Citation for the published Conference paper:

Title:
Predictive Models for Seamless Mobility

Author:
Said Rutabayiro Ngoga, David Erman, Adrian Popescu

Conference Name:
Sixth International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks, HET-NETs

Conference Year:
2010

Conference Location:
Zakopane

HET-NETs 2010
ISBN 978-83-926054-4-7
pp. 273–284

Predictive Models for Seamless Mobility

SAID RUTABAYIRO NGOGA   DAVID ERMAN   ADRIAN POPESCU [a]

[a]Dept. of Communications and Computer Systems
School of Computing
Blekinge Institute of Technology
371 79 Karlskrona, Sweden

Abstract: User's location modeling and prediction is complex and a challenge for seamless mobility in heterogeneous networks. Location prediction is fundamental for pro-active actions to be able to take place. Markov models and information-theoretic techniques are appropriate to perform location prediction. The paper characterizes user's location as a discrete sequence. We survey and describe Markovian methods and information-theoretic techniques for location prediction.
Keywords: : Location prediction, Discrete Markov models

1. Introduction

Today, various wireless technologies and networks exist and capture different user's preferences with different services these technologies provide. The range of wireless access network technologies includes GSM, GPRS, UMTS, WiMax, WLAN (802.11 a/b/g/h) and Bluetooth. On the other hand, a great part of today's mobile terminals are already capable of having more than one interface active at the same time. Since the available wireless networks are complementary to each other, the vision of the Next Generation Networks (NGN) is to combine them over a common IP-based core network to support high usability (any system, any time, any where). This will empower mobile users to choose from their terminals which network interface to use in order to connect to the best available access network that fits their preferences. Key features are user friendliness and personalization as well as terminal device and network heterogeneity [1].

Existing solutions attempt to handle mobility at the link layer (L2) and the network layer (L3) with a particular focus on the reduction of handoff latency. The basic idea is to rely on the capability of L2 to monitor and to trigger changes with regard to L2 or L1 conditions. This further assists IP in handover preparation and execution. However, mobility in heterogeneous networks is mainly user centric [1]. This calls for dynamic and adaptive mechanisms adapted to the current situation of users, which makes control and management difficult for operators. The IEEE 802.21 working group is working on the Media Independent Handover (MIH) standard to enhance the collaborative use of the information available at the mobile terminal and the network infrastructure. The primary focus is on the decision phase of handoffs.

In addition, QoS is an important part of distributed multimedia applications [2]. This calls for applications to specify their QoS requirements before any attempt for resource reservation. Guaranteed QoS requirements for the whole duration of the service delivery makes resource provisioning a complex task in traditional fixed Internet. Thus, the high heterogeneity supported in NGNs introduces additional complexities such as bandwidth fluctuations, temporary loss of connectivity when clients disconnect from one AP and connect to a new one (handoffs) [3].

Seamless mobility therefore requires continuous resource reservation and efficient context transfer for handover management as the mobile terminal moves. The work in [4] suggests to act pro-actively against handoffs as one way towards seamless mobility. Furthermore, due to the complexity of handoffs there is a need to consider solutions based on full context awareness implemented at L5 in the TCP/IP protocol stack [3]. In other words, the high heterogeneity supported in NGNs, along with their respective specific handoff related procedures, requires a complete visibility of all aspects related to the handover process before any attempt for handover management.

Similarly, Blekinge Institute of Technology (BTH) suggests an architectural solution based on middleware and overlays [1]. This is an architectural solution implemented at L5, with the objective to offer less dependence on physical parameters and more flexibility in the design of architectural solutions. By this, the convergence of different technologies is simplified. A number of research challenges have been identified by the authors and are being worked on. These regard SIP and delay measurements, security, quality of Experience (QoE) management, overlay routing, node positioning, mobility modeling and prediction, middleware and handover.

