HABITS: a Bayesian filter approach to indoor tracking and location

Eoghan Furey*, Kevin Curran and Paul McKevitt

Faculty of Computing and Engineering,
University of Ulster,
Derry, BT48 7JL, Northern Ireland
E-mail: furey-e1@email.ulster.ac.uk
E-mail: kj.curran@ulster.ac.uk
E-mail: p.mckevitt@ulster.ac.uk
*Corresponding author

Abstract: Using Wi-Fi signals is an attractive and reasonably affordable option to deal with the currently unsolved problem of widespread tracking in an indoor environment. History aware-based indoor tracking system (HABITS) models human movement patterns by applying a discrete Bayesian filter to predict the areas that will, or will not, be visited in the future. We outline here the operation of the HABITS real-time location system (RTLS) and discuss the implementation in relation to indoor Wi-Fi tracking with a large wireless network. Testing of HABITS shows that it gives comparable levels of accuracy to those achieved by doubling the number of access points. We conclude that HABITS improves on standard real-time location systems in term of accuracy (overcoming blackspots), latency (giving position fixes when others cannot), cost (less APs are required than are recommended by standard RTLS systems) and prediction (short, medium and longer-term predictions are available from HABITS).

Keywords: indoor location positioning; wireless tracking; Bayesian filters; human movement.

Reference to this paper should be made as follows: Furey, E., Curran, K. and McKevitt, P. (2012) 'HABITS: a Bayesian filter approach to indoor tracking and location', *Int. J. Bio-Inspired Computation*, Vol. 4, No. 2, pp.79–88.

Biographical notes: Eoghan Furey has recently graduated with his PhD degree from the School of Computing and Intelligent Systems at the University of Ulster. His research interests include local positioning systems, location-based software, wireless sensor networks, body area networks and machine learning.

Kevin Curran is a Reader in Computer Science at the University of Ulster. His achievements include winning and managing UK and European framework projects and technology transfer schemes. He has published over 600 research papers to date in the field of distributed computing. He is the Editor-in-Chief of the *International Journal of Ambient Computing and Intelligence* (*IJACI*). He is a Fellow of the Higher Education Academy, a Fellow of the British Computer Society and is listed by Marquis in their prestigious Who's Who in Science and Engineering. He is also listed in the Dictionary of International Biography and by Who's Who in the World.

Paul McKevitt is the Chair in Intelligent MultiMedia at the School of Computing and Intelligent Systems. He has been a Visiting Professor of Intelligent MultiMedia Computing in the Institute of Electronic Systems, Aalborg University, Denmark and a British EPSRC Advanced Fellow in the University of Sheffield, UK. He has been a Visiting Professor at LIMSI-CNRS, Visiting Fellow at Dublin City University, Ireland, and Research Fellow in Computer Science at New Mexico State University, USA. He has published numerous papers in international conferences, research books and journals and is the Editor-in-Chief of the *International Computer Science Research Journal* and *Artificial Intelligence Review* (Springer, Heidelberg). His primary research interests are in natural language processing (NLP), multimedia, creative technologies and the general area of artificial intelligence.

1 Introduction

To accurately position an object, intelligent prediction is required (Petzold et al., 2005; Vintan et al., 2004; Gellert and Vintan, 2006). These methods enable the accuracy levels of the estimates to be increased. When a human makes estimation about where a person or object will be located in the future, they automatically perform the complete calculation. To enable computers to replicate these calculations and to allow them to work with a number of different objects requires a number of artificial intelligence techniques. Technologies such as GPS and mobile phones have made this a hot area of research in recent years. This has been driven by commercial location-based software, from satellite navigation devices to targeted advertising on mobile phones. Application to indoor environments is a largely under researched area.

Outdoors, using GPS traces to try and learn next location has been attempted by Han (2004) for someone on foot and in Froehlich and Krumm (2008) for vehicles on a road. More recently, data gathered from mobile phone records has been mined to try and find patterns of movement which could be used to try and make next location predictions (González et al., 2008). Indoors, this is a largely under researched area, however a number of ‘smart environments’ have been setup such as the work (Petzold et al., 2005). Here, specific sensors on doors were utilised to provide movement patterns. A hidden Markov model (HMM) and a neutral network (NN) were applied to the data and successful predictions were made. Ashbrook and Starner (2002) used a Markov chain model and K-means clustering algorithm to attempt to predict future movement. They clustered GPS data to find significant locations and then built a first and second order Markov models using location as state to try and predict future movement. It is possible to create nth order Markov model where probability of the next state is dependent not only on the current state but on the previous n – 1 states. For some examples, considering the 2nd order can yield more accurate results as in the case of probability of transition from A $\rightarrow$ B is 70% but the probability of transitioning from B $\rightarrow$ A $\rightarrow$ B is 81%. This could be explained by a situation where A was a Shop and B was Home. If the shop was on the main road from Home then the probability of going from A to B (Shop to Home) is 70%. However, if the journey started at home and went to the shop, return to home could be more probable (perhaps getting something for dinner?). This demonstrates a situation when higher order models are useful and give extra information. It raises the question of which order of model is suitable for prediction. Ashbrook and Starner (2002) conclude that this depends on the quantity of data available. Other factors affecting their probabilities were due to the large distances travelled and the fact that their tests took place outdoors. They also found that changes in routine would take a long time to show up in their model and they suggested a possible method of weighting certain updates, but warned that this could lead to model that was somewhat skewed. Han (2004) attempts to build upon the work of Ashbrook and Starner (2002) by using a self-organising map (SOM) as a means of learning without pre-knowledge. To use a supervised learning method to learn patterns of movements, pre-knowledge of the person is required, however a SOM can overcome this. An SOM is an ‘unsupervised learning neural network’ which can preserve the topology of a map as it creates it. Sang uses an SOM to convert sequences of raw GPS data into meaningful patterns which are in turn applied to a Markov chain approach. They used the output from the SOM to learn a first order Markov model and to try and make predictions of next location from it. Their data was gathered based on a university campus. While their method looks promising, their results are very sparse and their conclusion of ‘acceptable’ prediction accuracy is of little value.

