilities of the transitions leaving one state is equal to one, meaning that $\sum_{j=1}^{n} t_{i,j} = 1$. Note that the probability of going from state p_i to state p_j (for $i \neq j$) is generally different from the probability of going from stage p_j to state p_i (e.g. the probability of going from "home" to "work" is not symmetric with the probability of going from "work" to "home"), therefore in general $t_{i,j} \neq t_{j,i}$. Once the states of the mobility Markov chain have been learnt from a trail of traces, the transition probabilities can be easily estimated by simply counting for each state the number of movements leaving to each other state and then dividing by the total number of movements leaving from this state.

The mobility Markov chain can be represented either as a *transition matrix* or as a *directed graph* in which nodes correspond to states and there is a directed weighted edge between two nodes if and only if the transition probability between these two nodes is non-null.

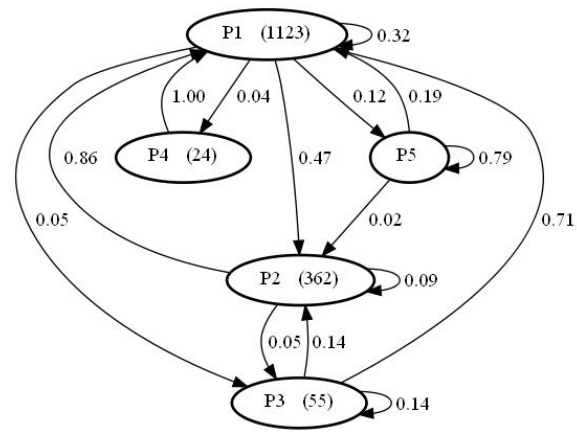

Fig. 1. Mobility Markov chain from user 1.

For instance, consider for illustration purpose, an individual, that we refer thereafter as "user 1", who has a set of 4 important POIs that he visits often (extracted by a clustering algorithm) plus some other POIs that are less important to him. Therefore, we could define a mobility Markov chain composed of 5 states, one for each important POI plus a last one that will contain all the unfrequent POIs. Thus, we have $P = \{p_1, p_2, p_3, p_4, p_5\}$. Suppose now that we have been able to learn the following mobility Markov chain (Figure 1) for this individual from his trail of traces. As additional information, in this Markov chain each state also has a weight associated to it in the form of an integer. We will see how to compute explicitly this weight in Section III but let suppose for now that this weight is given and is related (but not directly proportional) to the time spent by an individual in this state. As such the weight of a state gives an indication of the importance of a state and the states are ordered in decreasing importance of their weights, except the last state that is composed of all the unfrequent POIs and his weight is not considered as meaningful.

Simply by looking at the structure of the Markov chain, it is easy to realize that the state p_1 is the only one that can be reached from all states. Moreover, this happens often with a relatively high probability (except from state p_5). If we combine these observations with the high weight of state p_1, then we can infer with a relatively good confidence that p_1 might be the "home" of user 1. Afterwards, if we want to identify the place of "work" of user 1 and considering as a rule of thumb that "people often go from home to work and vice-versa", we can infer that state p_2 is a good candidate to be the "work" (which is also corroborate by the high weight of this state). Regarding state p_3, we can see that it is either reached from home or work and that a transition leaving this state is likely to take user 1 back to home. Applying now the rule of thumb that "at the end of the afternoon people often go to sport after work before coming back to home", we can infer that state p_3 is a place where user 1 plays sport on a

regular basis. Finally, state p_4 can only be reached from home and can only lead back to home, therefore it makes a good candidate for an activity done on a regular basis during the week-end such as leisure or shopping to a nearby supermarket for instance. To summarize, we can attach the semantic label "home" to p_1, "work" to p_2, "sport" to p_3 and "leisure" to p_4 and by default "unfrequent POIs" to p_5. Taking into account these labels and disregarding the weight of the states, we can now equivalently represent the mobility Markov chain of user 1 as the following transition matrix.