As a functional block within the BTH architecture, the “Mobility Modeling and Prediction” overlay has the main function to perform mobility modeling and prediction. Specifically, the “Node Positioning” overlay collects the mobile terminal location informations, and transforms them into GPS coordinates. Further, the “Mobility Modeling and Prediction” overlay relies on this location information to determine and to learn the user mobility behavior for the appropriate handoff process to take place pro-actively. The user behavior includes user’s location as well as user’s preferences (applications or terminal devices), which can be observed in terms of requests.

Mobility prediction is fundamental for pro-active actions to be able to take place. We envision to develop mobility predictors that exploit the past user’s behavior and deduce how the user will behave in the future. These will include on-line models as well as off-line models. Furthermore, mobility prediction will be dynamic with the prediction agent running on the network as well as on the mobile terminal. Previous work [4, 5, 6] has shown that Markovian models and information-theoretic techniques are appropriate for on-line location prediction in wireless and cellular networks. According to [7], we should envisage other prediction techniques that have been developed in other disciplines, such as computational biology, machine learning, and Wold Wide Web. In this paper we present an overview of the state-of-the-art techniques for location prediction in wireless and cellular networks. This paper focuses on the mobile terminal location, as an example of the mobile's context. We survey different models and algorithms to track a user's mobility by determining and predicting their location. We consider only infrastructure-based networks such that the user's location can be described in terms of the topology of the corresponding access infrastructure and not the actual location of mobile.

The rest of the paper is as follows. In section two the BTH architecture for seamless mobility is described. In section three mobility prediction techniques are presented. In section four we present the model for user mobility. In section five models for movement history are presented. Finally, section six concludes the paper.

2. Mobility prediction techniques

Mobility prediction is needed to estimate where and/or when the next handoff will occur. This can be done at the link layer (L2), network layer (L3), and application layer (L5) in a TCP/IP protocol stack [1]. This is achieved by monitoring the mobile, gathering related location information and inferring the associated model of the mobile motion. Mobility prediction techniques can be classified into:

- **User-Centric:** the mobile stores its most frequent path locally or gets them from a home repository and builds a model of its motion behavior [4].
- **AP-Centric:** prediction is performed locally at an access point (AP), which builds the model using the motion behavior of mobiles encountered in the past [4].

The concept behind all these techniques is that users' mobility present repetitive patterns, which can be observed, gathered and learned over time. These patterns are merely the favorite route of travel and the habitual residence time. Therefore, the underlying mobility model can be characterized as a stationary stochastic process whose output is a sequence of observable location information.

3. Mobility modeling

The mobility models reported in [8, 9, 10, 11, 12, 13] are based on the assumption that a user's location is a priori information, which allows for simplification and analysis. The problem however is that mobility is a stochastic process, which has to be learned. Thus we need a scenario to describe user's mobility.

Let us consider a mobile user walking down a street while connected to a wireless network. The mobile communicates with a given access point and can regularly measure some information directly related to the path it follows (e.g signal strength or GPS coordinates). By reporting these measurements at discrete time steps $t = 1, 2, 3, ...$ we get a sequence of observable $\mathcal{V} = V_1, V_2, V_3, ...$

Since we are interested in gathering location information, let us divide the zone near an AP in small areas $\vartheta = v_1, v_2, v_3, ...$ where v_i is an index identifying a zone-Id. Therefore, at any time a mobile can be described as being in one of the area zones v_i, thus ϑ is a state space. In addition, the road layout in the area near the AP represents the way area zones are interconnected. While a mobile is moving, it collects v_i at discrete time steps, thus any sequence $v_1, v_2, v_3, v_4, ...$ represents a movement history.

Here we are interested in matching the sequence of observable $\mathcal{V}$ and the movement history ϑ in order to build a model of the mobile motion. This is referred to as state-to-observation problem. It is the base of the learning process for Markovian models.