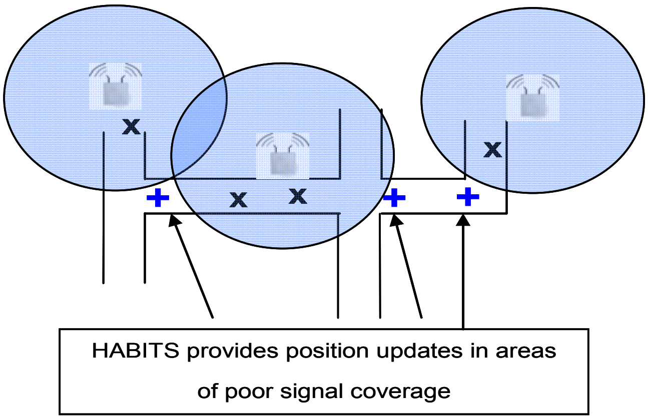
In indoor localisation, the area of movement prediction is sparsely researched. This is due to the fact that any sort of indoor localisation is a relatively recent phenomenon, however a number of research studies have been conducted in this area. One of the first research projects that considered future movement was Microsoft Research’s RADAR project (Bahl and Padmanabhan, 2000). This was the first significant attempts to track indoors using 802.1 Wi-Fi signals. Due to the severe problem of signal attenuation it was difficult to get an accurate fix on position using received signal strength (RSS) measurements alone. Position was occasionally reported in locations that were not possible or at least highly unlikely. An effort to overcome these problems is described in Bahl et al.’s (2000) paper. They concluded that the next location position should be close to the last reported one. Their Viterbi-like tracking algorithm deals with a situation of when two physically separate locations are close together in signal space (due to aliasing). The shortest path is depicted in bold. The likely trajectory is calculated based on the previous unambiguous location and a guess of somewhere in between the two is given. Between vertices i and j there is an edge d_{ij} whose weight is calculated based on the Euclidian distance between the locations i and j. This approach has been shown to significantly reduce the accuracy error in locating a user who is walking. They tested the Viterbi-like approach against a nearest neighbour in signal space (NNSS) and an NNSS-AVG (where the three nearest neighbours in signal space were averaged to estimate location) and it was found to significantly outperform the others. Median distance error for NNSS (3.59 m) and NNSS-AVG (3.32 m) are 51% and 40% worse, respectively compared with Viterbi (2.37 m) (Bahl et al., 2000).

2 Related work

Using past movements to improve localisation is an under researched area, although a number of useful studies have been conducted. Mature technologies, such as GPS navigation, have used this approach to predict where and when a user will re-emerge from a tunnel. Also, the approach is used in cellular systems to predict which cell a mobile user will enter next. Petzold et al. (2006) used various machine learning techniques and mathematical methods to model indoor movement patterns. Using these models, predictions of the next location of a certain user have been made with 69% accuracy without pre-training and 96% accuracy with pre-training. Another study by Zhou

(2006) has shown that by using knowledge of previous movements, overall accuracy could be improved by 14.3% and estimations of the wrong room and wrong floor could be improved by 69.7% and 50%, respectively. A recent study (Song et al., 2010) of past locations from mobile phone records, found that general human mobility patterns over a wide area were predictable 94% of the time. A related, relatively new field of reality mining (Eagle and Pentland, 2006) has been developed which records movements of people throughout the day with the intention of predicting future behaviour. These studies on learning human movements for prediction show that the research community is beginning to utilise movement information in a new way.

Figure 1 HABITS overcomes the need for extra APs (see online version for colours)



Using previous movements to help improve accuracy levels in Wi-Fi positioning has been attempted in a number of studies (Bahl and Padmanabhan, 2000; Bahl et al., 2000; Lassabe, 2009), but the focus has been on trying to improve the RSS-based problems.

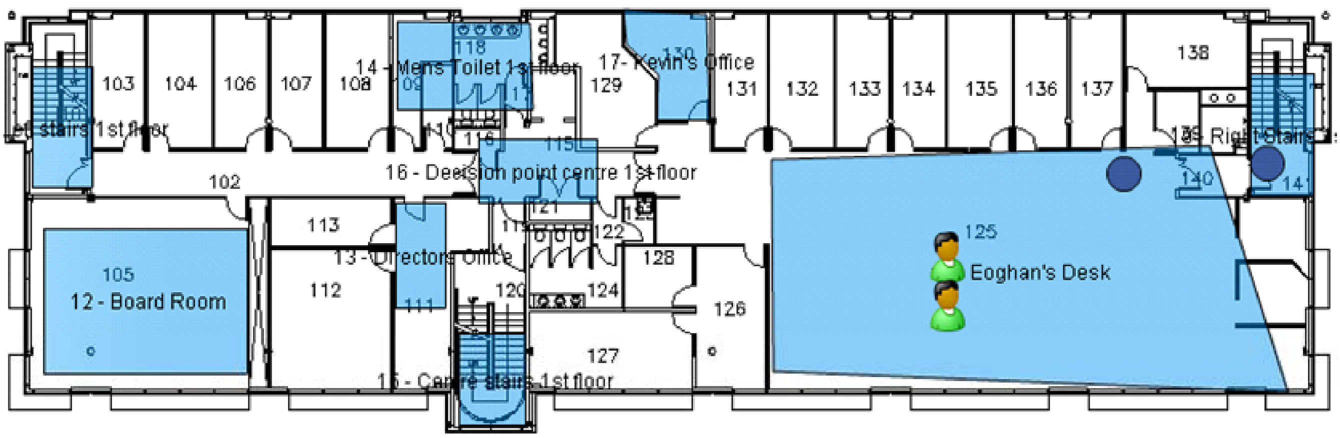
The HABITS framework which we introduce in the next section does not try and improve on the RSS methods but instead uses the movement habits of users as a means of adding intelligence to the system. This knowledge is then used to overcome signal black spots and to predict where the user will travel to next as Figure 3 shows.