	Home	Work	Sport	Leisure	Unf. POIs
Home	0.321	0.469	0.049	0.037	0.124
Work	0.86	0.093	0.047	0.0	0.0
Sport	0.714	0.143	0.143	0.0	0.0
Leisure	1.0	0.0	0.0	0.0	0.0
Unf. POIs	0.2	0.02	0.0	0.0	0.78

Moreover, we can also take the abstract structure of the mobility Markov chain and put it on a real map (disregarding the "Unfrequent POIs" state), which gives the following result (Figure 2). The mobility Markov chain is a data structure representing the mobility behaviour in a compact yet accurate manner and as such it can be used to perform several inference attacks. For instance, the states themselves directly represent the most significant POIs of an individual and therefore they can be used to derive information about his center of interests. Moreover, if the adversary knows the current position of the individual and if this position corresponds to a state of the Markov chain, he can predict the next movement of the individual by randomizing over the transition probabilities leaving from the current state. The same kind of reasoning can be used to predict the past locations visited by an individual or even guess his actual position. If a semantic label can be attached to some states of the mobility Markov chain (obtained for instance with the help of a reverse geocoding tool), then the mobility behaviour can be analyzed in a much deeper way. Imagine that we have been able to learn the mobility Markov chain of an individual and that we know that this individual is also contained inside a geolocated dataset that has been pseudonymized. The pseudonymity of the individual can be lift fairly easily by finding inside the dataset the individual whose mobility Markov chain is the most similar to the one learnt previously. Of course, this either requires to compute a metric measuring how two different mobility Markov chains are similar to each other or to evaluate the likelihood that a specific trail of traces is compatible with the Markov chain. It is even possible to imagine to use the Markov chain as a generative model for synthetizing artificial data of trail of traces.

III. Learning Mobility Markov Chains

In this section, we describe an algorithm for learning mobility Markov chains and we report on experimentations performed on real users' data.

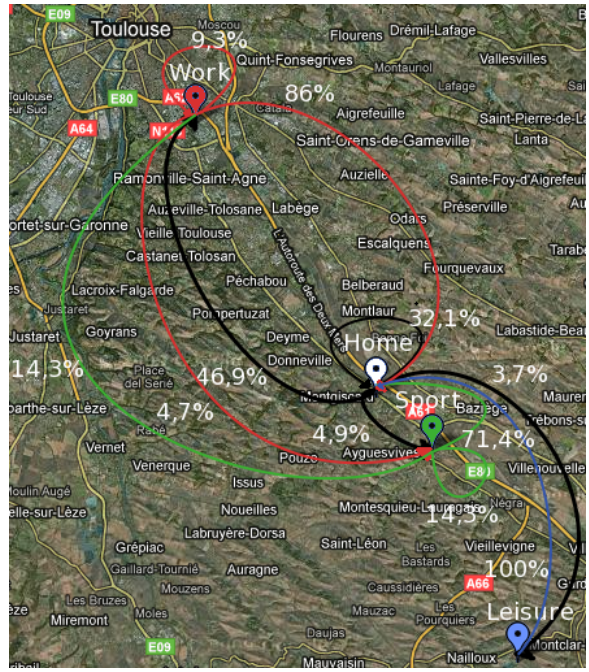

Fig. 2. Mobility Markov chain of user 1 displayed on a real map.

A. Description of the Learning Algorithm

At a high level, the algorithm starts by applying a clustering algorithm on a trail of traces of an individual in order to identify clusters of locations that are significant. Then, in order to reduce the number of resulting clusters, the algorithm merges clusters whose medoids are within a predefined distance d of each other. This merging is not performed in an agglomerative manner but rather a first pass is make on the clusters to determine which clusters are within d-distance from each other and then they are merged within a single global step. Each resulting cluster can be considered as a POI, for instance by taking the centroid or the medoid of the cluster to be the physical location of this POI. For each cluster, we compute the number of mobility traces inside the cluster (which we call the *density* of the cluster) and the *time interval* (measured in days) between the earliest and the latest mobility traces of the cluster. The POIs (i.e. clusters) are then split into two categories; the *frequent POIs* that correspond to POIs whose time interval is above or equal to a certain threshold $mintime$ and the *unfrequent POIs* whose time interval is below this threshold $mintime$. In the set of frequent POIs, we sort the POIs by decreasing order according to their densities. Therefore, the first POI will be the denser and the last POI the less dense.