4. Models for movement history

4.1. Discrete Markov models

Discrete Markov models relies on the simple mapping for the state-to-observation problem. Thus, given a state space $\vartheta = v_1, v_2, ..., v_N$ of size N, a mobile undergoes a change of state according to a set of probabilities associated with the state. If the time associated with state changes is $t = 1, 2, 3, ..., N$, V_t represent the actual state at time t. This mean, the movement history collected (L2 signals) correspond to the observable location information. In general, a full probabilistic description requires specification of the current state and all previous past states. However the first order Markov chain specifies only the current state and its previous state [14]:

$$\begin{aligned} P[V_n = v_n \mid V_1 = v_1, ..., V_{n-1} = v_{n-1}] &= P[V_n = v_n \mid V_{n-1} = v_{n-1})] \\ &= P[V_t = v_j \mid V_{t-1} = v_i] \end{aligned} \tag{1}$$

For a finite, irreducible ergodic Markov chain, the limiting probabily that the process will be in state j at time n, π_j, exists and is a unique non-negative solution of

$$\begin{aligned} \pi_j &= \sum_i \pi_i P_{ij}, \quad j \geq 0, \\ \sum_j \pi_j &= 1 \end{aligned} \tag{2}$$

Therefore, a complete specification of a first order Markov model requires specification of the model parameter N, the one-step transition propobility matrix $\mathbf{P}$ and the limiting probability vector $\mathbf{\Pi}$.

For the usage of Markov models, let us refer to the location tracking problem in cellular networks as mentioned in [5], where the movement history is recorded using time-movement based update scheme. The service area is composed of eight zones a, b, c, d, e, f, g, h interconnected with highway roads. Thus all location information (i.e., the sequence of observable information) related to the path a mobile follows corresponds to the movement history.

Given a typical sample path, e.g., $\mathcal{V} = aaababbbbbaabccddcbaaaa$, we are required to specify N, $\mathbf{P}$ and $\mathbf{\Pi}$ in order to build a first order Markov model, which can be used either as a model of this particular motion process or a generator of location information (sequence of observable). According to [5], a simple count approach is used to get values for these parameters.

The first order Markov model is appropriate for memoryless movement modeling, but in general the mobile user travels with a destination in mind. This requires considering the favorite route profiles of a user for the design of the motion model.

4.2. LeZi-update scheme

The mere purpose of this algorithm is to manage the uncertainty associated with the current location of a mobile based on specification of all previous past locations [5]. By doing this, we can only predict the current location of a mobile with a high degree of accuracy given its route prolife.

In other words, given a motion process $\mathcal{V} = v_1 v_2 v_3 ...$, we need to build candidate models based on previous last visited locations. The *null-order* model gathers all the *0-context* (no previous location) with their respective frequencies. Then the *first-order model* gathers all the *1-context* (one previous location) with their respective frequencies, and so on until we reach the *k-order* model, which collects the highest meaningful context. Further, the *entropy rate* of the motion process along each candidate model is calculated until we reach the limiting value $H(\mathcal{V})$ of the *entropy rate*, if it exists. Results in [5] show that for a stationary stochastic process the limit $H(\mathcal{V})$ exists and corresponds to the *universal model*.

For example, let us consider a movement history at a given time, $\mathcal{V} = aaababbbbbaabccddcbaaaa$, captured using L2 signals (see Fig.2). Under the concept of Markov model, candidate motion models of the movement history might be constructed by specifying, for the current location, the k-context locations ($k = 0, 1, 2, ...$). These are all the past locations of a user conditioned on the current location. Thus, for a sequence $aaababbbbbaabccddcbaaaa$, all the 0-context, 1-context, and 2-context along with their respective frequencies of occurrence are enumerated in Table.1. With reference to the illustration in [5], the entropy rate of the movement history, $\mathcal{V} = aaababbbbbaabccddcbaaaa$, are respectively: $H(\mathcal{V}|\text{ 0-order model}) = 1.742$, $H(\mathcal{V}|\text{ 1-order model}) = 1.182$ and $H(\mathcal{V})|\text{ 2-order model}) = 1.21$. Intuitively, the highest order model is directly linked to the highest meaningful context. This corresponds to the universal model.