Previously, we have implemented and tested a number of real-time location system (RTLS) systems and the results of these can be found in the study on behalf of JANet UK – location awareness trails (Furey et al., 2008). Of these, the Ekahau RTLS utilizes the existing Wi-Fi network and in our tests was the best overall indoor tracking system. For this reason, it was the chosen platform for implementing HABITS. A number of stages are involved in implementing this system and these are outlined in the next section.

3 HABITS

Past movement habits have been shown to be repeated by humans, usually to do necessary tasks or just to take what is felt to be the path of least resistance. These habits are often linked to particular tasks that need to be done regularly. Movement habits are the same as other types of habits in that they tend to be regularly repeated. While each of us has a number of habits or patterns that appear to be unique to us, much more probable is that we share habits with others. In our approach, history aware-based indoor tracking system (HABITS) is use to enhance an existing tracking system from Ekahau. The technology of the underlying tracking system or the positioning methods used is not relevant. HABITS is designed to be generic with application to many potential domains. The three main components of HABITS are a connected graph, a discrete Bayesian filer and a set of logic rules. The focus of HABITS is to combine these three methods in a novel way, allowing for predictions of human movement habits. These predictions overcome the latency of updates from currently available systems and enable them to make predictions of likely future movement.

Figure 2 Zones showing areas of interest (see online version for colours)



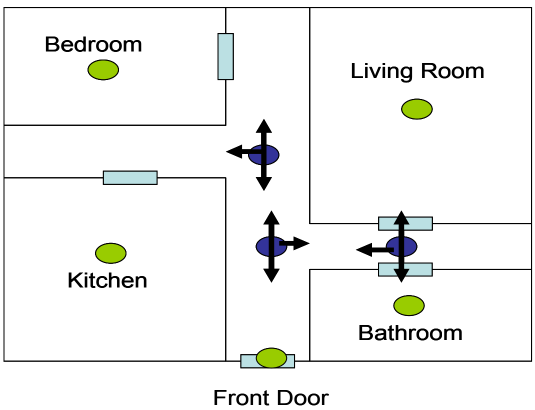
In order to collect historical movement data a topological map of the test area is created. A topological map is one which consists of a number of nodes representing places of interest which are connected by edges representing paths where a user may travel. These areas are covered by zones in Ekahau which allow for reporting of when a person carrying a mobile Wi-Fi device enters or leaves them. The zones shown in Figure 2 represent areas that are passed through frequently on the ground floor in the ISRC. Each of these zones can be considered to be a node in a connected graph. The positioning of these zones is a manual process based on expert knowledge of where a user is likely to stop and areas where they would pass through often. Also used are locations where a user has a number of options of where next to travel. The locations of these zones relative to one another can now be represented as an adjacency matrix and hence a connected graph. To do this each node in the graph representing a zone is given a unique ID between 1 and n, were n is the number of zones.

The underlying principle of our approach involves representing the movement areas as a graph which in turn is represented by a number of matrices; incidence, distance and transition.

These constraints show where it is possible for a user to go and where not, the distance between points of interest (for our purposes) and eventually represent the probability of going from one area to another. Methods of modelling the travel environment exist and of these, a graph structure closely represents the possible paths. The nodes in the graph can be positioned to represent areas of interest, decision points or places where the user stops. In between these locations are the paths that may be travelled.

The paths are edges and those locations of interest are the nodes/vertices of a connected graph. The graph structure clearly represents the connections between nodes and therefore areas in the real building. It shows which locations are connected either directly or indirectly.

Figure 3 Node positions in house (see online version for colours)

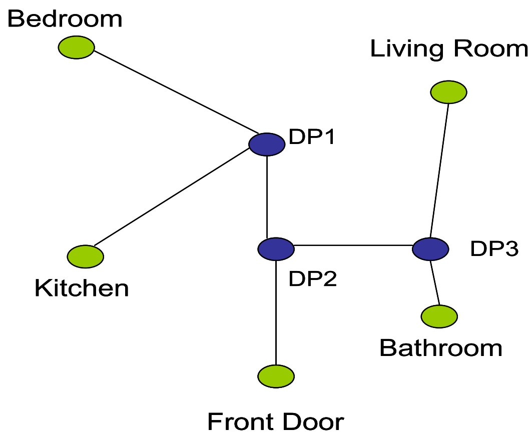


When studying a building plan or road map this information is normally clear to see however, in a new location, different methods need to be used to identify these areas of interest. Areas where a user stops for some reason may be thought of as 'base nodes'. Stopping for reasons such as sleep, eating, call of nature or work are some of the main reasons why humans would habitually stop at the same location. While for many people these may be in the same room or adjacent rooms, in the developed world, relatively large houses exist and these functions often occur in a number of different rooms with travel paths between. Examples of these rooms could be bedroom, kitchen, bathroom and living room. Movement between these rooms is often only possible by one or two different routes. The layout of a typical house (in the developed world) may be represented as a connected graph. In Figure 3, the green nodes represent stopping locations and the blue nodes represent decision points.