Now, we can start to build the mobility Markov chain by creating a state for each POI within the set of frequent POIs and also a last state representing the set of unfrequent POIs. As evoked in Section II, each state is then assigned a weight that we set to its density. Afterwards, we come back to the trail of traces that have been used to learn the POIs and we remove all the moving points (whose speed is above ϵ, for ϵ a small value). Then, we traverse the trail of traces in a chronological

order labelling each of the mobility traces either with the tag of closest state (POI) of the mobility Markov chain or with the tag "unknown" if the mobility trace is not within d-distance of one of the frequent or unfrequent POIs. From this labelling, we can extract sequences of locations that have been visited by the individual in which all the successive mobility traces sharing the same label are merged into a single occurence. For example, a typical day could be summarized as the following sequence "$p_1(home) \Rightarrow p_2(work) \Rightarrow p_3(sport) \Rightarrow$ "$unknown''$ $\Rightarrow p_1(home)$", which is quite similar in spirit to the concept of semantic trajectory [5], [6]. From the collection of sequences extracted, we can estimate the transition probabilities between the different states of the mobility Markov chain by counting the number of transitions between each pair of states and then normalizing these probabilities. If we observe a subsequence in the form of "$p_i \Rightarrow$ "$unknown'' \Rightarrow p_i$" then we increment the count from the state p_i to itself (which translates in the graph representation by a self-arrow). See Algorithm 1 for a bried description of this method.

B. Semantic Analysis of Mobility Behaviours

In principle, any clustering algorithm might be a valid candidate for building the initial clusters during the first part of the algorithm but in practice we have observed that out of the 3 clustering algorithms mentioned in Section **??**, DJ cluster was the one that lead to the most meaningful results. In the rest of the section, we report on experiments conducted on mobility data gather through the Phonetic project [25]. The aim of this project is to build realistic mobility models out of real data as well as to study the privacy risks associated with this type of data. Therefore, the goals of this project are closely related to the ones of GEPETO. In this project, Nokia 5800 smartphones have been distributed to registered participants. These smartphones are equipped with a a GPS chip, an accelerometer, a compass, a WiFi and a bluetooth interface. The Phonetic software installed on the smartphones measures every minute the GPS position of the owner of the smartphone as well as the bluetooth neighbourhood. Actually, the example used in Section II to illustrate the concept of mobility Markov chain was learnt from the mobility data collected from one of the user of Phonetic. In the rest of this section, we discuss the mobility Markov chains learnt from three other users of Phonetic that we refer thereafter as "user 2", "user 3" and "user 4". Contrary to the previous quantitative experiments conducted on the mobility traces of the taxi cabs of San Francisco, the following evaluation has a more "qualitative flavour" in the sense that we really focus on the study of the mobility of a few individuals through their mobility Markov chains. The mobility Markov chains have been learnt with a merging distance d of 200 meters and a value of the time interval threshold $mintime$ of 25 days. Previously, we have also tried to learn the mobility Markov chains of taxi drivers but as we expect the Markov chains we obtained were quite complex and difficult to interpret, with a big number of states mainly corresponding to hotspots in the city of San Francisco frequently visited by tourists.

Algorithm 1 Mobility Markov chain learning algorithm

Require: Trail of (mobility) traces M, merging distance d, speed threshold ϵ, time interval threshold $mintime$

Run a clustering algorithm on M to learn the most significant clusters
Merge all the clusters that are within d distance of each other
Let $listPOIs$ be the list of all remaining clusters
for each cluster C in $listPOIs$ **do**
  Compute the $time\_interval$ and the $density$ of C
end for
for each cluster C in $listPOIs$ **do**
  if $C.time\_interval > mintime$ **then**
    Add C to $listFrequentPOIs$ (the list of frequent POIs)
  else
    Add C to $listUnfrequentPOIs$ (the list of unfrequent POIs)
  end if
end for
Sort the clusters in $listFrequentPOIs$ by decreasing order according to their densities
for each cluster C_i in $listFrequentPOIs$ (for $1 \leq i \leq n-1$) **do**
  Create a state p_i in the mobility Markov chain
end for
Create a state p_n representing all the clusters within $listUnfrequentPOIs$
Let M' be the trail of traces obtained from M by removing all the traces whose speed is above ϵ
for each mobility trace in M' **do**
  if the distance between the trace and the state p_i is less than d and the state p_i is the closest state **then**
    labelled the trace with "p_i"
  else
    labelled the state with the value "unknown"
  end if
end for
Squash all the successive mobility traces sharing the same label into a single occurence
Compute all the transition probabilities between each pair of states of the Markov chain
return the mobility Markov chain computed