Again, under a *first order Markov Chain* the mobile terminal maintains a cache scheme for a user's location until the next update. By reporting this, the system uploads the new location and recomputes location statistics. The *universal model* relies on specification of the highest meaningful context. Therefore, the mobile is needed to maintain a *dictionary-like* scheme of route profile. At each location update, it reports a *phrase-like* message as a

0-context	1-context		2-contexts		
$a(10)$	$a\|a(6)$	$b\|c(1)$	$a\|aa(3)$	$a\|ba(2)$	$a\|cb(1)$
$b(8)$	$b\|a(3)$	$c\|c(1)$	$b\|aa(2)$	$b\|ba(1)$	$d\|cc(1)$
$c(3)$	$a\|b(3)$	$d\|c(1)$	$a\|ab(1)$	$a\|bb(1)$	$a\|cd(1)$
$d(2)$	$b\|b(4)$	$c\|d(1)$	$b\|ab(1)$	$b\|bb(3)$	$b\|dc(1)$
	$c\|b(1)$	$d\|d(1)$	$c\|ab(1)$	$c\|bc(1)$	$c\|dd(1)$

Table 1. Context locations and their frequencies for a sequence $aaababbbbbaabccddcbaaaa$ [14]

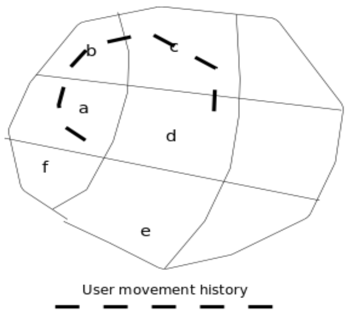

Fig. 1. A user movement history at a given time

new route explored. The system, on its turn retrieves the *phrase-like* message (new route), uploads the mobile's route *dictionary*, and estimates all predictable routes untill the next location update.

Thus, the degree of location uncertainty lies between two consecutive message updates. A good solution to reduce the number of updates is that upon detection of the same message, to delay message update until a new message arrives. According to [7], this is what achieves the Lempel-Ziv-78 scheme, a parsing technique for data compression. It is referred to as the *prefix-matching technique*.

The concept behind this formulation is that the movement history is compressible. The LeZi-update scheme, as reported in [5], is an interlaced learning and coding process that:

1. Uses the *prefix-matching technique*, during the learning process, to generate *phrase-like messages* of variable length.
2. Applies, during the encoding process, the *variable-to-fixed length coding scheme* and generates *fixed length dictionary indices* representing context statistics.
3. Stores these *fixed length dictionary indices* in a *symbol-wise model* that can be represented in a tree form.

With reference to illustrations presented in [5, 6], let us consider a typical movement history, e.g., $\mathcal{V} = aaababbbbbaabccddcbaaaa$, parsed as distinct *phrase-like* messages $a, aa, b, ab, bb, bba, abc, c, d, dc, baa, aaa$, by using a *prefix-matching technique*. Then, by applying the *variable-to-fixed length coding scheme*, each message is explored and symbol contexts together with their respective frequency of occurrence are stored in a *symbolwise context model*, which can be implemented by a tree (see Fig.2). Furthermore, context statistics are updated for each message entry in the tree using the following technique: *increment context statistics for every prefix of every suffix of the phrase*.

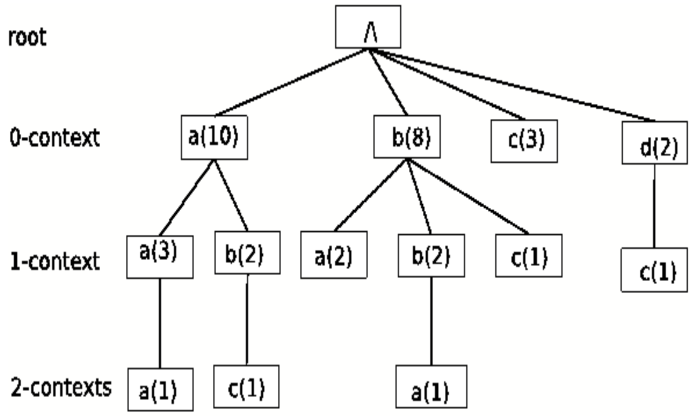

Fig. 2: **Symbolwise context model:** a contex represents a path from root to any node; a sub-tree rooted at a node represents a corresponding conditioned context; a path from root to leaves represents the largest context [14]