A connected graph or topological map of these nodes is shown in Figure 4. Learning the locations of these points can be performed automatically in a number of ways, all of which require an underlying tracking system to be installed.

Learning these significant locations can be carried out automatically by computers. One methods of achieving this is to plot the locations where there was a significant delay between movements. These would indicate the areas where a person was stationary. Even within the same room these points are not all likely to be in the exact same location. To extract wait nodes from a large number of estimates, clustering techniques are used to group the updates together, revealing the main stopping locations. When the nodes have been discovered and coded with numbers for name they may be represented as an $n \times n$ adjacency matrix where n is the number of nodes and the matrix details specific information about the graph. If a connection exists between the nodes then in the matrix location ij which represents the connection from i to j place a 1, if no connection exists then place a zero. This enables the paths between nodes to be represented mathematically and the matrix can easily be processed by a computer program.

Figure 4 Connected graph with node connections (see online version for colours)



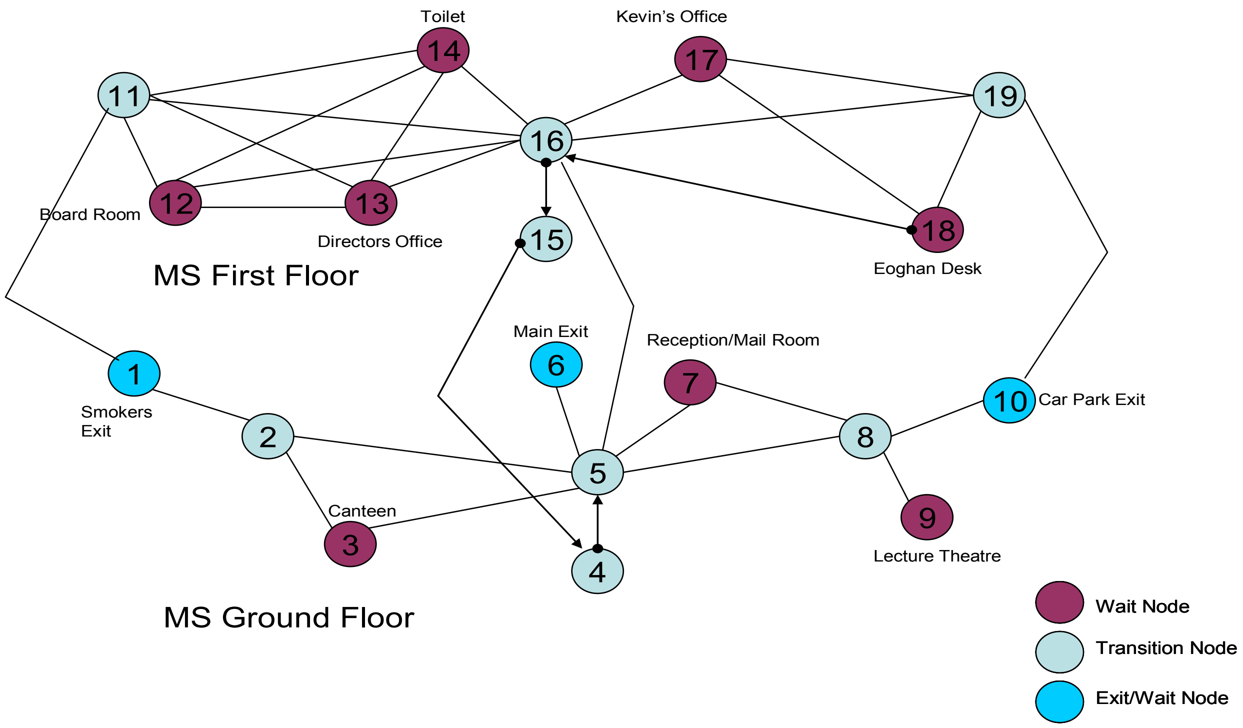
When the node locations have been discovered and the distance between two nodes is known, travel time between nodes may also be calculated automatically by the underlying tracking system Ekahau. Average walking or travelling speed for each user is estimated by using: speed = distance/time. Knowledge of the relative travel times between nodes is then used to generate a distance matrix with distances between each node being calculated based on average user speed. The distance matrix values are in the same positions in the matrix as the 1's are in the adjacency matrix. A transition matrix showing the probabilities of travelling from one node to the other is built up by monitoring the person's travel through the nodes. Again, a number of methods exist to solve this, but a straight forward method is to use the sequence of all nodes traversed for a day, a week or all travel time (depending on the application). String identification tools can be used giving the sequences of nodes and from this mathematical functions can generate a transition matrix. As before, the size of the matrix corresponds to n × n and at each location (node) a count is kept of the movement through it and where it goes to next. In the sample house scenario, consider movement from the kitchen, through a decision point to either the bedroom or another decision point. Hypothetically, it could be found that the probability of going from the kitchen to the bedroom was 12/50. This would equate to a situation where out of 50 times leaving the kitchen, 12 of these journeys were to the bedroom. 12/50 would give a probability of 0.2 of travelling to the bedroom meaning that 0.8 or 38 journeys went the other way to the next decision point. This is how transition matrices are created and knowledge of them gives a first order Markov chain.

By querying the EPE the list of zones is retrieved and each zone is allocated a unique ID. For the two floors in the MS building there are 19 zones in total. Using the zone map, an adjacency matrix of size n × n is manually created. The corresponding zone-node list allows all zone data from the EPE to be manipulated as if each zone was the node in the connected graph. Figure 5 shows a connected graph representation of the two floors in the ISRC. The edges between nodes show paths that may be travelled and represent the movements of Wi-Fi tracked people in the building. The numbers on the nodes are those used by the zone to node conversion table.