Figure 3 shows the mobility Markov chain learnt from the trail of traces of user 2. Contrary to user 1, the number of frequent POIs is much higher, thus indicating potentially a more complex mobility behaviour. However, it remains fairly easy to identify the home (state p_1) as the POI that has the biggest number of arrows pointing to it. The work (state p_2) can also be inferred straightforwardly by looking at the transition leaving from the home that has the heaviest weight. Some states such as p_3, p_4, p_5, p_6 and p_7 are more difficult to interpret although they both have a high density. However,

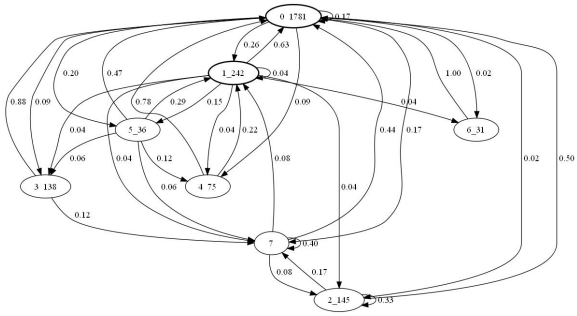

Fig. 3. Mobility Markov chain from user 2.

it is still possible to use a reverse geocoding tool such as Googlemaps to find the name of the closest physical address associated with the coordinates of this states. For state p_3, we obtain the address of a house in a small village approximately 150 kilometers from the home of user 2, which could be indicative of the home of some relative. The confidence in this guess could be strenghten if state p_3 is mainly visited during the week-end or holidays periods. For state p_4, the physical address corresponds to a plaza in the middle of the city in which user 2 lives, which could be for instance a frequent rendezvous where user 2 regularly meets with his friends. States p_5 and p_7 are located inside the university and can be accessed from home or work, which could be an indication that user 2 is either a student or a professor. Finally, state p_6 corresponds to the entrance of a park in a residential area close to which there are a few shops and schools and therefore is more ambiguous and difficult to interpret.

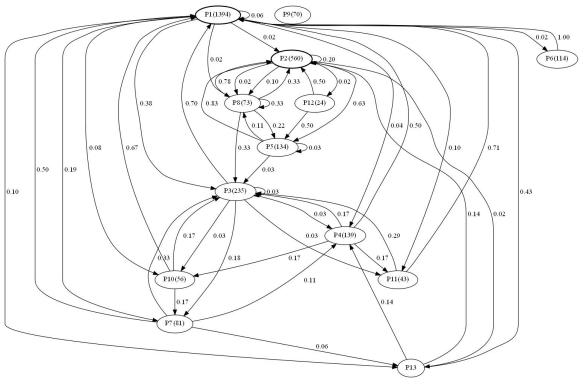

Fig. 4. Mobility Markov chain from user 3.