Therefore, as larger and larger messages arrives, the symbolwise context model, maintained in every mobile's location database, will contain sufficient information regarding user mobility profile. Further, the system uses this information to estimate the location of a mobile terminal until the next update message arrives, and this time period corresponds to the location uncertainty period. This gives insight to the predictability of higher order Markov models.

Another point to consider is with respect to specification of probability measures estimated based on context statistics stored in the tree form model. Since the system holds the mobile's route profile until the last update message (current context) in the form of all previous past locations, we are interested in estimating the next update message to be sent by the mobile. This is the condition probability distribution given the current context. This is what the *prediction by partial match* (PPM) techniques achieves for text compression [7]. However, due to space limitation, the work in [5] suggests the *blending-PPM* that work with total probability distributions instead of relying on conditional probability distributions.

4.3. Hidden Markov models

So far, we have mentioned the state-to-observation matching problem, which is the base of the learning process for the model of the mobile motion. By using the concept of Markov models, the movement history collected using L2 signals corresponds to the observable location information. In the following we extend this concept to include the case where the movement history information is absent, but can be generated through a set of stochastic processes. Thus, the observable location information is a probabilistic function of the physical location of a mobile. The physical locations of the mobile are hidden to the system, therefore the resulting model of the mobile motion is a doubly embedded stochastic processes called Hidden Markov model [15].

Typically for a L2/L3 handover case, characterization of user's mobility as a stochastic process relies on the ability of a mobile to monitor the user's motion behavior within the topology of the corresponding access infrastructure. Thus, the mobile collects some piece of information associated to its location along the route it follows. These informations can be very precise (GPS positions), less informative (signal strength), or very fragmentary (past pass by AP, time of day, or cell residence time) [4]. By reporting these informations to a handover management functional module such as L3 Mobile IP (MIP), the type of handover (horizontal or vertical) as well as the time to perform it can be determined. This is done based on knowledge of the access points surroundings. By using the concept of Hidden Markov models, on one side a prediction scheme can easily adapt any network, and on the other side it can accommodate any type of information emitted by the mobile terminal.

Therefore, a Hidden Markov model is characterized by:

- A hidden state space of size N, $\vartheta = \{v_1, v_2, ..., v_N\}$, where the state at time t is denoted q_t. The v_i(s) are the more likely physical locations of a user, a mobile terminal can collect.
- A set of observable location information per state of size M, $\mathcal{V} = \{V_1, V_2, ..., V_M\}$. The V_i(s) are any type of location information associated with a user location, a mobile terminal can send to the system.
- The single step state transition distribution, $\mathbf{P} = \{P_{ij}\}$

$$P_{ij} = P[V_t = v_j \mid V_{t-1} = v_i]. \tag{3}$$

- The observable location information distribution in state j, $\mathbf{B} = \{b_j(V_k)\}$

$$b_j(V_k) = P[V_k \text{ at } t | q_t = v_j] \qquad \begin{array}{c} 1 \leq j \leq N \\ 1 \leq k \leq M \end{array} \tag{4}$$

- The initial state probability vector, $\mathbf{\Pi} = \{\pi_i\}$

$$\pi_i = P[q_i = v_i] \tag{5}$$

Thus, given values for the parameters $\mathbf{N}, \mathbf{M}, \mathbf{P}, \mathbf{B}$ and $\mathbf{\Pi}$, a Hidden Markov model can be used as a generator or a model of a sequence of observable location information, $\mathcal{V} = V_1 V_2 V_3 ...$ For the usage of Hidden Markov Model, the work of [4] suggests a framework based on artificial intelligence technique that links directly a user movement to the handover mechanism.