In addition to the adjacency matrix, the distance between nodes is required to allow for dead reckoning when HABITS is in operation. Two methods are available to calculate this. The first involves manually taking measurements from the ESS. The second involves manual measurement of the distance from one node to another. By timing an averaged paced walk over this distance the speed of movement is calculated by the standard formula speed = distance/time. When this guide for speed is calculated, a standard walk is taken around the whole building with the time to move from one node to the next recorded. With the travel time from one node to the next available a simple calculation can convert this into distance measurements and the distance matrix is created with values in the same locations as the ones in the adjacency matrix.

At any time along the chain, only the current location gives the probability of going to the next location. A simple Markov chain like this gives some idea of the next node but alone it would not be enough to model real human movement habits. Raising the order of the model to consider the previous two nodes would help in some locations but (Froehlich and Krumm, 2008) proved this needs to be done with a large dataset which takes a considerable time to generate.

Figure 5 MS building represented as a graph (see online version for colours)

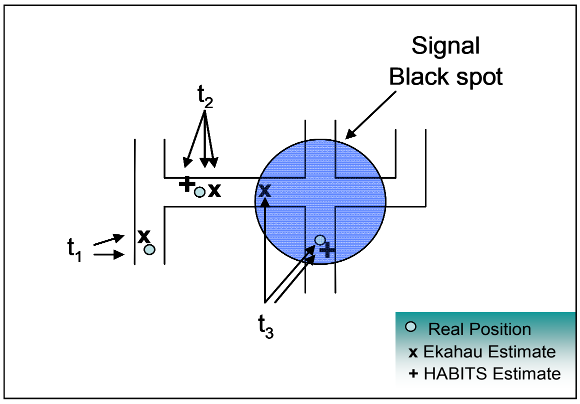


Maintaining a separate transition matrix for each day and/or each time period would improve the accuracy slightly but the system would not be expandable to a large area due to becoming overly complex. To predict the most likely next location with a useful degree of accuracy requires more than just a simple one state Markov chain. The movement habits of people are dependent on a variety of factors and to improve the accuracy of any model requires that more of these factors are considered.

The underlying Ekahau tracking system gives the initial location, bel(x_{t-1}). The transition matrix provides the belief, bel(x_t) when combined with the information in the perceptual model and the system dynamics. This outputs the probability of moving to the next node when given just the previous one and no other information. HABITS uses more information than just that provided by the first order Markov chain. As a Bayesian filter only works, for instances, that hold to the Markov assumption (meaning only a single order model), substantial information is being left out about commonly travelled paths or sequences of nodes. Froehlich and Krumm (2008) found that the more nodes they had information about (previously travelled), the higher the chances of predicting their final location. If an order (3 for example) Markov model was used, then for some paths, the predicted location probability would be much higher, however it would also take into account shorter journeys and could have sequences like 2-4-2 which would include changing direction completely. Taking into account higher order models results in overcomplicated calculations. The notion of 'preferred paths' (PPs), however allows for the same information to be gathered without keeping track of every path.

As part of the definition of a habit, it states that they are routines of behaviour that are repeated regularly. An approach to viewing habits could be that they take places between distinct locations, but it does not mean that those locations are necessarily adjacent locations. The paths may go through a number of intermediate nodes and a common journey could be kitchen to toilet in the example in Figure 2. This would involve travelling through fur different nodes but may be repeated a number of times a day. If a pattern occurred more often than a set number of times then it could be considered habitual. Habitual journeys of this sort we call 'PPs' and they can be mined from the string of all nodes visited. There could also be a temporal link between taking these PPs and a certain time period. This information can be used to adjust the output of HABITS prediction. It can also help with the identification of final destination which is another aim of HABITS. A PP is also stored as a vector and may be temporally linked to a specific time period if required. Some would be more frequently travelled at particular times than others. When on a PP, the information is used to increase the accuracy of the future location estimate. A last influencing factor to be considered in some instances is a rule that takes into account when people change their habits depending on who they are with. In largely populated environments certain people's movements have an influence on others. In, for example, going for lunch it may be that a particular person is a common factor in most locations. This is discovered by checking to see if people travel routes matched up temporally and if so, was one dominant over the other? When this is the case, a rule is applied in the same manner as the PPs, influencing the prediction. HABITS combines a number of different elements to produce future location predictions. The inputs to the Bayesian filter include the motion model showing where it is possible to go in the next step, the sensor model giving the accuracy of the updates from the underlying tracking system, the learnt historical belief and the location updates from the base system. When the filter has all the necessary information to give a prediction, it is run through a set of rules to improve the accuracy of its estimates. HABITS is designed to be able to operate on any type of tracking system to allow it to track between its updates and to give future predictions

Figure 6 HABITS overcomes need for extra APs (see online version for colours)



HABITS does not attempt to improve the underlying Wi-Fi positioning system but is used in conjunction with it to improve overall performance. While HABITS uses the same radio signals and equipment as other systems, it enables positioning and continuous real-time tracking with increased accuracy, and in areas that were not previously possible. However, HABITS will only work in certain environments where people follow particular habitual movement patterns. Examples include work environments such as factories or hospitals. When a mobile device is tracked by the Ekahau RTLS and the HABITS algorithm is applied, it can still be tracked when it is no longer within line of sight (LOS) of three or more access points (AP). This is normally the minimum required for accurate localisation. The highest frequency rate of position updates from the Ekahau RTLS has been found to be 5 s. These updates are often up to 15 seconds apart. Each update is sent to HABITS along with the learnt historical movement data and from this an intelligent prediction of the next likely location is given. Short-term predictions effectively fill in the blanks in between updates from the Ekahau system. HABITS does not try and improve on the RSS positioning methods currently in use, but instead uses knowledge of the movement habits of users as a means of adding intelligence to existing tracking systems. This knowledge is then used to overcome signal black spots where existing systems fail (Figure 6) and to predict where the tracked user will travel to next.