As shown by Figure 4, user 3 seems *a priori* to have a very complex mobility behaviour. We can start the analysis of this mobility Markov chain by labelling the state p_1 as "home" but then we are faced with the dilemma that state p_2 does not seem to be a valid candidate for "work" as the transition probability from state p_1 to p_2 is very low. Rather, it seems that state p_3 is a more likely candidate although, it is less dense that p_2, as the transition probability from p_1 to p_3 is greater than the transition probability from p_1 to p_2. To clarify this situation, we have use the reverse geocoding tool and observe that actually the states p_1 and p_2 are located in two different countries. Therefore, instead of being considered as "home" and "work", they should be labelled as "home from country 1" and "home from country 2". Taking this new knowledge into account, we can label as "work from country 1" and "work from country 2", respectively the states p_3 and p_5 as they can be reached by the heaviest transitions leaving from states p_1 and p_2. We can now separate the states depending on whether or not they correspond to POIs of "country 1" or "country 2". Of course, this can be done straightforwardly with the help of a reverse geocoding tool but instead this could be learn also directly from the structure of the mobility Markov chain. For instance, if we start a random walk from state p_1 (home of country 1) for a few steps then we are likely to end up in one of the following states: p_3, p_4, p_6, p_7, p_{10} or p_{11}. On the other hand, if we were to begin the random walk on state p_2, after walking for a few steps we have a high probability of ending in state p_5, p_9 or p_{12}. This means that from the structure of the graph we could potentially infer the existence of two highly connected components, one for "country 1" composed of states p_1, p_3, p_4, p_6, p_7, p_{10} and p_{11} and the other for "country 2" composed of states p_2, p_5, p_9 and p_{12}. State p_8 seems to be in none of the two components and indeed a query to the reverse geocoding tool reveals it to be a house in a small village located in "country 2" but quite far from the "home of country 2". As for user 2, this may be the home of a relative or close friend of user 3 that he visits either before or after going to "home of country 1" or "home of country 2". By using the reverse geocoding tool and combining it with the results of the mobility analysis, we can find a name for each state of the mobility Markov chain of user 3 (except state p_6) as illustrated by the following table.

State	Label	Country
p_1	Home of country A	A
p_3	Work of country A	A
p_4	Sport	A
p_6	???	A
p_7	Parking	A
p_{10}	Restaurant	A
p_{11}	Nightclub	A
p_2	Home of country B	B
p_5	Work of country B	B
p_9	Shopping mall	B
p_{12}	Plaza in city center	B
p_8	House of relative or close friend	B
p_{13}	Unf. POIs	

Finally, we finish by analysing the mobility Markov chain of user 4 (Figure 5). From the weight of the different states of the Markov chain, it is easy to see that user 4 is an individual that has contributed so far very seldomly to the Phonetic project and thus displays a very simplified mobility behaviour. Therefore, there is not much to be inferred from the his Markov chain except that state p_1 is likely to be his "home" and state p_2 should be his "work". If we query the reverse geocoding tool with the coordinates of p_3, we obtain a

street in the town center around which there is a high number of restaurants.

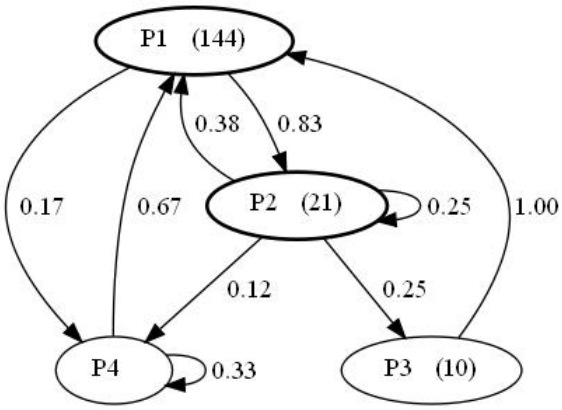

Fig. 5. Mobility Markov chain from user 4.

We have also tested conducted some preliminary experiments for evaluating the behaviour of the mobility Markov chain under some sanitization procedures such as perturbation and downsampling. Basically, we have observed that the mobility Markov chain is relatively robust and that even under significant perturbation or a low rate of sampling, it is still possible to identify the home and the place of work of an individual simply by looking at the structure of the chain. However, it can happen that the less dense states (i.e. POIs) are not preserved under high perturbation and also that the transition probabilities are slightly different from the original ones. This is especially true in the situation in which some states are not preserved and the probability mass of the transitions pointing to/leaving from them is redistributed over the other existing transitions.

IV. Conclusion

In this paper, we have seen that the mobility Markov chain is a highly compact yet relatively precise representation of the mobility behaviour of an individual. By analyzing the structure of the Markov chain (and this even without knowing the coordinates of its states), it is sometimes possible to derive non trivial information about an individual such as his home (i.e. the state that can be reached from almost all the states) and his work (i.e. the state that can be reached with the heaviest transition from the home). Moreover, more advanced knowledge might also be derived by looking for particular patterns in the graph such as the presence of a particular cycle or the existence of two different highly connected components that can indicate two different geographical areas. In the future, we plan to investigate in a more systematic and theoretical manner the knowledge that can be inferred from the mobility Markov chain of an individual.