5. Conclusion

Seamless mobility in heterogeneous networks requires continuous resource reservation and efficient context transfer for handover management as the mobile terminal moves. The BTH architectural solution for seamless mobility, as reported in [1], offers less dependence on physical parameters and more flexibility in the design of architectural solutions. Based on this, we envision to develop predictive models that exploit a user's previous behavior and deduce the user's future behavior. This paper focuses on user's location prediction as a fundamental for pro-active actions to be able to take place. We describe the user's locations as a discrete sequence. We present an overview of Markov models and information-theoretic techniques for location prediction.

Future work includes using simulation to perform comparative studies in order to evaluate various mobility prediction techniques in different mobility scenarios.

References

[1] A. Popescu, D. Ilie, D. Erman, M. Fiedler, A. Popescu, K. De Vogeleer: An Application Layer Architecture for Seamless Roaming, *Sixth International Conference on Wireless On-Demand Network Systems and Services (WONS 2009)*, Snowbird, Utah, USA, 2009.

[2] J. Jin, K. Nahrstedt: QoS Specification Languages for Distributed Multimedia Applications: A survey and Taxonomy, *IEEE MultiMedia*, 2004, vol. 11, no 3, pp. 74-87.

[3] P. Bellavista , A. Corradi, L. Foschini: Context-aware Handoff Middleware for Transparent Service Continuity in Wireless Networks, *Pervasive and Mobile Computing*, Elsevier, 2007, vol. 3, pp. 439-466.

[4] J.M. Francois, G. Leduc: Performing and Making Use of Mobility Prediction, *PhD Thesis*, Univesite de Liege, 2007.

[5] A. Bhattacharya, S.K. Das: LeZi-Update An Information-theoretic Framework for Personal Mobility Tracking in PCS Networks, *Springer Wireless Networks*, 2002, vol. 8, no. 2-3, pp. 121–135.

[6] F. Yu, V.C.M. Leung: Mobility-Based Predictive Call Admission Control and Bandwidth Reservation in Wireless Cellular Networks, *IEEE INFOCOM 2001*, 2001.

[7] D. Katsaros, Y. Manolopoulos: Prediction in wireless networks by Markov chains, *IEEE Wireless Communications*, 2009, vol. 16, no. 2, pp. 56-64.

[8] R. Thomas, H. Gilbert, G. Mazziotto: Infuence of the Moving of Mobile Stations on the Performance of a Radio Mobile Cellular Network, *Third Nordic Seminar on Digital Land Mobile Radio Communication*, 1988, pp. 1-9.

[9] I. Seskar, S.V. Maric, J. Holtzam, J. Wasserman: Rate of Location Area Updates in Cellular Systems, *IEEE Vehicular Technology Conference*, 1992, vol. 2, pp. 694–697.

[10] T. Camp, J. Boleng, V. Davies: A Survey of Mobility Models for Ad Hoc Network Research Wireless Communications and Mobile Computing (WCMC), *Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2002, vol. 2, pp.483-502.

[11] R. A. Guerin: Channel Occupancy Time Distribution in a Cellular Radio System, *IEEE Transanctions on Vehicular Technology*, 1987, vol. VT-35,no. 3, pp. 89-99.

[12] D. Hong, S. Rappaport: Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures, *IEEE Transanctions on Vehicular Technology*, 1986, vol. 35, no. 3, pp. 77-92.

[13] M.M. Zonoozi, P. Dassanayake: User Mobility Modelling and Characterization of Mobility Patterns, *IEEE Journal on Selected area in Communications*, 1997, vol. 35, no. 7, pp. 1239-1252.

[14] A. Bar-Noy, I. Kessler, M. Sidi: Mobile users: to update or not to upadate? *ACM Wireless Networks*, 1995, vol. 1, pp. 175-185.

[15] L. Rabiner: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Morgan Kaufmann Readings in Speech Recognition*, 1990, vol. 53, no. 3, pp. 267-296.