At time, t_1 (Figure 6) the Ekahau RTLS can give a position estimate that is close to the true position. At time t_2 both the standard Ekahau RTLS and the HABITS system also give an accurate estimate. However, at time t_3, the Ekahau system is no longer accurate due the user travelling through a signal black spot. This is where HABITS can dramatically improve standard location tracking systems and provide accurate updates of where the user is located.

4 Evaluating predictions with HABITS

The various inputs to HABITS are combined using a number of artificial intelligence techniques. The first is an idea described by Fox (2003), which is extensively used in robotics – that of a discrete Bayesian filter. This filter works in conjunction with the graph matrices and gives out a probability estimate for the next location or a number of possible locations when at a particular node. Pseudo code in Figure 7 shows the basic operation of a discrete Bayesian filter.

Figure 7 Discrete Bayesian filter (see online version for colours)

<u>General Algorithm for Bayes Filtering</u>

1 Algorithm_filter(bel(x_{t-1}), u_t, z_t):

2 for all xt do

3 $\overline{bel}(x_t) = \sum p(x_t \mid u_t, x_{t-1})\, bel\,(x_{t-1})$ **(PREDICTION STEP)**

4 $bel\,(x_t) = \eta\, p(z_t \mid x_t)\, \overline{bel}(x_t)$ **(UPDATE STEP)**

5 end for

6 return bel (x_t)

Inputs belief bel(x_{t-1}) at t-1; most recent control u_t + measurement z_t.
Output is the belief bel (x_t) at time t.

It is basically a data fusion technique which uses Bayes theorem as a means of predicting the probability of moving from one node to the next. The various movement and sensor constraints are represented as mathematical models (u_t) which work along with the updates from Ekahau (z_t) and the transition matrix data, $p(x_t \mid x_{t-1})$ to give a prediction of next location. The n symbol in Figure 11 – line 4 is used to normalise the result to 1. However, this prediction alone is not sufficient to model a user's movement habits accurately.

Fuzzy logic is derived from fuzzy set theory and is a technique used when reasoning is approximate rather than precise. *Fuzzy rules* are similar to normal rules except that there are degrees of correctness. In this way, we can represent ideas like "John *often* goes to the canteen for lunch". The addition of the *fuzzy rule base* is to overcome one of the weaknesses of the Bayesian filter. This weakness is that it is tied to the *Markov assumption* which states that all the necessary information needed to predict the next step is located in the current step. This makes the discrete Bayesian filter into a Markov chain, which is any random process that is bound by the Markov assumption. As the Markov assumption does not hold true in our case, it has been overcome by the creation of a hybrid Bayesian-fuzzy filter/rule base. This gives us the best of both and allows for extra habits, such as being on PPs, to be included which do not fit into the discrete Bayesian filter.

4.1 Operational scenario

The scenario next describes HABITS' operation in a real world scenario. The user is travelling from his desk to leave the building for lunch. The code and accompanying diagrams (Figures 8 to 10) show what the probabilities are of going to a particular node. This shows how the knowledge of a user's movement habits can be used to give predictions to a useful degree of accuracy. A possible use of these predictions is explained in the last section.

1 If tag = Eoghan

2 *node = 5 and previous node = 4*

3 node 5 NOT = wait node

4 Action = calc next node

5 *Next node = Either 2, 3, 6, 7, 8 (All have non-zero Probability) – Figure 13)*

6 Check time period = Lunch

7 *If time = Lunch THEN next node is 6 or 3 (Probability > 80%) – Figure 14 lunch temporal rule*

8 Check other users in area

9 *If with John THEN next node = 6 (John does not go to the canteen!) – Figure 15 other user rule*

10 *If with Mary THEN next node = 3 (Mary usually goes to the canteen!)*

11 If alone then next node = 6(40%) OR 3 (40%) – wait for more info!

12 Use speed and distance to calculate position at time t

13 *Calc and show positions at t + 1, t + 2 .. t + n.*

Long-term predictions are related to the likelihood that a particular node will be visited during a particular time period. This could be later the same day or later in the week. For example, HABITS tells us with 85% confidence that during the lunchtime period that users 1, 2 and 5 will all leave their base nodes and will exit the building through the front door. To calculate these movements, the repeatability of a PP within a time period (TP) is considered. Table 2 shows the frequency of each PP in each TP for User 1. The frequency tables for all users are automatically extracted from the learning data by HABITS. This works on the principle that if a journey has occurred every Tuesday morning for three weeks, then there is a high probability that it will occur the next Tuesday, all else being equal. There is no guarantee that this will occur but evidence from the tests show that it is highly probable. Users 3 and 4 both travel to node 3 (canteen in test area) on 13 out of the 15 days used for testing. These patterns allow HABITS to predict who will go where, when, for commonly repeated journeys with a useable degree of accuracy. The green boxes in Table 2 show PPs that have occurred twice during the same TP on the same day and the yellow boxes show those that have occurred on three or more occasions.