References

[1] S. Gambs, M.-O. Killijian, and M. N. del Prado, "GEPETO: a GEoPrivacy Enhancing Toolkit," in *Proceedings of the International Workshop on Advances in Mobile Computing and Applications: Security, Privacy and Trust, held in conjunction with the 24th IEEE AINA conference, Perth, Australia*, April 2010.

[2] J. H. Kang, B. Stewarta, G. Borriello, and W. Welbourne, "Extracting places from traces of locations," in *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, 2004, pp. 110–118.

[3] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[4] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, June 2008.

[5] L.O. Alvares, V. Bogorny, B. Kuijpers, J. Antônio Fernandes de Macêdo, B. Moelans and A.A. Vaisman, "A model for enriching trajectories with semantic geographical information," *ACM-GIS*, pp. 22, 2007.

[6] S. Spaccapietra, C. Parent, M.L. Damiani, J. Macedo, F. Porto and C. Vangenot, "A conceptual view on trajectories," *DKE Journal*, vol. 65, no. 1, pp. 126–146, 2008.

[7] L. Sweeney, "k-anonymity: a model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.

[8] P. Golle and K. Partridge, "On the anonymity of home/work location pairs," *Pervasive Computing*, pp. 390–397, May 2009.

[9] L. Jedrzejczyk, B. A. Price, A. K. Bandara, and B. Nuseibeh, "I know what you did last summer: risks of location data leakage in mobile and social computing," *Department of Computing Faculty of Mathematics, Computing and Technology The Open University*, November 2009.

[10] D. Ashbrook and T. Starner, "Learning significant locations and predicting user movement with *GPS*," in *Proceedings of the 6th IEEE International Symposium on Wearable Computers*, 2002, pp. 101–109.

[11] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 38–46, 2006.

[12] Z. Changqing, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering personal gazetteers: an interactive clustering approach," in *Proceedings of the ACM International Workshop on Geographic Information Systems*, 2004, pp. 266–273.

[13] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition using relational Markov networks," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005, pp. 773–778.

[14] D. Reiter, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[15] J. Krumm, "Inference attacks on location tracks," *Pervasive Computing*, pp. 127–143, 2007.

[16] A. Pfitzmann and M. Hansen, "Anonymity, unlinkability, unobservability, pseudonymity, and identity management – a consolidated proposal for terminology," February 2008.

[17] M. P. Armstrong, G. Rushton, and D. L. Zimmerman, "Geographically masking health data to preserve confidentiality," *Statistics in Medicine*, vol. 18, pp. 497–525, 1999.

[18] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," *Proceedings of the ACM/USENIX International Conference on Mobile Systems, Applications and Services (MobiSys)*, 2003.

[19] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Computing*, pp. 46–55, 2003.

[20] T.-H. You, W.-C. Peng, and W.-C. Lee, "Protecting moving trajectories with dummies," in *Proceedings of the 2007 International Conference on Mobile Data Management*. IEEE Computer Society, 2007, pp. 278–282.

[21] F. Giannotti and D. Pedreschi, *Mobility, Data Mining and Privacy Geographic Knowledge Discovery*, 2008.

[22] D. Matt, K. Lars, and B. Athol, "A spatiotemporal model of obfuscation strategies and counter strategies for location privacy," *Lectures Notes in Computer Science*, vol. 4197, no. 4, pp. 47–64, 2006.

[23] B. A. Price, K. Adam, and B. Nuseibeh, "Keeping ubiquitous computing to yourself: A practical model for user control of privacy," *Int. J. Human-Computer Studies*, pp. 228–253, 2005.

[24] R. Hariharan and K. Toyama, "Project lachesis: Parsing and modeling location histories," *Lecture notes in computer science - Geographic information science*, vol. 3, pp. 106–124, October 2004.

[25] M.O. Killijian, M. Roy and G. Trédan, "Beyond San Francisco Cabs: building a *-lity Mining Dataset," *Workshop on the Analysis of Mobile Phone Networks (NetMob)*, 2010.