Figure 8 Probability from Bayesian filter (see online version for colours)

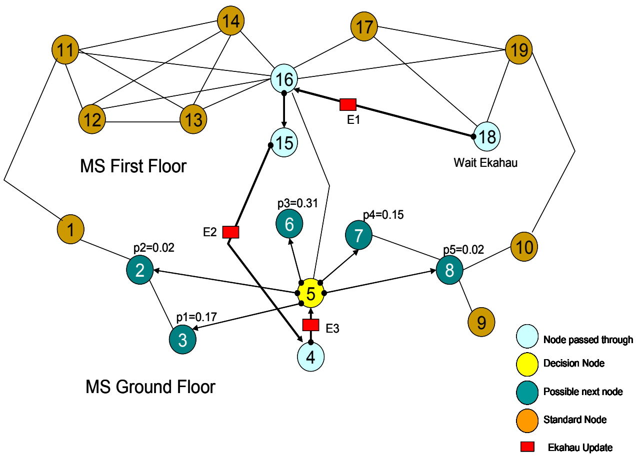


Figure 9 Probability from temporal fuzzy rule (see online version for colours)

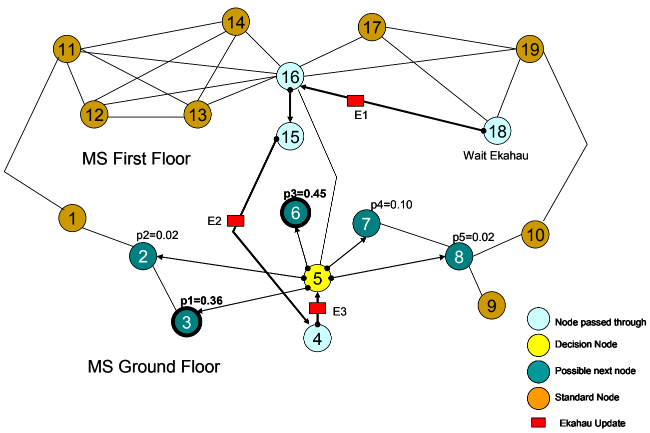
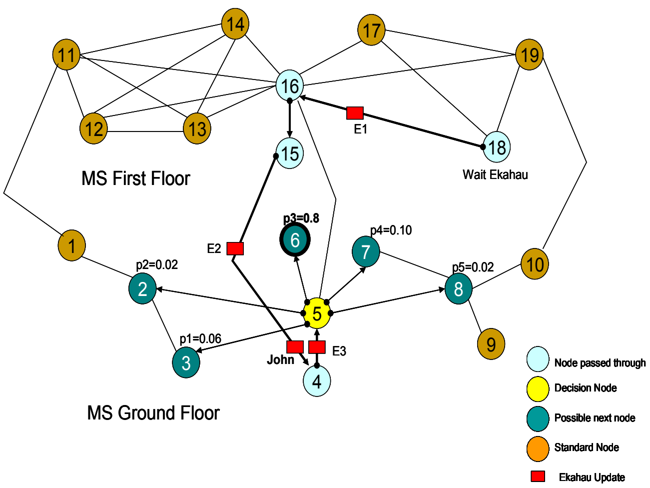


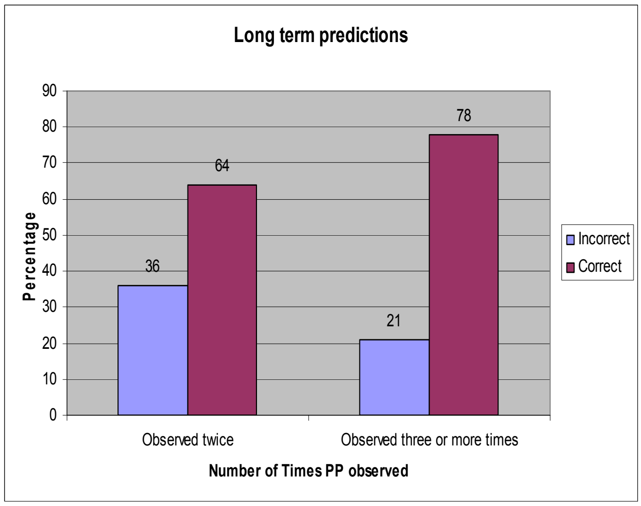
Figure 10 Probability from other user fuzzy rule (see online version for colours)



Applying this data to the test data for User 1 yields the results shown in Figure 11.

These show that when a PP has only been observed twice, the successful predictions occur 64% of the time meaning that 36% of the time the predictions are incorrect. However, when a PP has been observed three or more times within a time period during a particular day, then the predictions are correct 78% of the time. Table 2 lists the overall average predictions for all of the test subjects. These are compiled by running HABITS on the test data available. Overall User 2, an RA is the most predictable. The short and medium term predictions for all subjects are similar, however, the long-term predictions are much lower for User 3 (student) and User 5 (academic).

Figure 11 Long-term predictions from user 1 (see online version for colours)



The users base node (desk) is the key to making predictions with HABITS. Of the total number of journeys made during the test period, 42% had the base node as the destination and 47% had the base node as the starting point. This means that 89% of all journeys undertaken by our test subjects involved travel to or from their base node. All of the test subjects showed very high (> 89%) predictability when travelling to their own work station. When travelling from the base station, the final destination was more difficult to predict. However, HABITS still predicted the correct destination over 60% of the time for all users. User 4, the RA, was still predictable in over 90% of their journeys from their base station.

Other journeys in the building had a much lower predictability. Some small patterns were apparent such as going to the toilet after the canteen, but overall these journeys proved to be beyond the predictability of HABITS. The average predictability of final destination of any of the test subjects was almost 80%. This means, in our test week, for four out of every five journeys taken, HABITS correctly predicted the final destination. It must be noted that these results are for journeys of greater than two nodes.

Table 1 Frequency of PPs during time periods for user 1 (see online version for colours)

	PP	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*
	Morning	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	3
Mon	Lunch	0	0	0	0	4	3	0	0	0	0	0	0	4	3	1	1	2	3	2	0	0	0
	Evening	0	0	0	0	2	2	0	0	0	1	0	0	2	2	1	1	1	1	3	2	3	2
	Morning	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
Tue	Lunch	0	0	0	0	2	2	0	0	0	0	1	1	2	2	0	0	2	2	0	0	0	1
	Evening	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	2	2	4	2
	Morning	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Wed	Lunch	0	1	0	0	3	3	0	0	0	0	0	0	3	3	0	0	3	3	1	1	1	1
	Evening	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	2
	Morning	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Thur	Lunch	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	1	1	0	0	0	1	0
	Evening	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	2	0
	Morning	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	2
Fri	Lunch	0	0	0	0	1	3	0	1	0	1	0	0	1	3	0	0	2	1	0	0	1	0
	Evening	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0

Table 2 Predictions for all test subjects

Test subject	*Job*	*Predictability*		
		Short	*Medium*	*Long (3 or more)*
User 1	Research student	82%	83%	78%
User 2	Research associate	85%	88%	82%
User 3	Research student	81%	77%	55%
User 4	Research associate	87%	82%	76%
User 5	Academic	80%	78%	62%
Average		83%	81%	70%

The testing of HABITS revealed a number of interesting facts. HABITS is suitable in environments where people follow particular movement patterns. The two RAs (User 2 and User 4) proved to have much more predictable habits than the other three test subjects. It was concluded that this was because they were paid to sit in the same spot each day and had set times for breaks. User 5 (academic) and Users 1 and 3 (PhD students) did follow repeating movement patterns but these did not follow a rigid timetable. The conclusion from this was that the academic had a changeable meeting schedule, whereas the student made particular journeys when he/she felt like it.

5 Conclusions

Widely used techniques such as the Kalman and particle filters are probabilistic approaches to taking educated guesses of the future given relevant information. This paper outlines a system HABITS which aims at overcoming weaknesses in existing RTLSs by using the human approach of making educated guesses about future location. The hypothesis of this proposal is that knowledge of a person's historical movement habits allows for future location predictions to be made in the short, medium and long-term.

The research questions that were foremost are whether the tracking capabilities of existing real-time locating systems can be improved automatically by knowledge of historical movement and by the application of a combination of artificial intelligence approaches. We also considered whether this approach can allow for intelligent prediction of future locations. We conclude that HABITS improves on the standard Ekahau RTLS in term of accuracy (overcoming black spots), latency (giving position fixes when Ekahau cannot), cost (less APs are required than are recommended by Ekahau) and prediction (short, medium and longer term predictions are available from HABITS). These are features that no other indoor tracking system currently provides.

References

Ashbrook, D. and Starner, T. (2002) 'Learning significant locations and predicting user movement with GPS', *Wearable Computers, 2002, (ISWC 2002), Proceedings, Sixth International Symposium on*, pp.101–108.

Bahl, P. and Padmanabhan, V.N. (2000) 'RADAR: an in-building RF-based user location and tracking system', *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*, IEEE, Vol. 2, pp.775–784.

Bahl, P., Padmanabhan, V.N. and Balachandran, A. (2000) ‘Enhancements to the RADAR user location and tracking system’, *Microsoft Research*.

Eagle, N. and Pentland, A., (2006) ‘Reality mining: sensing complex social systems’, *Personal and Ubiquitous Computing*, Vol. 10, No. 4, pp.255–268.

Fox, D. (2003) ‘Bayesian filtering for location estimation’, *IEEE Pervasive Computing*, Vol. 2, No. 3, pp.24–33.

Froehlich, J. and Krumm, J. (2008) ‘Route prediction from trip observations’, *SAE SP*, Vol. 2193, p.53.

Furey, E., Curran, K. and McKevitt, P. (2008) HABITS: a history aware based Wi-Fi indoor tracking system’, *PGNet: The 9th Annual Postgraduate Symposium: The Convergence of Telecommunications, Networking and Broadcasting 2008*, Liverpool, John Moores University, UK, pp.225–229.

Gellert, A. and Vintan, L. (2006) ‘Person movement prediction using hidden Markov models’, *Studies in Informatics and Control*, Vol. 15, p.17.

González, M.C., Hidalgo, C.A. and Barabási, A.L. (2008) ‘Understanding individual human mobility patterns’, *Nature*, Vol. 453, No. 7196, pp.779–782.

Han, S.J. (2004) ‘Learning and predicting user’s movements for intelligent location-based services’, MSc dissertation, Yonsei University.

Lassabe, F. (2009) ‘Geolocalisation et prediction dans les reseaux Wi-Fi en interieur’, Unpublished PhD thesis, Université de Franche-Comté, Besançon.

Petzold, J., Bagci, F., Trumler, W. et al. (2005) ‘Next location prediction within a Smart Office Building’, *1st International Workshop on Exploiting Context Histories in Smart Environments (ECHISE’05) at the 3rd International Conference on Pervasive Computing*, Munich, Germany, May.

Petzold, J., Bagci, F., Trumler, W. et al. (2006) ‘Hybrid predictors for next location prediction’, *Lecture Notes in Computer Science*, Vol. 4159, p.125.

Song, C., Qu, Z., Blumm, N. and Barabasi, A.L. (2010) ‘Limits of predictability in human mobility’, *Science*. Vol. 327, No. 5968, pp.1018–1021.

Vintan, L., Gellert, A., Petzold, J. and Ungerer, T. (2004) ‘Person movement prediction using neural networks’, *First Workshop on Modeling and Retrieval of Context*, Ulm, Germany.

Zhou, R. (2006) ‘Enhanced wireless indoor tracking system in multi-floor buildings with location predication’, *EUNIS 2006: Proceedings of the Conference on European University Information Systems*, pp.448–453.